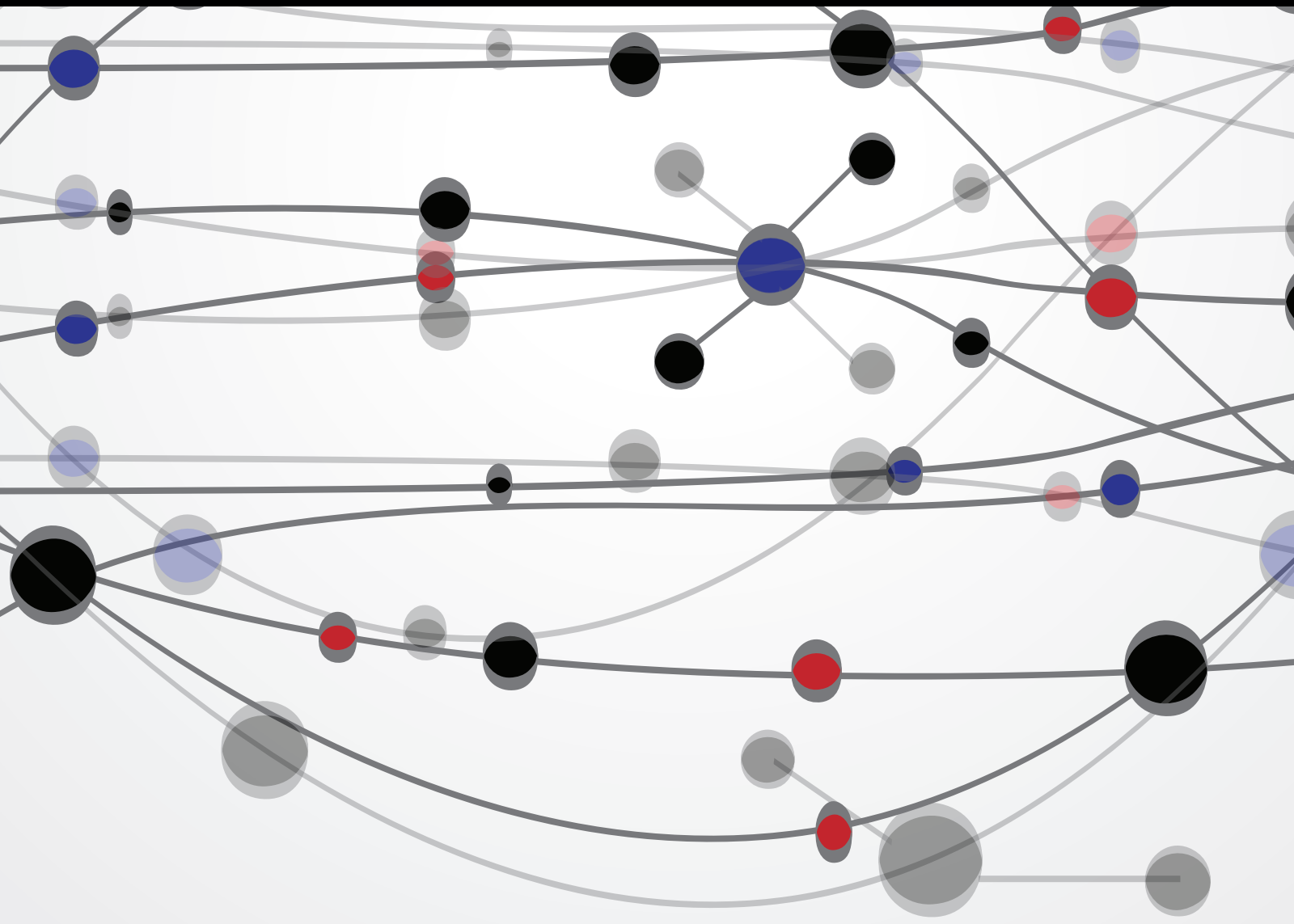
The Scientific World Journal

Multidimensional Signal Processing and Applications

Guest Editors: Salah Bourennane, Julien Marot, Caroline Fossati, Ahmed Bouridane, and Klaus Spinnler



Multidimensional Signal Processing and Applications

The Scientific World Journal

Multidimensional Signal Processing and Applications

Guest Editors: Salah Bourennane, Julien Marot, Caroline Fossati, Ahmed Bouridane, and Klaus Spinnler

Contents

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 365126, 2 pages
http://dx.doi.org/10.1155/2014/365126

Editorial
Multidimensional Signal Processing and Applications

Salah Bourennane,[1] Julien Marot,[1] Caroline Fossati,[1] Ahmed Bouridane,[2] and Klaus Spinnler[3]

[1] *Ecole Centrale Marseille, Université Aix-Marseille, Institut Fresnel-CNRS UMR 7249, Avenue Escadrille Normandie Niemen, 13013 Marseille, France*
[2] *Department of Computer Science and Digital Technologies, Northumbria University, Pandon Building, Newcastle upon Tyne NE2 1XE, UK*
[3] *Fraunhofer-Entwicklungszentrum Röntgentechnik EZRT, Fraunhofer-Instituts IIS, Flugplatzstr 75, 90768 Fürth, Germany*

Correspondence should be addressed to Salah Bourennane; salah.bourennane@fresnel.fr

Received 23 February 2014; Accepted 23 February 2014; Published 30 April 2014

In our daily lives and almost unconsciously, we deal with multidimensional data. From color images converted to the luminance and chrominance format to magnetic resonance images commonly acquired for health purposes, from different fashions to write an alphabet to array processing signals underlying any telecommunication system, we deal with multidimensional data.

In this special issue, we tried to show the variety of the topics which are currently investigated with multidimensional signal processing tools. The mathematical tools presented in this issue are as diverse as adaptive detectors, wavelet processing, principal component analysis, and improved classical image processing tools such as histogram equalization.

In the array processing paradigm, a two-dimensional matrix containing the data depends on the polarization properties of the sources, their number, and the number of sensors in the receiving antenna. Hence the interest of a multidimensional representation, including a polarization variable with two or three possible values, and a real and a complex part for the source amplitudes.

In the image processing paradigm, data are as various as magnetic resonance or color images, whose representation can be transferred from the RGB (red green blue) format to other spaces emphasizing for instance the luminance or the chrominance. It is shown how magnetic resonance brain images are classified with support vector machine. To avoid problems related to high dimensionality, which is current in big data processing, adequate features are extracted from the data by discrete wavelet transform and principal component analysis. Color spaces, which are useful for skin detection, for instance, are also further investigated: whatever the representation space is, a color image is a third order tensor, in other words, a three-dimensional data. It is shown how to detect image splicing with the help of merged features in the chrominance space: the relationships between pixels in a neighborhood are studied with a Markov process and the extraction of DCT features from the chrominance channel. Then, with the help of new color spaces, it is shown how evolved versions of neural networks called extreme learning machines can fuse multiple information such as color and local spatial information from face images. The "multi" aspect can also appear in the image processing paradigm when multiple images are obtained from several parameters. In images provided by synthetic aperture radar exploited for flood detection, contrast enhancement is achieved by an adjustable histogram equalization technique. For such an application where the visual aspect of the results are much important much, a color image, that is, a multidimensional signal, can be built from several two-dimensional result images, to get an informative map, where the color informs on the nature of the imaged scene, flooded or not, for instance. Starting from images, a set of multidimensional data is extracted from Serbian texts: the Serbian alphabet,

made of 30 letters, can be expressed in a Latin or in a Cyrillic fashion. All letters can be classified into four sets. By studying the frequencies of occurrence of each type of letter in a text, one can deduce that this text is written in the Latin or the Cyrillic fashion. In this application, matrices describing the cooccurrence in the distribution of the four types of letters are built out of any text, to make use of the classical texture features. Adapting the texture features to such a text recognition application, introducing a parameter which is the writing fashion, is a brand new idea. The "multi" aspect can also relate to multiresolution. Histogram of oriented gradients and hue descriptors can be merged to combine information related to the shape of an object and its color. By computing the merged data at several resolution levels, an innovative multidimensional descriptor is obtained. An application considered in this special issue is aircraft characterization and detection of images.

In a nutshell, the "multi" representation attracts the interest of researchers from very diverse application fields. Hopefully, this special issue will contribute in diffusing the models and tools of multidimensional signal processing to various application fields.

Salah Bourennane
Julien Marot
Caroline Fossati
Ahmed Bouridane
Klaus Spinnler

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 942923, 12 pages
http://dx.doi.org/10.1155/2014/942923

Research Article

The Generalized Sidelobe Canceller Based on Quaternion Widely Linear Processing

Jian-wu Tao[1] and Wen-xiu Chang[2]

[1] *Department of Flight Vehicle Control, Aviation University, Changchun, Jilin 130022, China*
[2] *College of Communication Engineering, Jilin University, Changchun, Jilin 130025, China*

Correspondence should be addressed to Jian-wu Tao; jianwu.tao@gmail.com

Received 4 August 2013; Accepted 2 January 2014; Published 29 April 2014

Academic Editors: S. Bourennane and J. Marot

We investigate the problem of quaternion beamforming based on widely linear processing. First, a quaternion model of linear symmetric array with two-component electromagnetic (EM) vector sensors is presented. Based on array's quaternion model, we propose the general expression of a quaternion *semiwidely* linear (QSWL) beamformer. Unlike the complex widely linear beamformer, the QSWL beamformer is based on the simultaneous operation on the quaternion vector, which is composed of two jointly proper complex vectors, and its involution counterpart. Second, we propose a useful implementation of QSWL beamformer, that is, QSWL generalized sidelobe canceller (GSC), and derive the simple expressions of the weight vectors. The QSWL GSC consists of two-stage beamformers. By designing the weight vectors of two-stage beamformers, the interference is completely canceled in the output of QSWL GSC and the desired signal is not distorted. We derive the array's gain expression and analyze the performance of the QSWL GSC in the presence of one type of interference. The advantage of QSWL GSC is that the main beam can always point to the desired signal's direction and the robustness to DOA mismatch is improved. Finally, simulations are used to verify the performance of the proposed QSWL GSC.

1. Introduction

As an important tool of multidimensional signal processing, the quaternion algebra has been applied to parameter estimation of 2D harmonic signals [1], DOA estimation of polarized signals [2–4], image processing [5], space-time-polarization block codes [6], Kalman filter [7], adaptive filter [8, 9], independent component analysis (ICA) algorithm [10], widely linear modeling and filtering [11–15], nonlinear adaptive filtering [16], and blind source separation [17]. In these applications, the interest in quaternion widely linear processing has recently increased due to the use of the full second-order statistical information in the quaternion domain. In [13], a widely linear quaternion least mean square (WL-QLMS) algorithm was presented to improve the accuracy in adaptive filtering of both second-order circular (Q-proper) and second-order noncircular (Q-improper) quaternion signals. In [14], a class of variable step-size algorithms were introduced into the WL-QLMS, so as to enhance the WL-QLMS tracking ability and sensibility to dynamically changing environments. In [15], a reduced-complexity WL-QLMS algorithm was developed to reduce computational cost and speed convergence of the WL-QLMS. In [16], a nonlinear quaternion adaptive filter and its widely linear version were proposed based on locally analytic nonlinear activation function. In [17], a quaternion widely linear predictor was applied to the blind source separation to extract both Q-proper and Q-improper sources. All the quaternion widely linear algorithms employ the quaternion widely linear model and the associated augmented quaternion statistics, which includes the information in both the standard covariance and the three pseudocovariances, so that their performance was enhanced. Motivated by the benefits of signal processing in quaternion domain, the quaternion beamformers [18–20] were recently developed. In [18], a quaternion minimum mean square error algorithm was proposed and applied to the beamforming of an airborne trimmed vector-sensor array. In [19], a quaternion-capon beamformer using a crossed dipole

array was proposed to improve the robustness of Capon beamformer. In [20], an interference and noise cancellation algorithm of quaternion MVDR beamformer was proposed to cancel the uncorrelated interference. Because the only information in the standard covariance of quaternion signals is used, the performance of these beamformers is not obviously enhanced. To make full use of the information in both the standard covariance and the three pseudocovariances of quaternion signals, the quaternion widely linear model may be introduced in the quaternion beamformer. Unfortunately, the quaternion beamformer based on widely linear model has received a little attention.

In this paper, we investigate the problem of quaternion beamforming based on widely linear processing. Because the quaternion output vector of array with two-component EM vector sensors is constructed by two jointly proper complex vectors, we present firstly the general expression of a quaternion *semiwidely* linear (QSWL) beamformer in this paper. Unlike the complex widely linear beamformers in [21–23], the QSWL beamformer can handle the proper complex vectors, whereas in this case the complex widely linear beamformer degenerates to conventional linear processing due to the vanishing of pseudocovariance. Thus, the QSWL beamformer can employ more information than the complex widely linear beamformers to improve the performance. Moreover, we propose a useful implementation of QSWL beamformer, that is, QSWL generalized sidelobe canceller (GSC), using a linear symmetric array. The QSWL GSC consists of two-stage beamformers. By designing the weight vectors of two-stage beamformers, the interference is completely canceled in the output of QSWL GSC and the desired signal is not distorted.

2. The Quaternion Widely Linear Beamformer for Vector-Sensor Array

2.1. Quaternion Model of Vector-Sensor Array. Consider that a scenario with one narrowband, completely polarized source, which is traveling in an isotropic and homogeneous medium, impinges on a uniform linear symmetric array from direction (θ, ϕ). This uniform linear symmetric array consists of $2M$ two-component vector sensors, which is depicted in Figure 1, and the spacing between the adjacent two vector sensors is assumed to be half wavelength. The array's reference point is at the array centroid and all vector sensors are indexed by $-M, \ldots, -1, 1, \ldots, M$ from left to right. Two complex series $x_{m1}(n)$ and $x_{m2}(n)$ are obtained from first and second components of the mth two-component vector sensor, respectively. The complex series $x_{m1}(n)$ and $x_{m2}(n)$ in the planes spanned by $\{1, i\}$, where i denotes a pure unit imaginary, are given by

$$\begin{bmatrix} x_{m1}(n) \\ x_{m2}(n) \end{bmatrix} = \begin{bmatrix} a_1(\theta, \phi, \alpha, \beta) \\ a_2(\theta, \phi, \alpha, \beta) \end{bmatrix} q_m(\theta, \phi)\, s(n), \quad (1)$$

where θ and ϕ denote the incidence source's elevation angle measured from the positive z-axis and azimuth angle measured from the positive x-axis, respectively. α represents

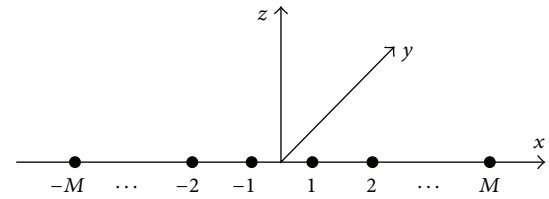


Figure 1: A uniform linear symmetric array.

the orientation angle, and β signifies the ellipticity angle. $a_1(\theta, \phi, \alpha, \beta)$ and $a_2(\theta, \phi, \alpha, \beta)$ are the responses on first and second components of two-component vector sensor, respectively. $q_m(\theta, \phi)$ is the spatial phase factor describing wavefield propagation along an array, and $q_m(\theta, \phi) = q^*_{-m}(\theta, \phi)$ due to the symmetric structure of uniform linear array. $s(n)$ is the complex envelope of the waveform, assumed to be a stationary stochastic process with zero mean and second-order circularity. Thus, $x_{m1}(n)$ and $x_{m2}(n)$ are also a stationary stochastic process with zero mean and second-order circularity.

Using $x_{m1}(n)$ and $x_{-m2}(n)$ $(m = -M, \ldots, M)$, the output of the mth two-component vector sensor can be written in the quaternion's *Cayley-Dickson* representation as

$$\begin{aligned} x_m(n) &= x_{m1}(n) + j x_{-m2}(n) \\ &= q_m(\theta, \phi)\left(a_1(\theta, \phi, \alpha, \beta) + j a_2(\theta, \phi, \alpha, \beta)\right) s(n) \\ &= q_m(\theta, \phi)\, P(\theta, \phi, \alpha, \beta)\, s(n), \end{aligned} \quad (2)$$

where j denotes a pure unit imaginary in a quaternion domain[1]. $P(\theta, \phi, \alpha, \beta) = a_1(\theta, \phi, \alpha, \beta) + j a_2(\theta, \phi, \alpha, \beta)$ is the quaternion-valued response on a two-component vector sensor. This transformation maps the complex signal $x_{m1}(n)$ on scalar and i imaginary fields of a quaternion, and the complex signal $x_{-m2}(n)$ is simultaneously mapped to the j and k imaginary fields of a quaternion. It is noted that $x_m(n)$ is C^i-proper[2] because two complex series $x_{m1}(n)$ and $x_{-m2}(n)$ are second-order circularity. When the quaternion-valued additive noise is considered, $x_m(n)$ can be rewritten as

$$x_m(n) = q_m(\theta, \phi)\, P(\theta, \phi, \alpha, \beta)\, s(n) + n_m(n), \quad (3)$$

where $n_m(n) = n_{m1}(n) + j n_{-m2}(n)$. $n_{m1}(n)$ and $n_{-m2}(n)$ are, respectively, the complex-valued additive noises at the first component of the mth vector sensor and the second component of the $-m$th vector sensor, which are assumed to be zero mean, Gaussian noise with identical covariance σ_n^2. And it is assumed that $n_m(n)$ and $n_k(n)$, where $m \neq k$, are uncorrelated.

In order to extend this model to the multisource case, we assume that K uncorrelated, completely polarized plane waves impinge on this array. One is the desired signal characterized by its arrival angles (θ_d, ϕ_d) and polarization parameters (α_d, β_d); the others are the interference characterized by its arrival angles (θ_u, ϕ_u) and polarization parameters

(α_u, β_u), where $u = 2, \ldots, K$. Thus, the quaternion-valued measurement vector of array can be written as

$$\begin{aligned}\mathbf{x}(n) &= \left[x_{-M}(n), \ldots, x_M(n)\right]^T \\ &= \mathbf{q}(\theta_d, \phi_d) P(\theta_d, \phi_d, \alpha_d, \beta_d) s_d(n) \\ &\quad + \sum_{u=2}^{K} \mathbf{q}(\theta_u, \phi_u) P(\theta_u, \phi_u, \alpha_u, \beta_u) s_u(n) + \mathbf{N}(n) \\ &= \mathbf{v}_d s_d(n) + \sum_{u=2}^{K} \mathbf{v}_u s_u(n) + \mathbf{N}(n),\end{aligned} \tag{4}$$

where $\mathbf{q}(\theta, \phi) = [q_{-M}(\theta, \phi), \ldots, q_M(\theta, \phi)]^T$ denotes the spatial phase vector. $\mathbf{v}_d = \mathbf{q}_d P_d$ and $\mathbf{v}_u = \mathbf{q}_u P_u$ are the quaternion-valued steering vector associated with the desired signal and the interference, respectively, where $\mathbf{q}(\theta_k, \phi_k)$ and $P(\theta_k, \phi_k, \alpha_k, \beta_k)$ are denoted by $\mathbf{q}_k$ and P_k $(k = d, u)$, respectively. $\mathbf{N}(n) = [n_{-M}(n), \ldots, n_M(n)]^T$ denotes the quaternion-valued additive noise vector. Because $x_m(n)$ $(m = -M, \ldots, M)$ is C^i-proper, $\mathbf{x}(n)$ is also C^i-proper vector.

2.2. The Quaternion Semiwidely Linear Beamformer. The most general linear processing is *full widely* linear processing, which consists in the simultaneous operation on the quaternion vector and its three involutions. Then, a quaternion widely linear beamformer can be written as

$$\begin{aligned}y(n) &= \mathbf{W}^{\triangleleft}\mathbf{x}(n) + \mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n) \\ &\quad + \mathbf{H}^{\triangleleft}\mathbf{x}^{(j)}(n) + \mathbf{F}^{\triangleleft}\mathbf{x}^{(k)}(n),\end{aligned} \tag{5}$$

where $\mathbf{W}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{F}$ are the quaternion-valued weight vectors. Superscript $(\cdot)^{\triangleleft}$ denotes the quaternion conjugate and transpose operator. $\mathbf{x}^{(i)}(n)$, $\mathbf{x}^{(j)}(n)$, and $\mathbf{x}^{(k)}(n)$ denote the quaternion involution of $\mathbf{x}(n)$ over a pure unit imaginary i, j, and k, respectively.

The *full widely* linear processing is optimal processing for the Q-improper quaternion vector. Since the quaternion-valued vector $\mathbf{x}(n)$ is C^i-proper vector, the optimal processing reduces to *semiwidely linear* processing[3]. Because the *semiwidely* linear processing consists only in the simultaneous operation on the quaternion vector and its involution over i, the general expression of a quaternion *semiwidely* linear (QSWL) beamformer can be written as

$$y(n) = \mathbf{W}^{\triangleleft}\mathbf{x}(n) + \mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n), \tag{6}$$

where $\mathbf{x}^{(i)}(n)$ is given by [11]

$$\begin{aligned}\mathbf{x}^{(i)}(n) &= -i\mathbf{x}(n)\, i \\ &= \mathbf{v}_d^{(i)} s_d(n) + \sum_{u=2}^{K} \mathbf{v}_u^{(i)} s_u(n) + \mathbf{N}^{(i)}(n).\end{aligned} \tag{7}$$

Moreover, we can write the quaternion-valued output series $y(n)$ in the following *Cayley-Dickson* representation:

$$\begin{aligned}y(n) &= \left(\left[\mathbf{W}^{\triangleleft}\mathbf{x}(n)\right]_1 + \left[\mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n)\right]_1\right) \\ &\quad + j\left(\left[\mathbf{W}^{\triangleleft}\mathbf{x}(n)\right]_2 + \left[\mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n)\right]_2\right) \\ &= y_1(n) + j y_2(n),\end{aligned} \tag{8}$$

where $[x]_1$ and $[x]_2$ denote, respectively, the first and the second complex-valued components of a quaternion x; that is, $x = [x]_1 + j[x]_2$. Thus, the QSWL beamformer has two complex-valued output series $y_1(n)$ and $y_2(n)$ in the planes spanned by $\{1, i\}$, where $y_1(n) = [\mathbf{W}^{\triangleleft}\mathbf{x}(n)]_1 + [\mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n)]_1$ and $y_2(n) = [\mathbf{W}^{\triangleleft}\mathbf{x}(n)]_2 + [\mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n)]_2$. Since the conventional "long vector" beamformer has only one complex-valued output series $y_1(n)$, the QSWL beamformer can obtain more information than the conventional "long vector" beamformer. The increase of information results in the improvement of QSWL beamformer's performance. In addition, we incorporate the information on both $\mathbf{x}(n)$ and $\mathbf{x}^{(i)}(n)$, so that the QSWL beamformer with different characteristics may be obtained by designing two weight vectors $\mathbf{W}$ and $\mathbf{G}$ under some different criterions.

3. The QSWL Generalized Sidelobe Canceller

In this section, a useful implementation of the QSWL beamformer, that is, QSWL generalized sidelobe canceller (GSC), is proposed. The QSWL GSC, which is depicted in Figure 2, consists of two-stage beamformers. In the first-stage beamformer (weight vector is $\mathbf{W}$), we attempt to extract a desired signal without any distortion from observed data. To cancel interference, we attempt to estimate interference in second-stage beamformer (weight vector is $\mathbf{G}$). By employing the output of the second-stage beamformer to cancel interference in the output of the first-stage beamformer, there is no interference in the output of the QSWL GSC. Compared with the conventional "long vector" beamformer, the advantages of two-stage beamformers are that the main beam can always point to desired signal's direction, even if the separation between the DOAs of the desired signal and interference is less, and the robustness to DOA mismatch is improved.

In the following, we derive the expressions of quaternion-valued weight vectors $\mathbf{W}$ in the first-stage beamformer and $\mathbf{G}$ in the second-stage beamformer. It is assumed that two $(K = 2)$ uncorrelated, completely polarized plane waves, whose waveform is unknown but whose DOA and polarization may be *priorly* estimated from techniques presented in [2, 4, 25–29], impinge on an array depicted in Figure 1. One plane wave is the desired signal and its complex envelope is denoted by $s_d(n)$; the other plane wave is the interference and its complex envelope is denoted by $s_u(n)$. Since $s_d(n)$ and $s_u(n)$ are complex series, the output of the QSWL GSC must be complex series. Because the quaternion-valued output has two complex-valued components in the planes spanned by $\{1, i\}$, we define the first complex-valued output component

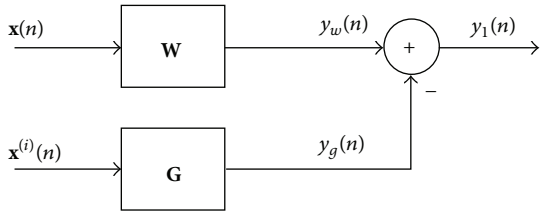


Figure 2: The structure of QSWL GSC.

as the output of the QSWL GSC. Thus, the complex-valued output of the QSWL GSC is written as

$$y_{\mathrm{GSC}}(n) = \left[y_Q(n)\right]_1 = y_w(n) - y_g(n), \tag{9}$$

where $y_w(n)$ is the complex-valued output of the first-stage beamformer; that is, $y_w(n) = [\mathbf{W}^{\triangleleft}\mathbf{x}(n)]_1$; $y_g(n)$ is the complex-valued output of the second-stage beamformer; that is, $y_g(n) = [\mathbf{G}^{\triangleleft}\mathbf{x}^{(i)}(n)]_1$.

3.1. The First-Stage Beamformer. From (4), we have

$$\begin{aligned} y_w(n) = {} & \left[\mathbf{W}^{\triangleleft}\mathbf{v}_d\right]_1 s_d(n) + \left[\mathbf{W}^{\triangleleft}\mathbf{v}_u\right]_1 s_u(n) \\ & + \left[\mathbf{W}^{\triangleleft}\mathbf{N}(n)\right]_1. \end{aligned} \tag{10}$$

In the first-stage beamformer, we attempt to minimize the interference-plus-noise energy in $y_w(n)$, subject to the constraint $[\mathbf{W}^{\triangleleft}\mathbf{v}_d]_1 = 1$.

Since the *Cayley-Dickson* representation of $\mathbf{W}$, $\mathbf{v}_d$, $\mathbf{v}_u$, and $\mathbf{N}(n)$ is, respectively, $\mathbf{W} = \mathbf{W}_1 + j\mathbf{W}_2$, $\mathbf{v}_d = \mathbf{v}_{d1} + j\mathbf{v}_{d2}$, $\mathbf{v}_u = \mathbf{v}_{u1} + j\mathbf{v}_{u2}$, and $\mathbf{N}(n) = \mathbf{N}_1(n) + j\mathbf{N}_2(n)$, we have

$$\left[\mathbf{W}^{\triangleleft}\mathbf{v}_d\right]_1 = \mathbf{W}_1^H\mathbf{v}_{d1} + \mathbf{W}_2^H\mathbf{v}_{d2} = \overline{\mathbf{W}}^H\overline{\mathbf{V}}_d, \tag{11}$$

$$\left[\mathbf{W}^{\triangleleft}\mathbf{v}_u\right]_1 = \mathbf{W}_1^H\mathbf{v}_{u1} + \mathbf{W}_2^H\mathbf{v}_{u2} = \overline{\mathbf{W}}^H\overline{\mathbf{V}}_u, \tag{12}$$

$$\left[\mathbf{W}^{\triangleleft}\mathbf{N}(n)\right]_1 = \mathbf{W}_1^H\mathbf{N}_1(n) + \mathbf{W}_2^H\mathbf{N}_2(n) = \overline{\mathbf{W}}^H\overline{\mathbf{N}}(n), \tag{13}$$

where

$$\begin{aligned} \overline{\mathbf{W}} &= \begin{bmatrix}\mathbf{W}_1 \\ \mathbf{W}_2\end{bmatrix}, \qquad \overline{\mathbf{V}}_d = \begin{bmatrix}\mathbf{v}_{d1} \\ \mathbf{v}_{d2}\end{bmatrix}, \\ \overline{\mathbf{V}}_u &= \begin{bmatrix}\mathbf{v}_{u1} \\ \mathbf{v}_{u2}\end{bmatrix}, \qquad \overline{\mathbf{N}}(n) = \begin{bmatrix}\mathbf{N}_1(n) \\ \mathbf{N}_2(n)\end{bmatrix}. \end{aligned} \tag{14}$$

Superscript $(\cdot)^H$ denotes the complex conjugate and transpose operator. Thus, (10) can be rewritten as

$$y_w(n) = \overline{\mathbf{W}}^H\overline{\mathbf{V}}_d s_d(n) + \overline{\mathbf{W}}^H\overline{\mathbf{V}}_u s_u(n) + \overline{\mathbf{W}}^H\overline{\mathbf{N}}(n). \tag{15}$$

Then, $\overline{\mathbf{W}}$ can be derived by solving the following constrained optimization problem:

$$J\left(\overline{\mathbf{W}}\right) = \min\left\{\overline{\mathbf{W}}^H\mathbf{R}_{in}\overline{\mathbf{W}}\right\}, \quad \text{subject to } \overline{\mathbf{W}}^H\overline{\mathbf{V}}_d = 1, \tag{16}$$

where

$$\mathbf{R}_{in} = \begin{bmatrix} E\left\{[\mathbf{x}_{in}(n)]_1[\mathbf{x}_{in}(n)]_1^H\right\} & E\left\{[\mathbf{x}_{in}(n)]_1[\mathbf{x}_{in}(n)]_2^H\right\} \\ E\left\{[\mathbf{x}_{in}(n)]_2[\mathbf{x}_{in}(n)]_1^H\right\} & E\left\{[\mathbf{x}_{in}(n)]_2[\mathbf{x}_{in}(n)]_2^H\right\} \end{bmatrix} \tag{17}$$

is the covariance matrix and $\mathbf{x}_{in}(n) = \mathbf{v}_u s_u(n) + \mathbf{N}(n)$ is the measurement vector of array in the absence of the desired signal. The solution of this constrained optimization problem is obtained by using Lagrange multipliers; that is,

$$\overline{\mathbf{W}} = \frac{\mathbf{R}_{in}^{-1}\overline{\mathbf{V}}_d}{\overline{\mathbf{V}}_d^H\mathbf{R}_{in}^{-1}\overline{\mathbf{V}}_d}. \tag{18}$$

If the interference is uncorrelated with the additive noise, $\overline{\mathbf{W}}$ can be written in the simple form (the proof is in Appendix A)

$$\overline{\mathbf{W}} = \frac{\varepsilon\overline{\mathbf{V}}_d - \left[P_u^{\triangleleft}P_d\right]_1\mathbf{q}_u^H\mathbf{q}_d\overline{\mathbf{V}}_u}{\mu}, \tag{19}$$

where

$$\begin{aligned} \mu &= 2M\left|P_d\right|^2\varepsilon - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2, \\ \varepsilon &= \xi_u^{-1} + 2M\left|P_u\right|^2, \end{aligned} \tag{20}$$

where ξ_u denotes the input interference-to-noise ratio (INR). Moreover, the quaternion-valued optimal weight vector $\mathbf{W}$ may be given by

$$\mathbf{W} = \mathbf{J}_1\overline{\mathbf{W}} + j\mathbf{J}_2\overline{\mathbf{W}}, \tag{21}$$

where $\mathbf{J}_1 = [\mathbf{I}_{2M\times 2M}, \mathbf{0}_{2M\times 2M}]$ and $\mathbf{J}_2 = [\mathbf{0}_{2M\times 2M}, \mathbf{I}_{2M\times 2M}]$ are two selection matrices. It is noted that in some applications, such as Radar, $\mathbf{R}_{in}$ may be estimated in intervals of no transmitted signal. But, $\mathbf{R}_{in}$ is not obtained in other applications, such as Communications. In these applications, we may replace $\mathbf{R}_{in}$ by $\mathbf{R}_x$, where

$$\mathbf{R}_x = \begin{bmatrix} E\left\{[\mathbf{x}(n)]_1[\mathbf{x}(n)]_1^H\right\} & E\left\{[\mathbf{x}(n)]_1[\mathbf{x}(n)]_2^H\right\} \\ E\left\{[\mathbf{x}(n)]_2[\mathbf{x}(n)]_1^H\right\} & E\left\{[\mathbf{x}(n)]_2[\mathbf{x}(n)]_2^H\right\} \end{bmatrix} \tag{22}$$

is the covariance matrix and $\mathbf{x}(n) = \mathbf{v}_d s_d(n) + \mathbf{v}_u s_u(n) + \mathbf{N}(n)$ is the measurement vector of array. When the distortionless constraint is perfectly matched with the desired signal, the weight vector $\mathbf{W}$ is identical in both $\mathbf{R}_x$ and $\mathbf{R}_{in}$ [30].

By using the optimal weight vector $\mathbf{W}$, the complex output of the first-stage beamformer can be given by

$$y_w(n) = s_d(n) + \left[\mathbf{W}^{\triangleleft}\mathbf{v}_u\right]_1 s_u(n) + \left[\mathbf{W}^{\triangleleft}\mathbf{N}(n)\right]_1. \tag{23}$$

3.2. The Second-Stage Beamformer. From (7), we have

$$\begin{aligned} y_g(n) = {} & \left[\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)}\right]_1 s_d(n) + \left[\mathbf{G}^{\triangleleft}\mathbf{v}_u^{(i)}\right]_1 s_u(n) \\ & + \left[\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1. \end{aligned} \tag{24}$$

In the second-stage beamformer, we attempt to minimize the noise energy in $y_g(n)$, subject to the constraints $[\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)}]_1 = 0$ and $[\mathbf{G}^{\triangleleft}\mathbf{v}_u^{(i)}]_1 = [\mathbf{W}^{\triangleleft}\mathbf{v}_u]_1$. In the following, two schemes are presented to implement this aim.

3.2.1. Scheme 1, That Is, Combined QPMC and MVDR. Let $\mathbf{G} = \mathbf{w}_{QS}\mathbf{w}_{MV}$, where $\mathbf{w}_{QS}$ is a quaternion-valued diagonal weight matrix and $\mathbf{w}_{MV}$ is a complex weight vector. In this scheme, the first step is to achieve the constraint $[\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)}]_1 = 0$ by designing $\mathbf{w}_{QS}$, and this step is referred to quaternion polarization matched cancellation (QPMC); the second step is to minimize the noise energy in $y_g(n)$ subject to the constraint $[\mathbf{G}^{\triangleleft}\mathbf{v}_u^{(i)}]_1 = [\mathbf{W}^{\triangleleft}\mathbf{v}_u]_1$ by designing $\mathbf{w}_{MV}$, and this step is referred to MVDR.

Let $\mathbf{w}_{QS} = \mathrm{diag}\{w_{QS}(-M), \ldots, w_{QS}(M)\}$; then we have

$$\begin{aligned}\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)} &= \mathbf{w}_{MV}^H\mathbf{w}_{QS}^{\triangleleft}\mathbf{v}_d^{(i)} \\ &= \mathbf{w}_{MV}^H\begin{bmatrix} w_{QS}^*(-M)\, q_{-M}(\theta_d,\phi_d)\, P_d^{(i)} \\ \vdots \\ w_{QS}^*(M)\, q_M(\theta_d,\phi_d)\, P_d^{(i)} \end{bmatrix},\end{aligned} \tag{25}$$

where superscript $(\cdot)^*$ denotes the quaternion conjugate operator. From (25) and the constraint $[\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)}]_1 = 0$, we have the constraint $[w_{QS}^*(m)q_m(\theta_d,\phi_d)P_d^{(i)}]_1 = 0$, where $m = -M, \ldots, M$. If $w_{QS}(m) = q_m(\theta_d,\phi_d)(a_{d2}^* + ja_{d1}^*)$, where $a_{d1} = a_1(\theta_d,\phi_d,\alpha_d,\beta_d)$ and $a_{d2} = a_2(\theta_d,\phi_d,\alpha_d,\beta_d)$, this constraint is satisfied. Thus, we can obtain

$$\mathbf{w}_{QS} = \mathrm{diag}\left(\mathbf{q}_d\right)\left(a_{d2}^* + ja_{d1}^*\right), \tag{26}$$

where $\mathrm{diag}(\mathbf{q}_d) = \mathrm{diag}\{q_{-M}(\theta_d,\phi_d), \ldots, q_M(\theta_d,\phi_d)\}$.

Inserting $\mathbf{G} = \mathbf{w}_{QS}\mathbf{w}_{MV}$ into (24), $y_g(n)$ can be rewritten as

$$y_g(n) = \mathbf{w}_{MV}^H\left[\mathbf{w}_{QS}^{\triangleleft}\mathbf{v}_u^{(i)}\right]_1 s_u(n) + \mathbf{w}_{MV}^H\left[\mathbf{w}_{QS}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1. \tag{27}$$

Then, $\mathbf{w}_{MV}$ can be derived by solving the following constrained optimization problem:

$$\begin{aligned} J(\mathbf{w}_{MV}) &= \min\left\{\mathbf{w}_{MV}^H\mathbf{R}_{QS}\mathbf{w}_{MV}\right\}, \\ &\text{subject to } \mathbf{w}_{MV}^H\widetilde{\mathbf{V}}_u = \overline{\mathbf{W}}^H\overline{\mathbf{V}}_u, \end{aligned} \tag{28}$$

where $\mathbf{R}_{QS} = E\{[\mathbf{w}_{QS}^{\triangleleft}\mathbf{x}^{(i)}(n)]_1[\mathbf{w}_{QS}^{\triangleleft}\mathbf{x}^{(i)}(n)]_1^H\}$ is the covariance matrix and $\widetilde{\mathbf{V}}_u = [\mathbf{w}_{QS}^{\triangleleft}\mathbf{v}_u^{(i)}]_1$. The solution of this constrained optimization problem is obtained by using Lagrange multipliers; that is,

$$\mathbf{w}_{MV} = \frac{\mathbf{R}_{QS}^{-1}\widetilde{\mathbf{V}}_u}{\widetilde{\mathbf{V}}_u^H\mathbf{R}_{QS}^{-1}\widetilde{\mathbf{V}}_u}\overline{\mathbf{V}}_u^H\overline{\mathbf{W}}. \tag{29}$$

If the desired signal and interference are uncorrelated with the additive noise, $\mathbf{w}_{MV}$ can be written in the simple form (the proof is in Appendix B)

$$\mathbf{w}_{MV} = \frac{g_1}{\kappa}\widetilde{\mathbf{V}}_u, \tag{30}$$

where

$$\begin{aligned}\kappa &= \widetilde{\mathbf{V}}_u^H\widetilde{\mathbf{V}}_u = 2M\left(|a_{d2}|^2|a_{u1}|^2 + |a_{d1}|^2|a_{u2}|^2\right) \\ &\quad - 2\Re\left(a_{d1}a_{d2}^*a_{u2}a_{u1}^*\left(\mathbf{q}_u^2\right)^H\mathbf{q}_d^2\right), \\ g_1 &= \left(\overline{\mathbf{W}}^H\overline{\mathbf{V}}_u\right)^H = \frac{\xi_u^{-1}[P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d}{\mu},\end{aligned} \tag{31}$$

where $\Re(\cdot)$ denotes the real part of a complex number. μ is given by (20).

3.2.2. Scheme 2, That Is, LCMV. In this scheme, we employ the LCMV beamformer as the second-stage beamformer. Since the *Cayley-Dickson* representation of $\mathbf{G}$, $\mathbf{v}_d^{(i)}$, $\mathbf{v}_u^{(i)}$, and $\mathbf{N}^{(i)}(n)$ are, respectively, $\mathbf{G} = \mathbf{G}_1 + j\mathbf{G}_2$, $\mathbf{v}_d^{(i)} = \mathbf{v}_{d1} - j\mathbf{v}_{d2}$, $\mathbf{v}_u^{(i)} = \mathbf{v}_{u1} - j\mathbf{v}_{u2}$, and $\mathbf{N}^{(i)}(n) = \mathbf{N}_1(n) - j\mathbf{N}_2(n)$, we have

$$\left[\mathbf{G}^{\triangleleft}\mathbf{v}_d^{(i)}\right]_1 = \mathbf{G}_1^H\mathbf{v}_{d1} - \mathbf{G}_2^H\mathbf{v}_{d2} = \overline{\mathbf{G}}^H\overline{\mathbf{V}}_d^{(i)}, \tag{32}$$

$$\left[\mathbf{G}^{\triangleleft}\mathbf{v}_u^{(i)}\right]_1 = \mathbf{G}_1^H\mathbf{v}_{u1} - \mathbf{G}_2^H\mathbf{v}_{u2} = \overline{\mathbf{G}}^H\overline{\mathbf{V}}_u^{(i)}, \tag{33}$$

$$\left[\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1 = \mathbf{G}_1^H\mathbf{N}_1(n) - \mathbf{G}_2^H\mathbf{N}_2(n) = \overline{\mathbf{G}}^H\overline{\mathbf{N}}^{(i)}(n), \tag{34}$$

where

$$\begin{aligned}\overline{\mathbf{G}} &= \begin{bmatrix}\mathbf{G}_1 \\ \mathbf{G}_2\end{bmatrix}, \qquad \overline{\mathbf{V}}_d^{(i)} = \begin{bmatrix}\mathbf{v}_{d1} \\ -\mathbf{v}_{d2}\end{bmatrix}, \\ \overline{\mathbf{V}}_u^{(i)} &= \begin{bmatrix}\mathbf{v}_{u1} \\ -\mathbf{v}_{u2}\end{bmatrix}, \qquad \overline{\mathbf{N}}^{(i)}(n) = \begin{bmatrix}\mathbf{N}_1(n) \\ -\mathbf{N}_2(n)\end{bmatrix}.\end{aligned} \tag{35}$$

Thus, (24) can be rewritten as

$$y_g(n) = \overline{\mathbf{G}}^H\overline{\mathbf{V}}_d^{(i)}s_d(n) + \overline{\mathbf{G}}^H\overline{\mathbf{V}}_u^{(i)}s_u(n) + \overline{\mathbf{G}}^H\overline{\mathbf{N}}^{(i)}(n). \tag{36}$$

Then, $\overline{\mathbf{G}}$ can be derived by solving the following constrained optimization problem:

$$J\left(\overline{\mathbf{G}}\right) = \min\left\{\overline{\mathbf{G}}^H\mathbf{R}_{inG}\overline{\mathbf{G}}\right\}, \quad \text{subject to } \overline{\mathbf{G}}^H\mathbf{C} = \mathbf{g}^H, \tag{37}$$

where

$$\mathbf{R}_{inG} = \begin{bmatrix} E\left\{\left[\mathbf{x}_{in}^i(n)\right]_1\left[\mathbf{x}_{in}^i(n)\right]_1^H\right\} & E\left\{\left[\mathbf{x}_{in}^i(n)\right]_1\left[\mathbf{x}_{in}^i(n)\right]_2^H\right\} \\ E\left\{\left[\mathbf{x}_{in}^i(n)\right]_2\left[\mathbf{x}_{in}^i(n)\right]_1^H\right\} & E\left\{\left[\mathbf{x}_{in}^i(n)\right]_2\left[\mathbf{x}_{in}^i(n)\right]_2^H\right\} \end{bmatrix} \tag{38}$$

is the covariance matrix and $\mathbf{x}_{in}^i(n) = \mathbf{v}_u^i s_u(n) + \mathbf{N}^i(n)$ is the quaternion involution of $\mathbf{x}_{in}(n)$. $\mathbf{C} = [\overline{\mathbf{V}}_u^{(i)}, \overline{\mathbf{V}}_d^{(i)}]$ and $\mathbf{g}^H = [g_1^H, 0]$, where g_1 is given by (31). The solution of (37) is given by [30]

$$\overline{\mathbf{G}} = \mathbf{R}_{inG}^{-1}\mathbf{C}\left(\mathbf{C}^H\mathbf{R}_{inG}^{-1}\mathbf{C}\right)^{-1}\mathbf{g}. \tag{39}$$

If the desired signal and interference are uncorrelated with the additive noise, $\overline{\mathbf{G}}$ can be written in the simple form (the proof is in Appendix C)

$$\overline{\mathbf{G}} = \frac{g_1}{\nu}\left(2M|P_d|^2\overline{\mathbf{V}}_u^{(i)} - \left[P_d^{\triangleleft}P_u\right]_1\mathbf{q}_d^H\mathbf{q}_u\overline{\mathbf{V}}_d^{(i)}\right), \tag{40}$$

where

$$\begin{aligned}\nu &= (2M)^2|P_d|^2|P_u|^2 - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2 \\ &= \mu - \xi_u^{-1}2M|P_d|^2,\end{aligned} \tag{41}$$

where μ is given by (20). Moreover, the quaternion-valued optimal weight vector $\mathbf{G}$ may be given by

$$\mathbf{G} = \mathbf{J}_1\overline{\mathbf{G}} + j\mathbf{J}_2\overline{\mathbf{G}}, \tag{42}$$

where $\mathbf{J}_1 = [\mathbf{I}_{2M\times 2M}, \mathbf{0}_{2M\times 2M}]$ and $\mathbf{J}_2 = [\mathbf{0}_{2M\times 2M}, \mathbf{I}_{2M\times 2M}]$ are two selection matrices.

By using the optimal weight vector $\mathbf{G}$, the complex output of second-stage beamformer can be given by

$$y_g(n) = \left[\mathbf{W}^{\triangleleft}\mathbf{v}_u\right]_1 s_u(n) + \left[\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1. \tag{43}$$

Thus, the complex output of QSWL GSC may be rewritten as

$$\begin{aligned} y_{\mathrm{GSC}}(n) &= y_w(n) - y_g(n) \\ &= s_d(n) + \left[\mathbf{W}^{\triangleleft}\mathbf{N}(n)\right]_1 - \left[\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1. \end{aligned} \tag{44}$$

From the above equation, we see that the interference component is completely canceled in the output $y_{\mathrm{GSC}}(n)$.

3.3. The Performance Analysis. Since the QSWL GSC can totally remove the interference, its output signal-to-interference ratio (SIR) tends to infinite. Thus, we focus our attention on the output signal-to-noise ratio (SNR) and array's gain. Let $\mathsf{p}_n = E\{|[\mathbf{W}^{\triangleleft}\mathbf{N}(n)]_1 - [\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)]_1|^2\}$ be the power of output noise. From (13) and (34), we have

$$\begin{aligned} &\left[\mathbf{W}^{\triangleleft}\mathbf{N}(n)\right]_1 - \left[\mathbf{G}^{\triangleleft}\mathbf{N}^{(i)}(n)\right]_1 \\ &\quad= \mathbf{W}_1^H\mathbf{N}_1(n) + \mathbf{W}_2^H\mathbf{N}_2(n) \\ &\qquad - \mathbf{G}_1^H\mathbf{N}_1(n) + \mathbf{G}_2^H\mathbf{N}_2(n) \\ &\quad= \left(\mathbf{W}_1^H - \mathbf{G}_1^H\right)\mathbf{N}_1(n) + \left(\mathbf{W}_2^H + \mathbf{G}_2^H\right)\mathbf{N}_2(n). \end{aligned} \tag{45}$$

Then, p_n can be written as

$$\begin{aligned} \mathsf{p}_n &= \sigma_n^2\left(\mathbf{W}_1^H - \mathbf{G}_1^H\right)\left(\mathbf{W}_1 - \mathbf{G}_1\right) \\ &\quad + \sigma_n^2\left(\mathbf{W}_2^H + \mathbf{G}_2^H\right)\left(\mathbf{W}_2 + \mathbf{G}_2\right) \\ &= \sigma_n^2\left(\overline{\mathbf{W}}^H\overline{\mathbf{W}} + \overline{\mathbf{G}}^H\overline{\mathbf{G}} - \mathbf{W}_1^H\mathbf{G}_1 + \mathbf{W}_2^H\mathbf{G}_2\right. \\ &\qquad \left. -\mathbf{G}_1^H\mathbf{W}_1 + \mathbf{G}_2^H\mathbf{W}_2\right). \end{aligned} \tag{46}$$

When the combined QPMC and MVDR are adopted in the second-stage beamformer, p_n can be written in the simple form (the proof is in Appendix D)

$$\mathsf{p}_n = \frac{\sigma_n^2}{\mu^2}\left(2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_q\right), \tag{47}$$

where

$$\gamma_q = \frac{\xi_u^{-2}|P_d|^2}{\kappa} - 2M|P_u|^2, \tag{48}$$

where κ is given by (31). From (44), the expression of output SNR and array's gain A_q may be written as

$$\begin{aligned} \mathrm{SNR}_o &= \xi_d\frac{\left(2M|P_d|^2\varepsilon - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\right)^2}{2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_q}, \\ A_q &= \frac{\left(2M|P_d|^2\varepsilon - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\right)^2}{2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_q}, \end{aligned} \tag{49}$$

where ξ_d denotes the input signal-to-noise ratio (SNR) and ε is given by (20).

When the LCMV is adopted in the second-stage beamformer, p_n can be written in the simple form (the proof is in Appendix E)

$$\mathsf{p}_n = \frac{\sigma_n^2}{\mu^2}\left(2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_l\right), \tag{50}$$

where

$$\gamma_l = \frac{2M\xi_u^{-2}|P_d|^2}{\nu} - 2M|P_u|^2, \tag{51}$$

where ν is given by (41). Then, the expression of output SNR and array's gain A_l may be written as

$$\begin{aligned} \mathrm{SNR}_o &= \xi_d\frac{\left(2M|P_d|^2\varepsilon - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\right)^2}{2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_l}, \\ A_l &= \frac{\left(2M|P_d|^2\varepsilon - \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\right)^2}{2M|P_d|^2\varepsilon^2 + \left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\left|\mathbf{q}_u^H\mathbf{q}_d\right|^2\gamma_l}. \end{aligned} \tag{52}$$

From (49) and (52), we can see that the output SNR and array's gain depend on not only separation between the DOA's of the desired signal and interference (i.e., $|\mathbf{q}_u^H\mathbf{q}_d|$), but also difference between the polarization of the desired signal and interference (i.e., $|[P_u^{\triangleleft}P_d]_1|$). The dependency of them on $|\mathbf{q}_u^H\mathbf{q}_d|$ and $|[P_u^{\triangleleft}P_d]_1|$ is shown in the following consequences.

(1) When $|\mathbf{q}_u^H\mathbf{q}_d| = 0$, the separation between the DOAs of the desired signal and interference reaches to the maximum. In this case, $A_q = A_l = 2M|P_d|^2$. Further, $|\mathbf{q}_u^H\mathbf{q}_d|$ increases with the decrease of the DOA's separation. Thus, the array's gain of A_q and A_l will reduce if $|P_d|^2$ is a constant. When $\mathbf{q}_u = \mathbf{q}_d$, $|\mathbf{q}_u^H\mathbf{q}_d| = 2M$. This implies that there is no separation between the DOAs of the desired signal and interference. In this case, the array's gain is given by

$$A_q = A_l = \frac{2M\left(|P_d|^2\varepsilon - 2M\left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\right)^2}{|P_d|^2\varepsilon^2 + 2M\left|\left[P_u^{\triangleleft}P_d\right]_1\right|^2\gamma}, \tag{53}$$

where

$$\gamma = \frac{\xi_u^{-2}|P_d|^2}{2M\left(|P_d|^2|P_u|^2 - |[P_u^{\triangleleft}P_d]_1|^2\right)} - 2M|P_u|^2. \quad (54)$$

Further, $P_u = P_d$ if $\alpha_d = \alpha_u$ and $\beta_d = \beta_u$. Thus, the array's gain $A_q = A_l = 0$ because $\gamma = \infty$. This implies that the QSWL GSC fails.

(2) When $|[P_u^{\triangleleft}P_d]_1| = 0$, $A_q = A_l = 2M|P_d|^2$. In the cases that $\theta_d = \theta_u \neq 0$ and $\phi_d = \phi_u \neq 0$ (i.e., $\mathbf{q}_u = \mathbf{q}_d$), we have $|[P_u^{\triangleleft}P_d]_1| = (\sin^2\theta_d\cos^2\phi_d + \sin^2\phi_d)\cos(\alpha_u - \alpha_d)\cos(\beta_u - \beta_d)$. If $\alpha_u - \alpha_d = \pm\pi/2$ or $\beta_u - \beta_d = \pm\pi/2$, then $|[P_u^{\triangleleft}P_d]_1| = 0$. This implies that, even though there is no separation between the DOAs of the desired signal and interference, the array's gain can also reach $2M|P_d|^2$ by using the orthogonality between the polarization of the desired signal and interference. Further, the array's gain decreases with the increase of $|[P_u^{\triangleleft}P_d]_1|$ if $|P_d|^2$ is a constant. When $P_u = P_d$, $|[P_u^{\triangleleft}P_d]_1| = |P_d|^2$. This implies that there is no difference between the polarization of the desired signal and interference. But, the array's gain is not equal to zero if $\mathbf{q}_u \neq \mathbf{q}_d$.

In addition, the output SNR and array's gain depend also on the input INR ξ_u, the array's element number $2M$, the interference response's power $|P_u|^2$, and the desired signal response's power $|P_d|^2$.

4. Monte Carlo Simulations

In this section, we investigate the performance of the proposed QSWL GSC (i.e., QSWL-QPMC-MVDR and QSWL-LCMV). In simulations, we consider a two-component vector-sensor array depicted in Figure 1, where each two-component vector sensor consists of one electric dipole and one magnetic loop coaligned along the x-axis. For each two-component vector sensor, the responses on the first and second components are given by

$$\begin{bmatrix} a_1(\theta,\phi,\alpha,\beta) \\ a_2(\theta,\phi,\alpha,\beta) \end{bmatrix} = \begin{bmatrix} -\sin\phi & -\sin\theta\cos\phi \\ -\sin\theta\cos\phi & \sin\phi \end{bmatrix} \times \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta \\ i\sin\beta \end{bmatrix}, \quad (55)$$

where $\theta \in (-\pi/2, \pi/2]$; $\phi \in (-\pi, \pi]$; $\alpha \in (-\pi/2, \pi/2]$; and $\beta \in [-\pi/4, \pi/4]$. The spatial phase factor vector $\mathbf{q}(\theta,\phi) = [e^{-i(M\pi/2)\sin\theta\cos\phi}, \ldots, e^{i(M\pi/2)\sin\theta\cos\phi}]^T$. To prevent performance degradation in the presence of DOA mismatch and/or array perturbations, we use the three main-lobe constraints [30], that is, $\mathbf{C} = [\overline{\mathbf{V}}_d(\theta_d), \overline{\mathbf{V}}_d(\theta_d - \theta_c), \overline{\mathbf{V}}_d(\theta_d + \theta_c)]$ and $\mathbf{g} = [1, 1, 1]^H$, in all experiments. It is assumed that $\xi_d = 10$ and $\xi_u = 1$. To compare the performance, the complex "long vector" LCMV (CLCMV) beamformer [30–32] is also included in simulation results.

4.1. Performance for One Type of Interference. In the presence of a single interference, we illustrate the performance of the proposed QSWL GSC by employing two experiments. We assume that $M = 6$, $\phi_u = \phi_d = 60^\circ$, $\alpha_u = \alpha_d = 30^\circ$, $\beta_u = \beta_d = 30^\circ$, and $\theta_c = 1^\circ$. In the first experiment, we investigate the effect of the angular separation $\Delta\theta$ between the desired signal and the interference, where $\Delta\theta = \theta_u - \theta_d$. Figure 3 displays the output SINR as a function of $\Delta\theta$, where $\theta_u = \Delta\theta$, $\theta_d = 0^\circ$. From Figure 3, it is seen that the output SINR of the proposed QSWL GSC is larger than that of the CLCMV in small angular separation (i.e., $\Delta\theta < 40^\circ$). Compared with the QSWL-QPMC-MVDR, the QSWL-LCMV has a little increase of the output SINR. This implies that the QSWL GSC has the better performance in small angular separation.

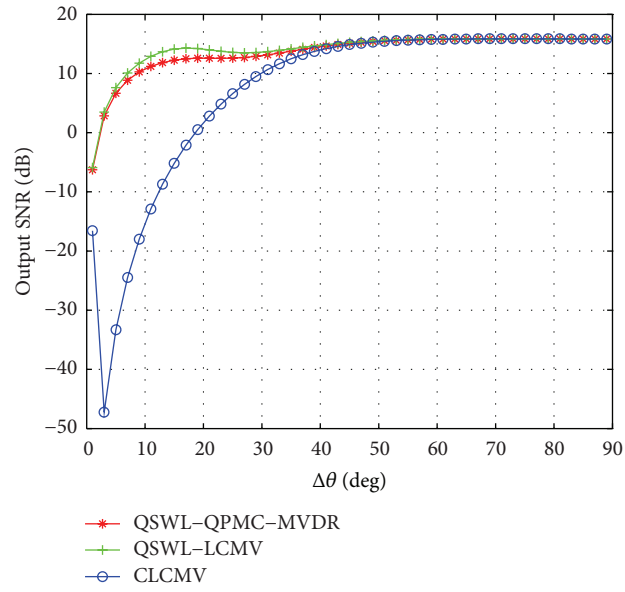


Figure 3: Effect of the angular separation $\Delta\theta$ between the interference and the desired signal.

In the second experiment, we assume that the covariance matrix $\mathbf{R}_x$, instead of $\mathbf{R}_{in}$, is available. Figure 4 displays the power patterns for three values of $|\Delta\theta|$: 60°, 20°, and 10°, where $\theta_d = |\Delta\theta|$, $\theta_u = 0^\circ$. From Figure 4, it is seen that the power is almost zero towards the interference's DOA (located at 0°) in all cases. When $|\Delta\theta|$ decreases, the main lobe of the QSWL GSC points almost to the source location, but the main lobe of the CLCMV is away from the source location. This implies that the QSWL GSC outperforms obviously CLCMV as the desired signal moves towards the interference. In addition, the side lobes are amplified with a decrease of $|\Delta\theta|$. These side lobes lead the beamformer to capture the white noise which spans the whole space, so that the performance of beamformer degrades.

4.2. Performance for Two Types of Interference. In the presence of several types of interference, the performance analysis of the proposed QSWL GSC is not so easy in theory. Thus, we employ two experiments to illustrate the performance of the proposed QSWL GSC. We assume that there are two types of interference with identical input INR, $\phi_d = \phi_{u1} = \phi_{u2} = 60^\circ$, $\alpha_d = \alpha_{u1} = \alpha_{u2} = 30^\circ$, $\beta_d = \beta_{u1} = \beta_{u2} = 30^\circ$, and $\theta_c = 3^\circ$.

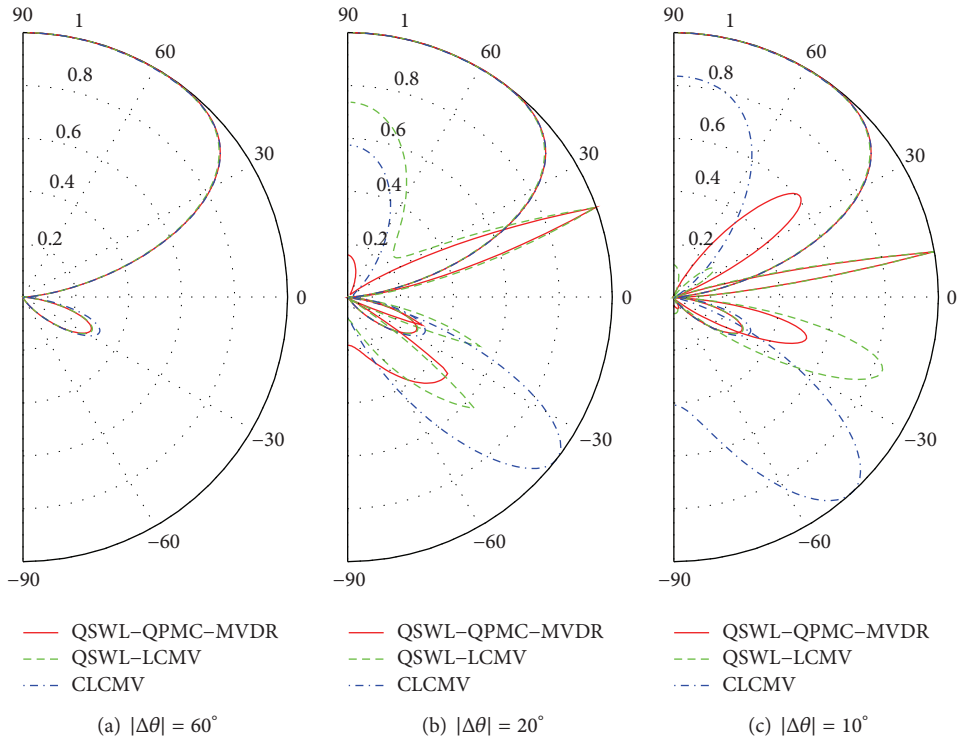


FIGURE 4: The power patterns at $\theta_u = 0^\circ$ and $\theta_d = |\Delta\theta|$.

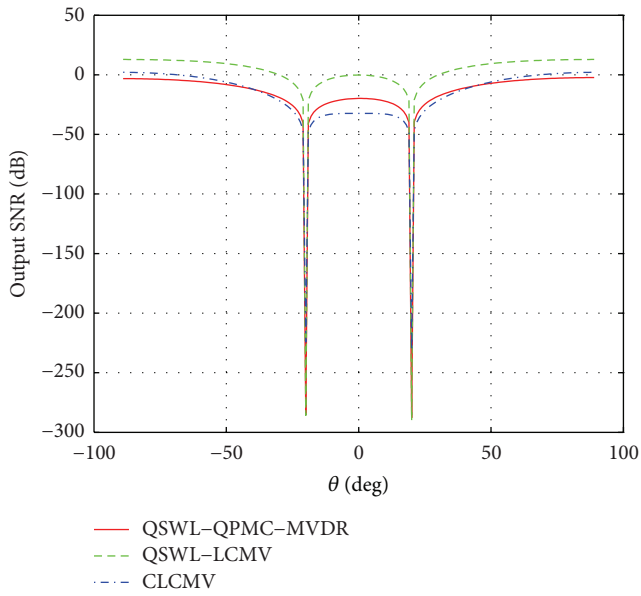


FIGURE 5: Effect of the angular mismatch between the distortionless constraint direction and the real arrival direction of the desired signal at $M = 3$.

The covariance matrix $\mathbf{R}_x$ is used. In the first experiment, we investigate the effect of DOA mismatch, where θ is used to denote the real presumed DOA of the desired signal. Figures 5 and 6 display, respectively, the output SINR as a function of θ at $M = 3$ and $M = 6$, where $\theta_d = 0^\circ$, $\theta_{u1} = -20^\circ$, and $\theta_{u2} = 20^\circ$. From Figures 5 and 6, it is seen that the QSWL-LCMV beamformer outperforms obviously both the QSWL-QPMC-MVDR and the CLCMV. In the case of no DOA mismatch and $M = 6$, the output SINR of the QSWL-LCMV is 25 dB more than that of the CLCMV, but the output SINR of the CLCMV is larger than that of the QSWL-QPMC-MVDR. When the array's element number reduces, the performance of the CLCMV degrades. At $M = 3$, the output SINR of the QSWL-QPMC-MVDR are 12 dB more than that of the CLCMV. The cause of this behavior is that the QSWL GSC has larger degrees of freedom than the CLCMV. Thus, the QSWL GSC has the better performance when many constraints are simultaneously imposed and/or the array's element number is reduced. In addition, when the DOA mismatch is little, such as $\theta \leq 10^\circ$, the output SINR of three beamformers do not obviously degrade.

In the second experiment, we investigate the effect of snapshots n, where the covariance matrix $\mathbf{R}_x$ is replaced by the sample covariance matrix $\widehat{\mathbf{R}}_x$. Figure 7 displays the output SINR as a function of n at $M = 6$, where $\theta_d = 0^\circ$, $\theta_{u1} = -30^\circ$, and $\theta_{u2} = 30^\circ$. From Figure 7, it is seen that the QSWL GSC has slower convergence than the CLCMV. But the output SINR of the QSWL-LCMV are obviously more than that of the CLCMV.

5. Conclusion

In this paper, we propose a quaternion *semiwidely* linear beamformer and its useful implementation based on a

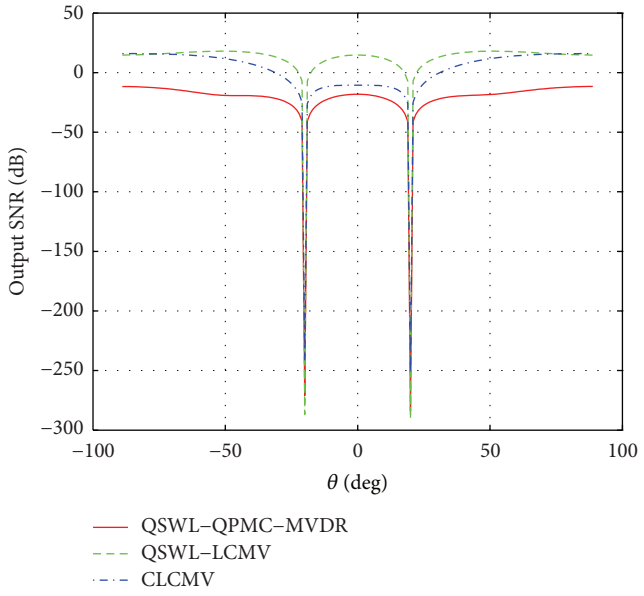


FIGURE 6: Effect of the angular mismatch between the distortionless constraint direction and the real direction of the desired signal at $M = 6$.

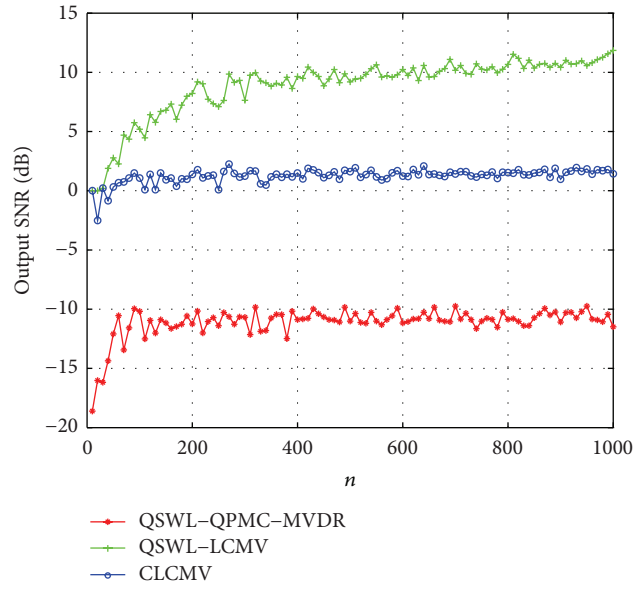


FIGURE 7: Effect of the number of snapshots at $M = 6$.

quaternion model of linear symmetric array with $2M$ two-component EM vector sensors. Since the QSWL GSC consists of two-stage beamformers, it has more information than the complex "long vector" beamformer. The increase of information results in the improvement of the beamformer's performance. By designing the weight vectors of two-stage beamformers, the interference is completely canceled in the output of QSWL GSC and the desired signal is not distorted. Simulation results reveal that the proposed QSWL GSC has the better performance in small angular separation and the robustness to DOA mismatch.

Appendices

A. Derivation for (19)

If the interference and additive noise are uncorrelated, we have

$$\mathbf{R}_{in} = \sigma_u^2 \overline{\mathbf{V}}_u \overline{\mathbf{V}}_u^H + \sigma_n^2 \mathbf{I}, \tag{A.1}$$

where σ_u^2 and σ_n^2 are the covariance of interference and additive noise, respectively. By making use of the matrix inversion lemma (see Appendix in [30]), we find

$$\mathbf{R}_{in}^{-1} = \frac{1}{\sigma_n^2}\left(\mathbf{I} - \frac{\overline{\mathbf{V}}_u \overline{\mathbf{V}}_u^H}{\xi_u^{-1} + \overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_u}\right), \tag{A.2}$$

where $\xi_u = \sigma_u^2/\sigma_n^2$. Plugging (A.2) into (18), we obtain

$$\overline{\mathbf{W}} = \frac{\overline{\mathbf{V}}_d - \left(\overline{\mathbf{V}}_u \overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_d / \left(\xi_u^{-1} + \overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_u\right)\right)}{\overline{\mathbf{V}}_d^H \overline{\mathbf{V}}_d - \left(\overline{\mathbf{V}}_d^H \overline{\mathbf{V}}_u \overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_d / \left(\xi_u^{-1} + \overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_u\right)\right)}. \tag{A.3}$$

Plugging $\overline{\mathbf{V}}_k^H \overline{\mathbf{V}}_k = 2M|P_k|^2$ $(k = d, u)$ and $\overline{\mathbf{V}}_u^H \overline{\mathbf{V}}_d = [P_u^{\triangleleft} P_d]_1 \mathbf{q}_u^H \mathbf{q}_d$ into (A.3), we can obtain (19) after some manipulation.

B. Derivation for (30)

If the interference and additive noise are uncorrelated, we have

$$\mathbf{R}_{QS} = \sigma_u^2 \widetilde{\mathbf{V}}_u \widetilde{\mathbf{V}}_u^H + \sigma_n^2 |P_d|^2 \mathbf{I}. \tag{B.1}$$

By making use of the matrix inversion lemma, we find

$$\mathbf{R}_{QS}^{-1} = \frac{1}{\sigma_n^2 |P_d|^2}\left(\mathbf{I} - \frac{\widetilde{\mathbf{V}}_u \widetilde{\mathbf{V}}_u^H}{\xi_u^{-1}|P_d|^2 + \widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u}\right). \tag{B.2}$$

Plugging (B.2) into (29), we obtain

$$\mathbf{w}_{MV} = \frac{\widetilde{\mathbf{V}}_u - \left(\widetilde{\mathbf{V}}_u \widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u / \left(\xi_u^{-1}|P_d|^2 + \widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u\right)\right)}{\widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u - \left(\widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u \widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u / \left(\xi_u^{-1}|P_d|^2 + \widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u\right)\right)} \overline{\mathbf{V}}_u^H \overline{\mathbf{W}}. \tag{B.3}$$

Plugging $\widetilde{\mathbf{V}}_u^H \widetilde{\mathbf{V}}_u = \kappa$ and $\overline{\mathbf{V}}_u^H \overline{\mathbf{W}} = g_1$ into (B.3), we can obtain (30) after some manipulation.

C. Derivation for (40)

Since $\mathbf{C} = [\overline{\mathbf{V}}_u^{(i)}, \overline{\mathbf{V}}_d^{(i)}]$, the inverse of 2×2 matrix $\mathbf{C}^H \mathbf{R}_{inG}^{-1} \mathbf{C}$ is given by

$$\left(\mathbf{C}^H \mathbf{R}_{inG}^{-1} \mathbf{C}\right)^{-1} = \frac{1}{\tau}\begin{bmatrix} \overline{\mathbf{V}}_d^{(i)H} \mathbf{R}_{inG}^{-1} \overline{\mathbf{V}}_d^{(i)} & -\overline{\mathbf{V}}_u^{(i)H} \mathbf{R}_{inG}^{-1} \overline{\mathbf{V}}_d^{(i)} \\ -\overline{\mathbf{V}}_d^{(i)H} \mathbf{R}_{inG}^{-1} \overline{\mathbf{V}}_u^{(i)} & \overline{\mathbf{V}}_u^{(i)H} \mathbf{R}_{inG}^{-1} \overline{\mathbf{V}}_u^{(i)} \end{bmatrix}, \tag{C.1}$$

where

$$\tau = \overline{\mathbf{V}}_u^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_u^{(i)}\overline{\mathbf{V}}_d^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_d^{(i)} - \overline{\mathbf{V}}_u^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_d^{(i)}\overline{\mathbf{V}}_d^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_u^{(i)} \tag{C.2}$$

is the determinant of 2×2 matrix $\mathbf{C}^H\mathbf{R}_{inG}^{-1}\mathbf{C}$. Plugging (C.1) into (39) and after some manipulation, we obtain

$$\overline{\mathbf{G}} = \frac{g_1}{\tau}\left(\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_u^{(i)}\overline{\mathbf{V}}_d^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_d^{(i)}\right) - \frac{g_1}{\tau}\left(\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_d^{(i)}\overline{\mathbf{V}}_d^{(i)H}\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_u^{(i)}\right). \tag{C.3}$$

By making use of the matrix inversion lemma, we find

$$\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_u^{(i)} = \frac{1}{\sigma_n^2}\frac{\xi_u^{-1}\overline{\mathbf{V}}_u^{(i)}}{\xi_u^{-1} + \overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_u^{(i)}},$$
$$\mathbf{R}_{inG}^{-1}\overline{\mathbf{V}}_d^{(i)} = \frac{1}{\sigma_n^2}\left(\overline{\mathbf{V}}_d^{(i)} - \frac{\overline{\mathbf{V}}_u^{(i)}\overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_d^{(i)}}{\xi_u^{-1} + \overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_u^{(i)}}\right). \tag{C.4}$$

Plugging (C.4) into (C.3) and using the results $\overline{\mathbf{V}}_k^{(i)H}\overline{\mathbf{V}}_k^{(i)} = 2M|P_k|^2$ $(k = d, u)$; $\overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_d^{(i)} = [P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d$; and $\overline{\mathbf{V}}_d^{(i)H}\overline{\mathbf{V}}_u^{(i)} = (\overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_d^{(i)})^H = [P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u$, we can obtain (40) after some manipulation.

D. Derivation for (47)

Using (19), we can easily obtain

$$\overline{\mathbf{W}}^H\overline{\mathbf{W}} = \frac{2M\varepsilon^2|P_d|^2}{\mu^2} - \frac{\left(2\xi_u^{-1} + 2M|P_u|^2\right)|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2}. \tag{D.1}$$

Using the results $\mathbf{w}_{QS}^{\triangleleft}\mathbf{w}_{QS} = |P_d|^2$ and $\mathbf{w}_{MV}^H\mathbf{w}_{MV} = |g_1|^2/\kappa$, we can easily obtain

$$\overline{\mathbf{G}}^H\overline{\mathbf{G}} = \mathbf{G}^{\triangleleft}\mathbf{G} = \left(\mathbf{w}_{QS}\mathbf{w}_{MV}\right)^{\triangleleft}\left(\mathbf{w}_{QS}\mathbf{w}_{MV}\right) = \frac{|P_d|^2\xi_u^{-2}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2\kappa}. \tag{D.2}$$

Using (19) and (26), we can easily obtain

$$\mathbf{W}_1^H\mathbf{G}_1 = \frac{\varepsilon\mathbf{V}_{d1}^H\,\mathrm{diag}\left(\mathbf{q}_d\right)\mathbf{w}_{MV}a_{d2}^*}{\mu} - \frac{\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H\mathbf{V}_{u1}^H\,\mathrm{diag}\left(\mathbf{q}_d\right)\mathbf{w}_{MV}a_{d2}^*}{\mu}$$
$$= \frac{\varepsilon a_{d1}^*a_{d2}^*}{\mu}\mathbf{q}_d^H\,\mathrm{diag}\left(\mathbf{q}_d\right)\mathbf{w}_{MV} - \frac{\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H a_{u1}^*a_{d2}^*}{\mu}\mathbf{q}_u^H\,\mathrm{diag}\left(\mathbf{q}_d\right)\mathbf{w}_{MV}, \tag{D.3}$$

$$\mathbf{W}_2^H\mathbf{G}_2 = \frac{\varepsilon\mathbf{V}_{d2}^H\,\mathrm{diag}\left(\mathbf{q}_d^*\right)\mathbf{w}_{MV}a_{d1}^*}{\mu} - \frac{\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H\mathbf{V}_{u2}^H\,\mathrm{diag}\left(\mathbf{q}_d^*\right)\mathbf{w}_{MV}a_{d1}^*}{\mu}$$
$$= \frac{\varepsilon a_{d1}^*a_{d2}^*}{\mu}\left(\mathbf{q}_d^H\,\mathrm{diag}(\mathbf{q}_d)\right)^*\mathbf{w}_{MV} - \frac{\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H a_{u2}^*a_{d1}^*}{\mu}\left(\mathbf{q}_u^H\,\mathrm{diag}\left(\mathbf{q}_d\right)\right)^*\mathbf{w}_{MV}. \tag{D.4}$$

Using $\mathbf{q}_d^H\,\mathrm{diag}(\mathbf{q}_d) = (\mathbf{q}_d^H\,\mathrm{diag}(\mathbf{q}_d))^*$ and $\mathbf{q}_{ud} = \mathbf{q}_u^H\,\mathrm{diag}(\mathbf{q}_d)$, we can obtain

$$\mathbf{W}_1^H\mathbf{G}_1 - \mathbf{W}_2^H\mathbf{G}_2 = \frac{-\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H}{\mu} \times \left(a_{u1}^*a_{d2}^*\mathbf{q}_{ud} - a_{u2}^*a_{d1}^*\mathbf{q}_{ud}^*\right)\mathbf{w}_{MV}. \tag{D.5}$$

Using (30) and (31), we can easily obtain $(a_{u1}^*a_{d2}^*\mathbf{q}_{ud} - a_{u2}^*a_{d1}^*\mathbf{q}_{ud}^*)\mathbf{w}_{MV} = g_1$. Thus, we have

$$\mathbf{W}_1^H\mathbf{G}_1 - \mathbf{W}_2^H\mathbf{G}_2 = -\frac{\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H g_1}{\mu} = -\frac{\xi_u^{-1}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2},$$
$$\mathbf{G}_1^H\mathbf{W}_1 - \mathbf{G}_2^H\mathbf{W}_2 = \left(\mathbf{W}_1^H\mathbf{G}_1 - \mathbf{W}_2^H\mathbf{G}_2\right)^H = -\frac{\xi_u^{-1}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2}. \tag{D.6}$$

Plugging (D.1), (D.2), and (D.6) into (46), we can obtain (47) after some manipulation.

E. Derivation for (50)

Using the results $\overline{\mathbf{V}}_k^{(i)H}\overline{\mathbf{V}}_k^{(i)} = 2M|P_k|^2$ $(k = d, u)$; $\overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_d^{(i)} = [P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d$; and $\overline{\mathbf{V}}_d^{(i)H}\overline{\mathbf{V}}_u^{(i)} = (\overline{\mathbf{V}}_u^{(i)H}\overline{\mathbf{V}}_d^{(i)})^H$, we have

$$\begin{aligned}2M|P_d|^2\nu = &\left(2M|P_d|^2\overline{\mathbf{V}}_u^{(i)} - \left[P_d^{\triangleleft}P_u\right]_1\mathbf{q}_d^H\mathbf{q}_u\overline{\mathbf{V}}_d^{(i)}\right)^H\\ &\times\left(2M|P_d|^2\overline{\mathbf{V}}_u^{(i)} - \left[P_d^{\triangleleft}P_u\right]_1\mathbf{q}_d^H\mathbf{q}_u\overline{\mathbf{V}}_d^{(i)}\right).\end{aligned} \tag{E.1}$$

Using (40), we can easily obtain

$$\begin{aligned}\overline{\mathbf{G}}^H\overline{\mathbf{G}} &= 2M|P_d|^2\frac{|g_1|^2}{\nu}\\ &= \frac{2M|P_d|^2\xi_u^{-2}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2\nu}.\end{aligned} \tag{E.2}$$

Using (19) and (40), we can easily obtain

$$\begin{aligned}\mathbf{W}_1^H\mathbf{G}_1 = &\frac{g_1\varepsilon 2M|P_d|^2\mathbf{V}_{d1}^H\mathbf{V}_{u1}}{\mu\nu}\\ &-\frac{g_1\varepsilon[P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u\mathbf{V}_{d1}^H\mathbf{V}_{d1}}{\mu\nu}\\ &-\frac{g_1\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H2M|P_d|^2\mathbf{V}_{u1}^H\mathbf{V}_{u1}}{\mu\nu}\\ &+\frac{g_1\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H[P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u\mathbf{V}_{u1}^H\mathbf{V}_{d1}}{\mu\nu},\end{aligned} \tag{E.3}$$

$$\begin{aligned}\mathbf{W}_2^H\mathbf{G}_2 = &-\frac{g_1\varepsilon 2M|P_d|^2\mathbf{V}_{d2}^H\mathbf{V}_{u2}}{\mu\nu}\\ &+\frac{g_1\varepsilon[P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u\mathbf{V}_{d2}^H\mathbf{V}_{d2}}{\mu\nu}\\ &+\frac{g_1\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H2M|P_d|^2\mathbf{V}_{u2}^H\mathbf{V}_{u2}}{\mu\nu}\\ &-\frac{g_1\left([P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d\right)^H[P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u\mathbf{V}_{u2}^H\mathbf{V}_{d2}}{\mu\nu}.\end{aligned} \tag{E.4}$$

Using the results $\mathbf{V}_{k1}^H\mathbf{V}_{k1} + \mathbf{V}_{k2}^H\mathbf{V}_{k2} = 2M|P_k|^2$ $(k = d, u)$; $\mathbf{V}_{d1}^H\mathbf{V}_{u1} + \mathbf{V}_{d2}^H\mathbf{V}_{u2} = [P_d^{\triangleleft}P_u]_1\mathbf{q}_d^H\mathbf{q}_u$; and $\mathbf{V}_{u1}^H\mathbf{V}_{d1} + \mathbf{V}_{u2}^H\mathbf{V}_{d2} = [P_u^{\triangleleft}P_d]_1\mathbf{q}_u^H\mathbf{q}_d$, we have

$$\begin{aligned}\mathbf{W}_1^H\mathbf{G}_1 - \mathbf{W}_2^H\mathbf{G}_2 &= -\frac{\xi_u^{-1}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2},\\ \mathbf{G}_1^H\mathbf{W}_1 - \mathbf{G}_2^H\mathbf{W}_2 &= \left(\mathbf{W}_1^H\mathbf{G}_1 - \mathbf{W}_2^H\mathbf{G}_2\right)^H\\ &= -\frac{\xi_u^{-1}|[P_u^{\triangleleft}P_d]_1|^2|\mathbf{q}_u^H\mathbf{q}_d|^2}{\mu^2}.\end{aligned} \tag{E.5}$$

Plugging (D.1), (E.2), and (E.5) into (46), we can obtain (50) after some manipulation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants 60872088 and 61172126.

Endnotes

1. According to the property of quaternion shown in [24], we have $jc = c^*j$ for any complex number c. Thus, $jx_{-m2}(t) = q_m(\theta,\phi)ja_2(\theta,\phi,\alpha,\beta)s(t)$. Since the spatial phase factor $q_m(\theta,\phi)$ in $x_{m1}(t)$ is identical with that in $jx_{-m2}(t)$, we use $x_{-m2}(t)$ instead of $x_{m2}(t)$ to construct the quaternion-valued signal $x_m(t)$.
2. According to the definition in [11], $\mathbf{x}$ is C^i-proper if it can be represented by means of two jointly proper complex vectors in the plane spanned by $\{1, i\}$.
3. According to the definition in [11], the augmented covariance matrix of a C^i-proper quaternion vector $\mathbf{x}$ can be written as

$$\mathbf{R}_{\overline{\mathbf{x}},\overline{\mathbf{x}}} = E\left\{\overline{\mathbf{x}}\,\overline{\mathbf{x}}^{\triangleleft}\right\} = \begin{bmatrix}\mathbf{R}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{x}}} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}^{(j)}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{x}}}\end{bmatrix}, \tag{$*$}$$

where $\overline{\mathbf{x}} = [\mathbf{x}^T, \mathbf{x}^{(i)T}, \mathbf{x}^{(j)T}, \mathbf{x}^{(k)T}]$ is the augmented quaternion vector; $\widetilde{\mathbf{x}} = [\mathbf{x}^T, \mathbf{x}^{(i)T}]$ is the semiaugmented quaternion vector; and $\mathbf{R}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{x}}} = E\{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}^{\triangleleft}\}$ is the semiaugmented covariance matrix of quaternion vector $\mathbf{x}$. In comparison with the semiaugmented covariance matrix $\mathbf{R}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{x}}}$, the augmented covariance matrix $\mathbf{R}_{\overline{\mathbf{x}},\overline{\mathbf{x}}}$ has no more extra information. In other words, the *full widely* linear processing is equivalent to the *semiwidely* linear processing in handling the C^i-proper quaternion vector. We should not expect that the performance is improved by replacing *semiwidely* linear processing with *full widely* linear processing.

References

[1] T. Bülow and G. Sommer, "Hypercomplex signals—a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2844–2852, 2001.

[2] S. Miron, N. le Bihan, and J. I. Mars, "Quaternion-MUSIC for vector-sensor array processing," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1218–1229, 2006.

[3] N. le Bihan, S. Miron, and J. I. Mars, "MUSIC algorithm for vector-sensors array using biquaternions," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4523–4533, 2007.

[4] X. Gong, Y. Xu, and Z. Liu, "Quaternion ESPRIT for direction finding with a polarization sentive array," in *Proceedings of the*

9th International Conference on Signal Processing (ICSP '08), pp. 378–381, October 2008.

[5] T. A. Ell and S. J. Sangwine, "Hypercomplex Fourier transforms of color images," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 22–35, 2007.

[6] J. Seberry, K. Finlayson, S. S. Adams, T. A. Wysocki, T. Xia, and B. J. Wysocki, "The theory of quaternion orthogonal designs," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 256–265, 2008.

[7] D. Choukroun, I. Y. Bar-Itzhack, and Y. Oshman, "Novel quaternion Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 174–190, 2006.

[8] C. C. Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex processes," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1316–1327, 2009.

[9] C. C. Took and D. P. Mandic, "Quaternion-valued stochastic gradient-based adaptive IIR filtering," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3895–3901, 2010.

[10] J. Vía, D. P. Palomar, L. Vielva, and I. Santamaría, "Quaternion ICA from second-order statistics," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1586–1600, 2011.

[11] J. Vía, D. Ramírez, and I. Santamaría, "Properness and widely linear processing of quaternion random vectors," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3502–3515, 2010.

[12] C. C. Took and D. P. Mandic, "Augmented second-order statistics of quaternion random signals," *Signal Processing*, vol. 91, no. 2, pp. 214–224, 2011.

[13] C. Cheong Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4427–4431, 2010.

[14] M. Wang, C. C. Took, and D. P. Mandic, "A class of fast quaternion valued variable stepsize stochastic gradient learning algorithms for vector sensor processes," in *Proceedings of the International Joint Conference on Neural Network (IJCNN '11)*, pp. 2783–2786, August 2011.

[15] F. G. A. Neto and V. H. Nascimento, "A novel reduced-complexity widely linear QLMS algorithm," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP '11)*, pp. 81–84, June 2011.

[16] B. C. Ujang, C. C. Took, and D. P. D. Mandic, "Quaternion-valued nonlinear adaptive filtering," *IEEE Transactions on Neural Networks*, vol. 22, no. 8, pp. 1193–1206, 2011.

[17] S. Javidi, C. Cheong Took, C. Jahanchahi, N. le Bihan, and D. P. Mandic, "Blind extraction of improper quaternion sources," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 3708–3711, May 2011.

[18] J. Tao and W. Chang, "Quaternion MMSE algorithm and its application in beamforming," *Acta Aeronautica et Astronautica Sinica*, vol. 32, no. 4, pp. 729–738, 2011.

[19] X. Gou, Y. Xu, Z. Liu, and X. Gong, "Quaternion-capon beamformer using crossed-dipole arrays," in *Proceedings of the 4th IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE '11)*, pp. 34–37, November 2011.

[20] J. W. Tao and W. X. Chang, "The MVDR beamformer based on hypercomplex processes," in *Proceedings of the 2012 IEEE International of Conference on Computer Science and Electronic Engineering*, pp. 273–277, March 2012.

[21] P. Chevalier and A. Blin, "Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5323–5336, 2007.

[22] P. Chevalier, J. P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3573–3576, May 2009.

[23] J. Benesty, J. Chen, and Y. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 469–472, 2010.

[24] J. P. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*, Kluwer, Boston, Mass, USA, 1997.

[25] J. Li, P. Stoica, and D. Zheng, "Efficient direction and polarization estimation with a COLD array," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 4, pp. 539–547, 1996.

[26] A. Manikas and J. W. P. Ng, "Crossed-dipole arrays for asynchronous DS-CDMA systems," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 1, pp. 122–131, 2004.

[27] H. S. Mir and J. D. Sahr, "Passive direction finding using airborne vector sensors in the presence of manifold perturbations," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 156–164, 2007.

[28] L. Liang, T. Jian-Wu, and C. Wei, "DOA estimation based on sparse and nonuniform COLD array," in *Proceedings of the WRI International Conference on Communications and Mobile Computing (CMC '09)*, pp. 391–395, January 2009.

[29] J. W. Tao, L. Liu, and Z. Y. Lin, "Joint DOA, range, and polarization estimation in the fresnel region," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2657–2672, 2011.

[30] H. van Trees, *Optimum Array Processing (Detection Estimation, and Modulation Theory)*, part 4, Wiley-Interscience, New York, NY, USA, 2002.

[31] A. Nehorai, K. C. Ho, and B. T. G. Tan, "Minimum-noise-variance beamformer with an electromagnetic vector sensor," *IEEE Transactions on Signal Processing*, vol. 47, no. 3, pp. 601–618, 1999.

[32] K. T. Wong, "Blind beamforming geolocation for wideband-FFHs with unknown hop-sequences," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 65–76, 2001.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 608158, 5 pages
http://dx.doi.org/10.1155/2014/608158

Research Article

Generalized Discriminant Orthogonal Nonnegative Tensor Factorization for Facial Expression Recognition

Zhang XiuJun[1,2] and Liu Chang[1,2]

[1] *College of Information Science and Technology, Chengdu University, Chengdu 610106, China*
[2] *Key Laboratory of Pattern Recognition and Intelligent Information Processing in Sichuan, Chengdu 610106, China*

Correspondence should be addressed to Liu Chang; chang.liu.scu@gmail.com

Received 4 August 2013; Accepted 6 January 2014; Published 26 March 2014

Academic Editors: S. Bourennane and J. Marot

In order to overcome the limitation of traditional nonnegative factorization algorithms, the paper presents a generalized discriminant orthogonal non-negative tensor factorization algorithm. At first, the algorithm takes the orthogonal constraint into account to ensure the nonnegativity of the low-dimensional features. Furthermore, the discriminant constraint is imposed on low-dimensional weights to strengthen the discriminant capability of the low-dimensional features. The experiments on facial expression recognition have demonstrated that the algorithm is superior to other non-negative factorization algorithms.

1. Introduction

Over the past few years, the nonnegative matrix factorization algorithm (NMF) [1] and its variants have proven to be useful for several problems, especially in facial image characterization and representation problems [2–8]. The idea of nonnegative factorization is partly motivated by the biological fact that the firing rates in visual perception neurons are nonnegative.

However, NMF and its variants have some drawbacks. First of all, NMF requires that all object images should be vectorized in order to find the non-negative decomposition. This vectorization leads to information loss, since the local structure of the image is lost. Moreover, NMF is not unique [9, 10]. In order to remedy these drawbacks, non-negative tensor factorization (NTF) has been proposed [11–13]. NTF represents a facial expression database as a three-order tensor. The tensor representation avoids the vectorization operation and preserves the structure of the data. Under some mild conditions, NTF is unique. Existing NMF and NTF algorithms project data into low-dimensional space with the inverse or pseudoinverse of the basis images, so both of them cannot guarantee the nonnegativity of low-dimensional features, which restricts the application of non-negative factorization in real world. Furthermore, NTF do not take into account class information in data samples. Actually, it is believed that those features with discriminant constraints are of great importance for pattern recognition. Reference [14] develops a discriminant non-negative tensor factorization algorithm (DNTF), which adds fisher discriminant constraint into the objective function. But like other discriminant non-negative matrix factorizations [6, 15–18], DNTF employed discriminant analysis on the representation coefficients and not on the actual features used in the recognition procedure. The actual features used for recognition are derived from the projection of data samples to the bases matrix and only implicitly depend on the representation coefficients.

Based on the above analysis, the paper proposes a generalized discriminant orthogonal non-negative tensor factorization algorithm (GDONTF), which makes full use of the class information and imposes the orthogonal constraint to the objective function. The algorithm not only guarantees the non-negativity of low-dimensional features, but also generalizes discriminant constraints to low-dimension features. The experiments on facial expression recognition indicate that GDONTF achieves better performance than other non-negative factorization algorithms.

2. Generalized Discriminant Orthogonal Non-Negative Tensor Factorization

Consider an N order tensor $X \in \mathbb{R}^{d_1 \times d_2 \cdots \times d_N}$, every data sample X_i is an $n-1$ order tensor; that is, $X_i \in \mathbb{R}^{d_1 \times d_2 \cdots \times d_{n-1}}$, in which $d_1, d_2 \cdots d_{N-1}$, is the dimensionality and d_N is the number of data set. The data set is divided into C classes. Data samples belonging to class c denote $V(c)$; the number of data samples in $V(c)$ is N_c. In order to guarantee the non-negativity of low-dimensional features and take use of the class information, we propose generalized discriminant orthogonal non-negative tensor factorization algorithm; the objective function of which is defined as follow:

$$O = \left\| X - \sum_{r=1}^{R} U_{:,r}^{(1)} \circ U_{:,r}^{(2)} \cdots \circ U_{:,r}^{(N)} \right\|^2 + \alpha tr\left[S_w\right] - \beta tr\left[S_b\right] \quad \text{s.t.} \quad U^{(n)} U^{(n)T} = I \quad n = 1, 2, \ldots, N-1. \tag{1}$$

In which, $\alpha \geq 0$, $\beta \geq 0$, $U_{ij}^{(n)} \geq 0$, $n = 1, 2, \ldots, N$, I is the identity matrix and S_w and S_b are the within- and between-class scatter matrices of the low-dimensional features, respectively. Because $U^{(n)} U^{(n)T} = I$, low-dimensional features can be computed as follows:

$$h_i = \left(U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(2)} \odot U^{(1)}\right)^T \overline{X_i} = W^T \overline{X_i}, \tag{2}$$

where the basis matrix $W = U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(2)} \odot U^{(1)}$. Let h_i be the low-dimensional features of the sample X_i; then the feature matrix $H \in \mathbb{R}^{d \times M}$ consists of all low-dimensional features, d is the low dimensionality of samples, and M is the number of all samples. Actually, the separability of the weight coefficient has nothing to do with the recognition accuracy, while the class separability of the low-dimensional features has a great influence on the recognition accuracy. Consequently, the within- and between-class scatter matrices are defined as follows:

$$\begin{aligned} S_w &= \sum_{c=1}^{C} \sum_{u_i \in V(c)} \left(h_i - m_c\right)\left(h_i - m_c\right)^T \\ &= \sum_{c=1}^{C} \sum_{u_i \in V(c)} \left(W^T \overline{X_i} - m_c\right)\left(W^T \overline{X_i} - m_c\right)^T, \\ S_b &= \sum_{c=1}^{C} N_c \left(m_c - m\right)\left(m_c - m\right)^T, \end{aligned} \tag{3}$$

where m_c is the mean of the low-dimensional features in the class c and m is the mean of all low-dimensional features. The objective function in (1) can be written as the following optimization problem:

$$\begin{aligned} \min_{U^{(n)}} \quad & \left\| X - \sum_{r=1}^{R} U_{:,r}^{(1)} \circ U_{:,r}^{(2)} \cdots \circ U_{:,r}^{(N)} \right\|^2 + \alpha tr\left[S_w\right] - \beta tr\left[S_b\right] \\ \text{s.t.} \quad & U_{ij}^{(n)} \geq 0 \\ & U^{(n)T} U^{(n)} = I, \quad n = 1, 2, \ldots, N-1. \end{aligned} \tag{4}$$

Since the basis matrix W consists of the projection matrices $U^{(n)}$, $n = 1, 2, \ldots, N-1$, we solve the projection matrices $U^{(n)}$, $n = 1, 2, \ldots, N-1$, and the weight matrix $U^{(N)}$, respectively, to deal with the optimization problem (4). First of all, we formulate the Lagrange multipliers out of the constrained optimization problem in (4):

$$\begin{aligned} f\left(U^{(n)}, \lambda\right) = & \left\| X_{(n)} - U^{(n)} G^T \right\|^2 + \alpha tr\left[S_w\right] - \beta tr\left[S_b\right] \\ & + Tr\left[\lambda\left(U^{(n)T} U^{(n)} - I\right)\right], \end{aligned} \tag{5}$$

where $G = U^{(N)} \odot \cdots \odot U^{(n+1)} \odot U^{(1)} \odot U^{(2)} \odot \cdots \odot U^{(n-1)}$.

Take the derivative of $f(U^{(n)}, \lambda)$ with respect to $U^{(n)}$ and λ, $n = 1, 2, \ldots, N-1$; we have

$$\begin{aligned} \frac{\partial f}{\partial U^{(n)}} &= -2X_{(n)}G + 2U^{(n)}G^TG + 2U^{(n)}\lambda + \alpha\frac{\partial Tr\left(S_w\right)}{\partial U^{(n)}} \\ &\quad - \beta\frac{\partial Tr\left(S_b\right)}{\partial U^{(n)}} \\ &= -2X_{(n)}G + 2U^{(n)}G^TG + 2U^{(n)}\lambda + \alpha\nabla Tr\left(S_w\right) \\ &\quad - \beta\nabla Tr\left(S_b\right), \end{aligned} \tag{6}$$

$$\frac{\partial f}{\partial \lambda} = U^{(n)T} U^{(n)} - I. \tag{7}$$

Set (6) and (7) to zeros; we get

$$\begin{aligned} & -2X_{(n)}G + 2U^{(n)}G^TG + 2U^{(n)}\lambda + \alpha\nabla Tr\left(S_w\right) \\ & \quad - \beta\nabla Tr\left(S_b\right) = 0, \end{aligned} \tag{8}$$

$$U^{(n)T} U^{(n)} = I. \tag{9}$$

Left multiply both side of (8) by $U^{(n)T}$; we immediately have

$$\begin{aligned} & -2U^{(n)T}X_{(n)}G + 2U^{(n)T}U^{(n)}G^TG + 2U^{(n)T}U^{(n)}\lambda \\ & \quad + \alpha U^{(n)T}\nabla Tr\left(S_w\right) - \beta U^{(n)T}\nabla Tr\left(S_b\right) = 0. \end{aligned} \tag{10}$$

Therefore, the update rule for $U^{(n)}$ is

$$U^{(n)^{t+1}} = U^{(n)^t} * \frac{2X_{(n)}G + \beta\nabla Tr\left(S_b\right)}{2U^{(n)}G^TG + 2U^{(n)}\lambda + \alpha\nabla Tr\left(S_w\right)}. \tag{11}$$

The gradient $[\nabla Tr(S_b)]_{i,j} = \partial Tr[S_w]/\partial u_{ij}^{(n)}$ is given by

$$\begin{aligned} \frac{\partial Tr\left[S_w\right]}{\partial u_{ij}^{(n)}} &= \frac{\partial \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(h_p - m_c\right)\left(h_p - m_c\right)^T}{\partial u_{ij}^{(n)}} \\ &= \frac{\partial \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(W^T\overline{X_p} - m_c\right)\left(W^T\overline{X_p} - m_c\right)^T}{\partial u_{ij}^{(n)}}. \end{aligned} \tag{12}$$

Because $W = (U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(2)} \odot U^{(1)})$, we can get

$$\begin{aligned}
&\frac{\partial Tr\left[S_w\right]}{\partial u_{ij}^{(n)}} \\
&= \left(\partial \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(\left(U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(2)} \odot U^{(1)} \right)^T \right. \right. \\
&\qquad \left. \left. \times \overline{X_p} - m_c \right)^2 \right) \times \left(\partial u_{ij}^{(n)} \right)^{-1} \\
&= 2 \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(W^T \overline{X_p} - m_c \right) \\
&\quad \times \left(\partial \left(\left(U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(1)} \right)^T \right. \right. \\
&\qquad \left. \left. \times \left(\overline{X_p} - \sum_{\substack{q=1 \\ h_q \in V(c)}}^{N_c} \overline{X_q} \right) \right) \right) \times \left(\partial u_{ij}^{(n)} \right)^{-1}.
\end{aligned} \tag{13}$$

Let $U^{pn} = U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(n+1)}$ and $U^{an} = U^{(n-1)} \odot U^{(n-2)} \odot \cdots \odot U^{(1)}$; we have

$$\begin{aligned}
&\frac{\partial Tr\left[S_w\right]}{\partial u_{ij}^{(n)}} \\
&= 2 \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(W^T \overline{X_p} - m_c \right) \\
&\quad \times \left(\partial \left(\left(U^{pn} \odot U^{(n)} \odot U^{an} \right)^T \right. \right. \\
&\qquad \left. \left. \times \left(\overline{X_p} - \sum_{\substack{q=1 \\ h_q \in V(c)}}^{N_c} \overline{X_q} \right) \right) \right) \times \left(\partial u_{ij}^{(n)} \right)^{-1};
\end{aligned} \tag{14}$$

Since

$$\begin{aligned}
&U^{pn} \odot U^{(n)} \odot U^{an} \\
&\quad = \left[U_1^{pn} \otimes U_1^{(n)} \otimes U_1^{an}, \ldots, U_j^{pn} \otimes U_j^{(n)} \otimes U_j^{an}, \ldots \right] \\
&\quad = \left[U_{:,1}^{pn} \otimes U_{:,1}^{(n)} \otimes U_{:,1}^{an}, \ldots, U_{:,j}^{pn} \otimes U_{:,j}^{(n)} \otimes U_{:,j}^{an}, \ldots \right] \\
&\quad = \left[u_{11}^{pn} u_{11}^{(n)} u_{11}^{an}, \ldots, u_{1j}^{pn} u_{1j}^{(n)} u_{1j}^{an}, \ldots, \right. \\
&\qquad \left. u_{1j}^{pn} u_{ij}^{(n)} u_{1j}^{an}, \ldots, u_{ij}^{pn} u_{ij}^{(n)} u_{ij}^{an}, \ldots \right].
\end{aligned} \tag{15}$$

We have

$$\begin{aligned}
\frac{\partial Tr\left[S_w\right]}{\partial u_{ij}^{(n)}} &= 2 \sum_p \sum_{c=1}^{C} \sum_{h_p \in V(c)} \left(W^T \overline{X_p} - m_c \right) \\
&\quad \times \left[0, \ldots, 0, u_{1j}^{pn} u_{1j}^{an}, \ldots, u_{ij}^{pn} u_{ij}^{an}, \ldots, 0, \ldots, 0 \right] \overline{X_p}.
\end{aligned} \tag{16}$$

Similarly, we have

$$\begin{aligned}
&\frac{\partial Tr\left[S_b\right]}{\partial u_{ij}^{(n)}} \\
&= \frac{\partial \sum_p \sum_{c=1}^{C} N_c \left(m_p^c - m_p \right) \left(m_p^c - m_p \right)^T}{\partial u_{ij}^{(n)}} \\
&= \frac{\partial \sum_p \sum_{c=1}^{C} N_c \left(\sum_{h_p \in V(c)}^{N_c} W^T \overline{X_p} - \sum_{q=1}^{M} W^T \overline{X_q} \right)^2}{\partial u_{ij}^{(n)}} \\
&= 2 \sum_p \sum_{c=1}^{C} N_c \left(m_p^c - m_p \right) \\
&\quad \times \frac{\partial \left(\sum_{h_p \in V(c)}^{N_c} W^T \overline{X_p} - \sum_{q=1}^{M} W^T \overline{X_q} \right)}{\partial u_{ij}^{(n)}} \\
&= 2 \sum_p \sum_{c=1}^{C} N_c \left(m_p^c - m_p \right) \\
&\quad \times \left(\partial \left(U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(1)} \right)^T \right. \\
&\qquad \left. \times \left(\sum_{h_p \in V(c)}^{N_c} \overline{X_p} - \sum_{q=1}^{M} \overline{X_q} \right) \right) \times \left(\partial u_{ij}^{(n)} \right)^{-1} \\
&= 2 \sum_p \sum_{c=1}^{C} N_c \left(m_p^c - m_p \right) \\
&\quad \times \left[0, \ldots, 0, u_{1j}^{pn} u_{1j}^{an}, \ldots, u_{ij}^{pn} u_{ij}^{an}, \ldots, 0, \ldots, 0 \right] \\
&\quad \times \left(\sum_{h_p \in V(c)}^{N_c} \overline{X_p} - \sum_{q=1}^{M} \overline{X_q} \right).
\end{aligned} \tag{17}$$

To solve the weight matrix $U^{(N)}$, the objective function is

$$f\left(U^{(N)}\right) = \left\| X_{(N)} - U^{(N)} \left(U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(1)} \right)^T \right\|^2. \tag{18}$$

The gradient functon is

$$g\left(U^{(N)}\right) = -2X_{(N)}W + 2U^{(N)}W^TW, \tag{19}$$

where $W = U^{(N-1)} \odot U^{(N-2)} \odot \cdots \odot U^{(1)}$.

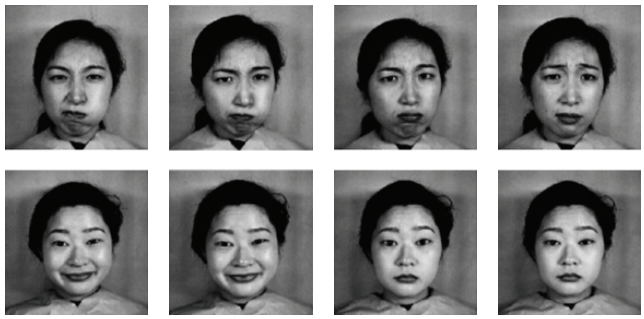


Figure 1: Some images in the Jaff facial expression database.

Table 1: Comparison of the best recognition rates for all tested algorithms.

Algorithms	Recognition rate	Algorithms	Recognition rate
NMF	79.19%	NMFOS	89.06%
DNMF	92.06%	FisherNMF	92.06%
DNTF	95.24%	GDONTF	97.07%

Consequently, the update rules of $U^{(N)}$ are

$$U_{ij}^{(N)^{t+1}} = U_{ij}^{(N)^t} \frac{\left(X_{(N)}W\right)_{ij}}{\left(U^{(N)}W^TW\right)_{ij}}. \quad (20)$$

3. Experiments

We have conducted facial expression recognition in order to compare the GDONTF with other algorithms such as NMFOS [19], DNMF [6], FisherNMF [16], and DNTF [14]. Because these algorithms calculate low-dimension features in iteration form, the iteration number is 100. For NMFOS and GDONTF, $\lambda = 1$. $\gamma = 0.5$ in DNMF and $\alpha = 1$ in FisherNMF. All low-dimension features are classified by SVM with linear kernel.

The database used for the facial expression recognition experiments is Jaff facial expression database [20]. The database contains 213 images of ten Japanese women. Each person has two to four images for each of the seven expressions: neutral, happy, sad, surprise, anger, disgust, and fear. Each image is resized into 32 × 32. A few examples are shown in Figure 1. We randomly select 20 images from each expression for training; the rest is used for testing. The recognition rates with various dimensionalities of different algorithms are shown in Figure 2. Table 1 shows the best recognition rates of the above algorithms. Because NMF is unsupervised learning algorithm, it has the lowest recognition rates. DNMF and FisherNMF have better recognition rates with supervised learning. It is interesting that NMFOS is superior to DNMF and FisherNMF when the feature dimensionality is from 16 to 160 and is better than DNTF when the feature dimensionality is from 16 to 40, which also illustrates the validity of the orthogonal constraint. It is obvious that GDONTF outperforms other algorithms and the best recognition rate is up to 97.07%.

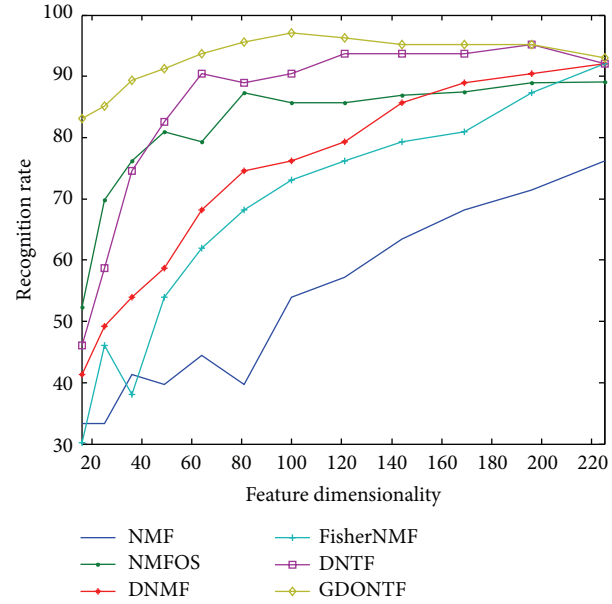


Figure 2: Facial expression recognition rate versus dimensionality in Jaff database.

4. Conclusion

In this paper, a generalized discriminant orthogonal nonnegative tensor factorization algorithm is proposed considering the orthogonal constraint and the discriminant constraint. For the algorithm, the non-negativity of the low-dimensional features is preserved due to the orthogonal constraint for either training samples or testing samples. In order to enhance the recognition accuracy, the discriminant is conducted on low-dimensional features instead of the weight coefficient of the basis images. The experiments also validate the performance of the algorithm.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] I. Buciu and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 288–291, August 2004.

[3] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.

[4] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.

[5] I. Kotsia, S. Zafeiriou, and I. Pitas, "Novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 588–595, 2007.

[6] B. Ioan and N. Ioan, "Non-negative matrix factorization methods for face recognition under extreme lighting variations," in *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS '09)*, Iasi, Romania, July 2009.

[7] J. Lu and Y.-P. Tan, "Doubly weighted nonnegative matrix factorization for imbalanced face recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 877–880, IEEE Computer Society, Taipei, Taiwan, April 2009.

[8] H. Tong and C.-Y. Lin, "Non-negative residual matrix factorization: problem definition, fast solutions, and applications," *Statistical Analysis and Data Mining*, vol. 5, no. 1, pp. 3–15, 2012.

[9] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts," *Advances in Neural Information Processing Systems*, vol. 16, pp. 1141–1148, 2004.

[10] J. Liu, J. Liu, P. Wonka, and J. Ye, "Sparse non-negative tensor factorization using columnwise coordinate descent," *Pattern Recognition*, vol. 45, no. 1, pp. 649–656, 2012.

[11] T. Van De Cruys, "A non-negative tensor factorization model for selectional preference induction," *Natural Language Engineering*, vol. 16, no. 4, pp. 417–437, 2010.

[12] S. Zafeiriou, "Discriminant nonnegative tensor factorization algorithms," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 217–235, 2009.

[13] C. Liu, K. He, J.-L. Zhou, and Y.-L. Zhu, "Facial Expression recognition based on Orthogonal Nonnegative CP Factorization," *Advanced Materials Research*, vol. 143-144, pp. 111–115, 2011.

[14] S. Zafeiriou, "Discriminant nonnegative tensor factorization algorithms," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 217–235, 2009.

[15] I. Buciu and I. Pitas, "NMF, LNMF, and DNMF modeling of neural receptive fields involved in human facial expression perception," *Journal of Visual Communication and Image Representation*, vol. 17, no. 5, pp. 958–969, 2006.

[16] Y. Wang, Y. Jia, H. U. Changbo, and M. Turk, "Non-negative matrix factorization framework for face recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 4, pp. 495–511, 2005.

[17] X. Li and K. Fukui, "Fisher non-negative matrix factorization with pairwise weighting," in *Proceedings of the IAPR Conference on Machine Vision Applications*, MVA Press, Tokyo, Japan, 2007.

[18] A. Vilamala, P. J. G. Lisboab, S. Ortega-Martorellc, and A. Vellidoa, "Discriminant convex non-negative matrix factorization for the classification of human brain tumours," *Pattern Recognition Letters*, vol. 34, no. 14, pp. 1734–1747, 2013.

[19] Z. Li, X. Wu, and H. Peng, "Nonnegative matrix factorization on orthogonal subspace," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 905–911, 2010.

[20] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 130512, 8 pages
http://dx.doi.org/10.1155/2014/130512

Research Article

Global Systems for Mobile Position Tracking Using Kalman and Lainiotis Filters

Nicholas Assimakis[1] and Maria Adam[2]

[1] *Department of Electronic Engineering, Technological Educational Institute of Central Greece, 3rd km Old National Road Lamia-Athens, 35100 Lamia, Greece*
[2] *Department of Computer Science and Biomedical Informatics, University of Thessaly, 2-4 Papasiopoulou Street, 35100 Lamia, Greece*

Correspondence should be addressed to Maria Adam; madam@dib.uth.gr

Received 3 August 2013; Accepted 2 January 2014; Published 23 February 2014

Academic Editors: S. Bourennane and J. Marot

We present two time invariant models for Global Systems for Mobile (GSM) position tracking, which describe the movement in x-axis and y-axis simultaneously or separately. We present the time invariant filters as well as the steady state filters: the classical Kalman filter and Lainiotis Filter and the Join Kalman Lainiotis Filter, which consists of the parallel usage of the two classical filters. Various implementations are proposed and compared with respect to their behavior and to their computational burden: all time invariant and steady state filters have the same behavior using both proposed models but have different computational burden. Finally, we propose a Finite Impulse Response (FIR) implementation of the Steady State Kalman, and Lainiotis filters, which does not require previous estimations but requires a well-defined set of previous measurements.

1. Introduction

The Global Positioning System (GPS) is the most popular positioning technique in navigation providing reliable mobile location estimates in many applications [1–4]. Thus wireless location systems offering reliable mobile location estimates have been studied by researchers and engineers over the past few years. Various techniques require one base station or at least two base stations or more than three base stations in order to determine the location of the user. The accuracy of the positioning results is affected by many interference sources as the signals propagate in the atmosphere. So, techniques were developed using filters to estimate the location of the user through the location information exchanged between the handset and the base station. Kalman filter has been used in the localization process [4–6], due to the following advantages mentioned in [5]: (a) Kalman filter [7–9] processes noisy measurements and so it can smooth out the effects of noise in the estimated state variables by integrating more information from reliable data more than unreliable data and (b) Kalman filter allows the combination of measurements from different sources (locomotion data) and different times. Kalman filter was implemented for Global Systems for Mobile (GSM) position tracking in [5]: Kalman filter was used for tracking in two dimensions and it was stated that Kalman filter is very powerful due to its reliable performance, because it yielded enhanced position tracking results.

In this paper we extend the ideas in [5] in two fields: (a) by using two models for GSM position tracking, which describe the movement in x-axis and y-axis simultaneously or separately and (b) by using the Kalman filter and the Lainiotis filter [8, 10]. The paper is organized as follows. In Section 2, we present two time invariant models for Global Systems for Mobile (GSM) position tracking, which describe the movement in x-axis and y-axis. In Section 3, we present the time invariant filters: Kalman filter, Lainiotis Filter and Join Kalman Lainiotis Filter. In Section 4, we present the corresponding steady state filters. In Section 5, various implementations are proposed. In Section 6, we compare the filters with respect to their behavior and to their computational

burden. In Section 7, we propose a Finite Impulse Response (FIR) implementation of the Steady State Kalman and Lainiotis Filters. Finally, Section 8 summarizes the conclusions.

2. Time Invariant Models

Linear estimation is associated with time invariant systems described by the following state space equations:

$$\begin{aligned} x(k+1) &= Fx(k) + Gw(k), \\ z(k) &= Hx(k) + v(k) \end{aligned} \tag{1}$$

for $k \geq 0$, where $x(k)$ is the n-dimensional state vector at time k, $z(k)$ is the m-dimensional measurement vector at time k, F is the $n \times n$ system transition matrix, H is the $m \times n$ output matrix, $w(k)$ is the plant noise at time k, $v(k)$ is the measurement noise at time k. Also, $\{w(k)\}$ and $\{v(k)\}$ are Gaussian zero-mean white random processes with covariance matrices Q and R, respectively. The initial state $x(0)$ is a Gaussian random variable with mean x_0 and covariance P_0 and is assumed to be independent of $w(k)$ and $v(k)$.

In this paper we consider two models.

Model A. The first model (model A) describes the movement in x-axis and y-axis simultaneously and follows the ideas in [5].

The state vector is of dimension $n = 4$ and contains the position and the velocity in x-axis and y-axis: $x(k) = \left[s_x(k) \;\; v_x(k) \;\; s_y(k) \;\; v_y(k)\right]^T$. The measurement vector is of dimension $m = 2$ and contains the measured position x-axis and y-axis: $z(k) = \left[z_x(k) \;\; z_y(k)\right]^T$.

Then we take:

$$F = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad G = \begin{bmatrix} \frac{1}{2}\Delta t & 0 \\ 1 & 0 \\ 0 & \frac{1}{2}\Delta t \\ 0 & 1 \end{bmatrix}, \tag{2}$$

$$H = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & \Delta t \end{bmatrix}.$$

The plant noise $w(k) = \left[w_x(k) \;\; w_y(k)\right]^T$ is Gaussian zero-mean with covariance matrix $Q = \begin{bmatrix} \sigma_{xq}^2 & 0 \\ 0 & \sigma_{yq}^2 \end{bmatrix}$.

The measurement noise $v(k) = \left[v_x(k) \;\; v_y(k)\right]^T$ is Gaussian zero-mean with covariance matrix $R = \begin{bmatrix} \sigma_{xr}^2 & 0 \\ 0 & \sigma_{yr}^2 \end{bmatrix}$.

Model B. The second model (model B) describes the movement in x-axis and y-axis separately. In each axis, the state vector is of dimension $n = 2$ and contains the position and the velocity: $x(k) = \left[s(k) \;\; v(k)\right]^T$. The measurement vector vector is of dimension $m = 1$ and contains the measured position $z(k)$.

Then we take:

$$F = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \qquad G = \begin{bmatrix} \frac{1}{2}\Delta t \\ 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1 & \Delta t \end{bmatrix}. \tag{3}$$

The plant noise $w(k)$ is Gaussian zero-mean with covariance matrix $Q = \sigma_q^2$.

The measurement noise $v(k)$ is Gaussian zero-mean with covariance matrix $R = \sigma_r^2$.

It is obvious that we are able to describe the movement in both axes using two separate state vectors: $x_x(k) = \left[s_x(k) \;\; v_x(k)\right]^T$ for the x-axis and $x_y(k) = \left[s_y(k) \;\; v_y(k)\right]^T$ for the y-axis. If we merge these two state vectors, we take the state vector $x(k) = \left[s_x(k) \;\; v_x(k) \;\; s_y(k) \;\; v_y(k)\right]^T$ of model A.

3. Time Invariant Kalman and Lainiotis Filters

In this section, we present the classical time invariant Kalman filter [7–9] and Lainiotis Filter [8, 10], which are the most well-known algorithms that solve the filtering problem. Both algorithms compute the estimation $x(k/k)$ and the corresponding estimation error covariance $P(k/k)$. We also propose the Join Kalman-Lainiotis Filter which consists of the parallel (with the same measurements) usage of *two* filters (one Kalman filter and one Lainiotis Filter) and combination of the results (weight 50% for each filter).

Kalman Filter (KF). The following equations constitute the KF:

$$\begin{aligned} x(k+1/k) &= Fx(k/k), \\ P(k+1/k) &= \left(GQG^T\right) + FP(k/k)F^T, \\ K(k+1) &= P(k+1/k)H^T\left[HP(k+1/k)H^T + R\right]^{-1}, \\ x(k+1/k+1) &= [I - K(k+1)H]\,x(k+1/k) \\ &\quad + K(k+1)z(k+1), \\ P(k+1/k+1) &= [I - K(k+1)H]\,P(k+1/k), \end{aligned} \tag{4}$$

for $k \geq 0$, with initial conditions $x(0/0) = x_0$ and $P(0/0) = P_0$.

The Kalman filter computes the estimation $x(k/k)$ and the estimation error covariance $P(k/k)$ through the prediction $x(k+1/k)$ and the corresponding prediction error covariance $P(k+1/k)$ using the Kalman filter gain $K(k)$.

Lainiotis Filter (LF). The following equations constitute the LF:

$$\begin{aligned} x(k+1/k+1) &= K_n z(k+1) + F_n\left[I + P(k/k)O_n\right]^{-1} \\ &\quad \times \left[P(k/k)K_m z(k+1) + x(k/k)\right], \end{aligned} \tag{5}$$

$$P(k+1/k+1) = P_n + F_n\left[I + P(k/k)O_n\right]^{-1}P(k/k)F_n^T \tag{6}$$

for $k \geq 0$, with initial conditions $x(0/0) = x_0$ and $P(0/0) = P_0$, where

$$\begin{aligned} A &= \left[H\left(GQG^T\right)H^T + R\right]^{-1}, \\ K_n &= \left(GQG^T\right)H^T A, \\ K_m &= F^T H^T A, \end{aligned}$$

$$P_n = \left(GQG^T\right) - K_nH\left(GQG^T\right),$$
$$F_n = F - K_nHF,$$
$$O_n = F^TH^TAHF. \tag{7}$$

Join Kalman-Lainiotis Filter (JKLF). The filter consists of the parallel usage of *two* filters (one Kalman filter and one Lainiotis Filter) with the same measurements and combination of the results (weight 50% for each filter):

$$x(k/k) = \frac{1}{2}x_{\mathrm{KF}}(k/k) + \frac{1}{2}x_{\mathrm{LF}}(k/k),$$
$$P(k/k) = \frac{1}{2}P_{\mathrm{KF}}(k/k) + \frac{1}{2}P_{\mathrm{LF}}(k/k). \tag{8}$$

4. Steady State Kalman and Lainiotis Filters

For time invariant systems, it is well known [7] that there exists a steady state value P_p of the prediction error covariance matrix, if the signal process model is asymptotically stable, or if the signal process model is not necessarily asymptotically stable, but the pair $[F, H]$ is completely detectable and the pair $[F, GG_1]$ is completely stabilizable for any G_1 with $G_1G_1^T = Q$. Then there also exist a steady state value P_e of the estimation error covariance matrix and a steady state value K of the Kalman filter gain.

In this section we present the Steady State Kalman filter and Lainiotis Filter. Both algorithms compute the estimation $x(k/k)$ using the previous estimation and the current measurement. We also propose the Join Steady State Kalman-Lainiotis Filter, which consists of the parallel usage of *two* filters (one Steady State Kalman filter and one Steady State Lainiotis Filter) with the same measurements and combination of the results (weight 50% for each filter).

Steady State Kalman Filter (SSKF). The following equation constitutes the SSKF:

$$x(k+1/k+1) = A_{\mathrm{KF}}x(k/k) + B_{\mathrm{KF}}z(k+1) \tag{9}$$

for $k \geq 0$, with initial condition $x(0/0) = x_0$, where

$$A_{\mathrm{KF}} = [I - KH]F,$$
$$B_{\mathrm{KF}} = K. \tag{10}$$

The steady state Kalman filter gain K is computed by $K = P_pH^T[HP_pH^T + R]^{-1}$, where P_p is the steady state prediction error covariance computed by solving the *Riccati equation emanating from Kalman filter (REKF):*

$$P_p = \left(GQG^T\right) + \mathrm{FP}_pF^T - \mathrm{FP}_pH^T\left[\mathrm{HP}_pH^T + R\right]^{-1}\mathrm{HP}_pF^T. \tag{11}$$

In view of the importance of the Riccati equation emanating from Kalman filter, there exists considerable literature on its algebraic solutions [7, 11] or iterative solutions [7, 12–15] concerning per step or doubling algorithms.

Steady State Lainiotis Filter (SSLF). The following equation constitutes the SSLF:

$$x(k+1/k+1) = A_{\mathrm{LF}}x(k/k) + B_{\mathrm{LF}}z(k+1), \tag{12}$$

for $k \geq 0$, with initial condition $x(0/0) = x_0$, where

$$A_{\mathrm{LF}} = F_n\left[I + P_eO_n\right]^{-1},$$
$$B_{\mathrm{LF}} = K_n + F_n\left[I + P_eO_n\right]^{-1}P_eK_m, \tag{13}$$

and P_e is the steady state estimation error covariance computed by solving the *Riccati equation emanating from Lainiotis filter (RELF):*

$$P_e = P_n + F_n\left[I + P_eO_n\right]^{-1}P_eF_n^T. \tag{14}$$

In view of the importance of the Riccati equation emanating from Lainiotis Filter, there exists considerable literature on its algebraic or iterative solutions [12, 14–16] concerning per step or doubling algorithms.

Note that in [8] it is shown that SSKF is equivalent to SSLF, since

$$A_{\mathrm{KF}} = A_{\mathrm{LF}},$$
$$B_{\mathrm{KF}} = B_{\mathrm{LF}}. \tag{15}$$

Join Steady State Kalman-Lainiotis Filter (JSSKLF). The filter consists of the parallel usage of *two* steady state filters (one Steady State Kalman filter and one Steady State Lainiotis Filter) with the same measurements and combination of the results (weight 50% for each filter):

$$x(k/k) = \frac{1}{2}x_{\mathrm{KF}}(k/k) + \frac{1}{2}x_{\mathrm{LF}}(k/k), \tag{16}$$

for $k \geq 0$.

5. Implementations

In this section, we propose various implementations.

The use of model A which describes the movement in x-axis and y-axis simultaneously requires the use one filter; we are able to use KF/LF/SSKF/SSLF/JKLF in order to compute the estimation and the corresponding estimation error covariance.

The use of model B, which describes the movement in x-axis and y-axis separately, requires the use of two filters KF/LF/SSKF/SSLF/JSSKLF in order to compute the estimation and the corresponding estimation error covariance for each movement. It is obvious that, if we merge the estimation $x_x(k/k) = \left[s_x(k/k)\ \ v_x(k/k)\right]^T$ for the movement in x-axis and the estimation $x_y(k/k) = \left[s_y(k/k)\ \ v_y(k/k)\right]^T$ for the movement in y-axis, we take the state vector of model A:

$$x(k/k) = \left[s_x(k/k)\ \ v_x(k/k)\ \ s_y(k/k)\ \ v_y(k/k)\right]^T$$
$$= \left[x_x(k/k)\ \ x_y(k/k)\right]^T. \tag{17}$$

Table 1: GSM position tracking implementations.

Implementation	Model	System	Filter
1	Model A	Time invariant	KF
2			LF
3			JKLF
4		Steady state	SSKF
5			SSLF
6			JSSKLF
7	Model B	Time invariant	KF
8			LF
9			JKLF
10		Steady state	SSKF
11			SSLF
12			JSSKLF

Also, the estimation error covariances $P_x(k/k)$ and $P_y(k/k)$ for each movement can be merged to the estimation error covariance of model A:

$$P(k/k) = \begin{bmatrix} P_x(k/k) & 0 \\ 0 & P_y(k/k) \end{bmatrix}. \quad (18)$$

Thus, we propose various implementations for Global Systems for Mobile (GSM) position tracking, as it is shown in Table 1.

6. Comparison of the Filters

In this section we compare the filters with respect to their behavior and to their computational burden.

Example 1. We implemented the filters with the following parameters:

(i) discretization factor: $\Delta t = 1$,

(ii) movement reliability: $\sigma_{xq}^2 = \sigma_{yq}^2 = 0.01$,

(iii) measurements reliability: $\sigma_{xr}^2 = \sigma_{yr}^2 = 0.1$,

(iv) initial conditions: $x_0 = 0$ and $P_0 = 0$.

Concerning the behavior of the filters, we found that

(i) the time invariant filters KF, LF and JKLF are equivalent, since they compute the same outputs (estimation and estimation error covariance), using model A or model B,

(ii) the steady state filters SSKF, SSLF and JSSKLF are equivalent, since they compute the same outputs (estimation and estimation error covariance), using model A or model B,

(iii) the steady state filters and the time invariant filters compute outputs very close to each other,

(iv) model A is equivalent to model B, since they produce the same outputs.

These results are depicted in Figure 1.

Concerning the computational burden of the filters, we compared the filters with respect to their per-iteration calculation burdens, computed using the ideas in [8], as shown in Table 2.

Table 3 summarizes the per-iteration calculation burden of all implementations, using model A and model B.

We observe that

(i) KF is faster than LF,

(ii) JKLF is slower than KF and LF (since the join filter requires the implementation of both the Kalman and Lainiotis filters),

(iii) SSKF is as fast as SSLF,

(iv) SSKF and SSLF are faster than KF and LF,

(v) JSSKLF is slower than SSKF and SSLF,

(vi) the filters using model B are faster than the same filters using model A.

Table 4 summarizes speedup between the various implementations.

We observe that

(i) KF is faster than LF,

speedup (LF model A to KF model A) = 1.316,
speedup (LF model B to KF model B) = 1.290.

(ii) Model B is faster than model A,

speedup (KF model A to SSKF model B) = 25.450,
speedup (LF model A to SSLF model B) = 33.500.

7. FIR Steady State Kalman and Lainiotis Filters

In this section we propose an FIR implementation of the Steady State Kalman filter and the Steady State Lainiotis Filter.

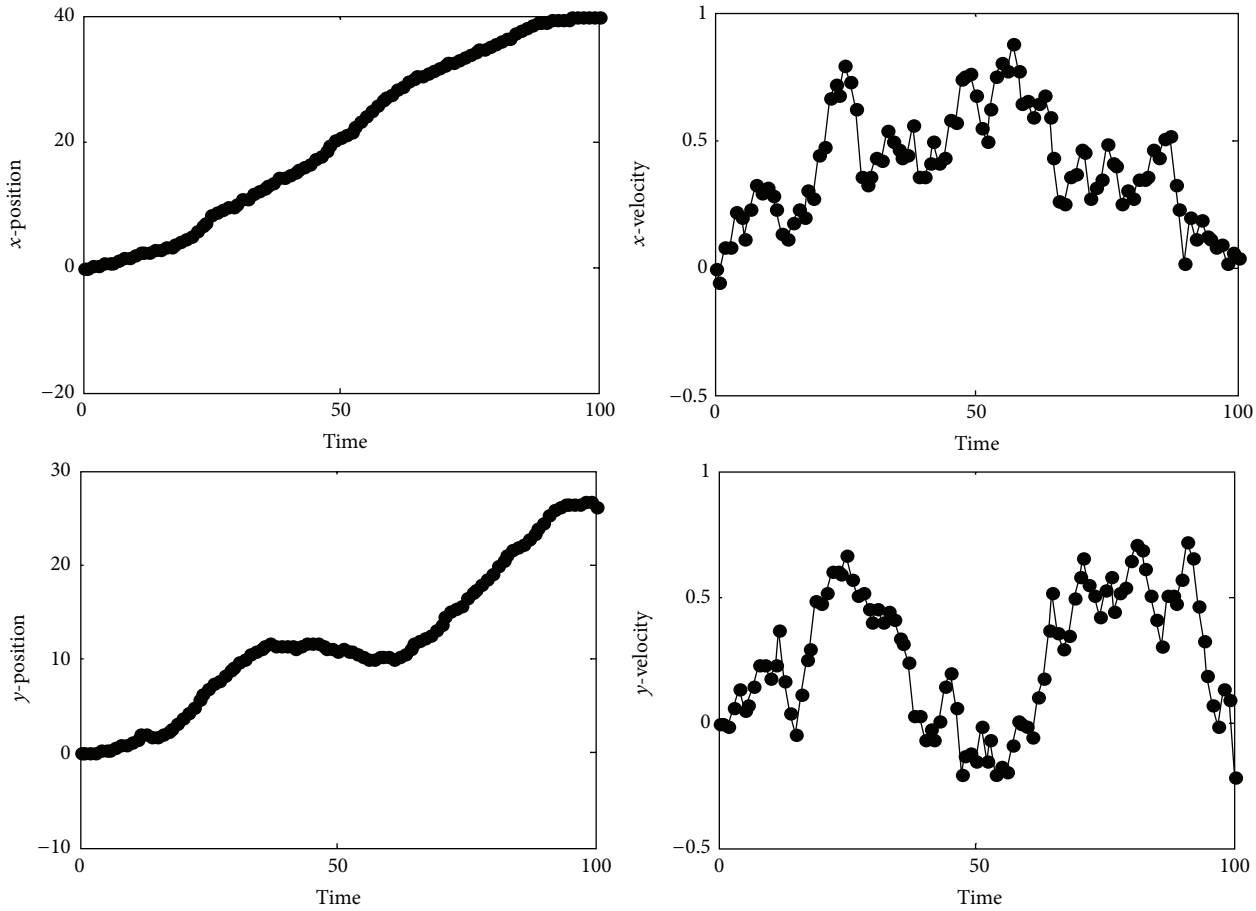


Figure 1: Position and velocity estimation solid line: KF/LF/JKLF, dashed line: SSKF/SSLF/JSSKLF.

Table 2: Per-iteration calculation burden of filters.

KF	$4n^3 + 3.5n^2 - 1.5n + 4n^2m + nm + 3nm^2 + (16m^3 - 3m^2 - m)/6$
LF	$4nm + (58n^3 + 9n^2 - 7n)/6$
JKLF	$n^2 + 3n$ (join procedure)
SSKF	$2n^2 + 2nm - n$
SSLF	$2n^2 + 2nm - n$
JSSKLF	$2n$ (join procedure)

Table 3: Per-iteration calculation burden of implementations.

Implementation	Model	System	Filter	Calculation burden
1	Model A	Time invariant	KF	509
2			LF	670
3			JKLF	1207
4		Steady state	SSKF	44
5			SSLF	44
6			JSSKLF	96
7	Model B	Time invariant	KF	138
8			LF	178
9			JKLF	326
10		Steady state	SSKF	20
11			SSLF	20
12			JSSKLF	44

Table 4: Speedup.

			1	2	3	4	5	6	7	8	9	10	11	12
			Model A						Model B					
			KF	LF	JKLF	SSKF	SSLF	JSSKLF	KF	LF		SSKF	SSLF	JSSKLF
1	Model A	KF		0,760	0,427	11,568	11,568	4,314	3,688	2,860	0,795	25,450	25,450	5,194
2		LF	1,316		0,563	15,227	15,227	5,678	4,855	3,764	1,047	33,500	33,500	6,837
3		JKLF	2,340	1,778		27,068	27,068	10,093	8,630	6,691	1,861	59,550	59,550	12,153
4		SSKF	0,086	0,066	0,037		1,000	0,373	0,319	0,247	0,069	2,200	2,200	0,449
5		SSLF	0,086	0,066	0,037	1,000		0,373	0,319	0,247	0,069	2,200	2,200	0,449
6		JSSKLF	0,232	0,176	0,099	2,682	2,682		0,855	0,663	0,184	5,900	5,900	1,204
7	Model B	KF	0,271	0,206	0,116	3,136	3,136	1,169		0,775	0,216	6,900	6,900	1,408
8		LF	0,350	0,266	0,149	4,045	4,045	1,508	1,290		0,278	8,900	8,900	1,816
9		JKLF	1,257	0,955	0,537	14,545	14,545	5,424	4,638	3,596		32,000	32,000	6,531
10		SSKF	0,039	0,030	0,017	0,455	0,455	0,169	0,145	0,112	0,031		1,000	0,204
11		SSLF	0,039	0,030	0,017	0,455	0,455	0,169	0,145	0,112	0,031	1,000		0,204
12		JSSKLF	0,193	0,146	0,082	2,227	2,227	0,831	0,710	0,551	0,153	4,900	4,900	

Table 5: Calculation burden of classical and FIR Steady State filters.

SSKF/SSLF	$2n^2 + 2nm - n$	Per-iteration
FIR SSKF/SSLF	$2n^2 + 2nm - n + (M-1)n$	When required

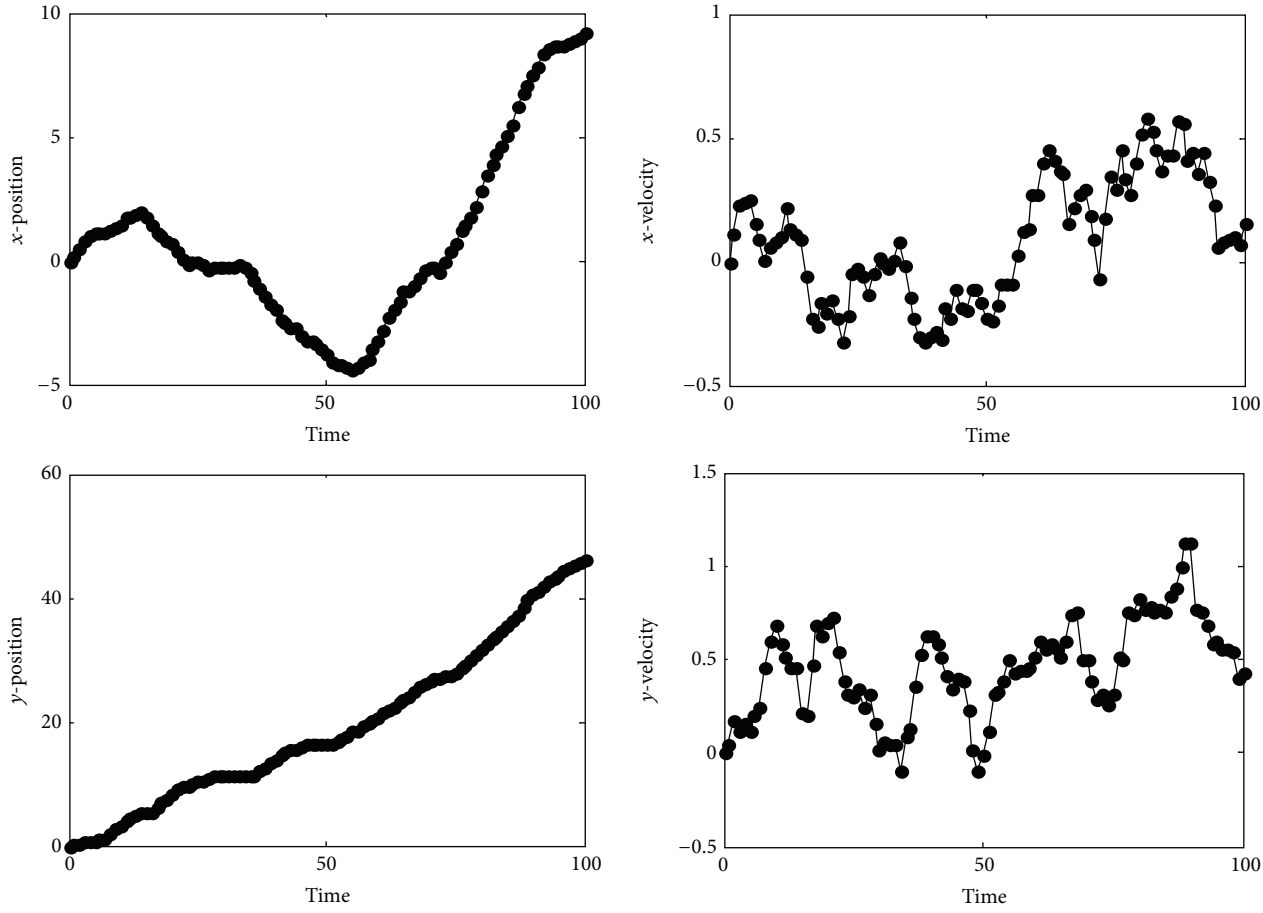


Figure 2: Position and velocity estimation solid line: SSKF/SSLF, dashed line: FIR SSKF/SSLF.

Recall that SSKF and SSLF have equal parameters:

$$A = A_{\mathrm{KF}} = A_{\mathrm{LF}}, \qquad B = B_{\mathrm{KF}} = B_{\mathrm{LF}}. \tag{19}$$

Then we are able to write

$$x(k+1/k+1) = Ax(k/k) + Bz(k+1), \tag{20}$$

for $k \geq 0$ with initial condition $x(0/0) = x_0$.

Then we take:

$$\begin{aligned} x(1/1) &= Ax_0 + Bz(1) \\ x(2/2) &= A^2 x_0 + ABz(1) + Bz(2) \\ &\vdots \\ x(\nu/\nu) &= A^{\nu} x_0 + A^{\nu-1} Bz(1) + \cdots + ABz(\nu-1) + Bz(\nu). \end{aligned} \tag{21}$$

Using the ideas in [17], the resulting FIR SSKF/SSLF is formulated as

$$x(k/k) = \sum_{j=0}^{M-1} A^j Bz(k-j), \quad \text{for } k \geq 1, \tag{22}$$

where $A = A_{\mathrm{KF}} = A_{\mathrm{LF}}$ and $B = B_{\mathrm{KF}} = B_{\mathrm{LF}}$ and M is the FIR filter order defined by $\|A^M\| < \varepsilon$ and $\|A^{M-1}\| \geq \varepsilon$, with ε a small real value.

Remarks. (1) The FIR steady state filter coefficients can be calculated off-line by solving the corresponding Riccati equation.

(2) The FIR steady state filter does not require previous estimations but it requires a well-defined set of M previous measurements. This means that we have to wait for M time moments in order to produce the results. Alternatively, we are able to use only the available measurements until time M is reached or to use SSKF until time M.

Table 6: Calculation burden of classical and FIR steady state filters implementations.

	Model A	Model B
SSKF/SSLF	44	20
FIR SSKF/SSLF	$40 + 4M$	$16 + 4M$

Table 7: FIR results for Example 1.

$M = 17$	Model A	Model B
SSKF/SSLF	44	20
FIR SSKF/SSLF	108	84
μ	3	5

We implemented the FIR filter for the parameters of the example in Section 6. We used $\varepsilon = 10^{-3}$ and we found $M = 17$.

The steady state filters and FIR steady state filters compute outputs very close to each other, as depicted in Figure 2.

Concerning the computational burden, the FIR steady state filter possesses a constant burden while the classical steady state filter (SSKF/SSLF) possesses a constant per-iteration computational burden, as it is shown in Table 5.

Table 6 summarizes the calculation burden of the classical and FIR steady state filters implementations, using model A and model B.

Thus, FIR SSKF/SSLF can be faster than SSKF/SSLF, if we take results using FIR SSKF/SSLF every (or more than) μ time lags, where μ is the nearest integer greater than or equal to the ratio $(40 + 4M)/44 = (10 + M)/11$ for model A and $(16 + 4M)/20 = (4 + M)/5$ for model B.

For our example, we take the results, which are appeared in Table 7.

8. Conclusions

In this paper we presented two time invariant models for Global Systems for Mobile (GSM) position tracking, which describe the movement in x-axis and y-axis simultaneously or separately. We presented the time invariant filters as well as the steady state filters: the classical Kalman filter and Lainiotis Filter and the Join Kalman Lainiotis Filter, which consists of the parallel usage of the two classical filters. Various implementations are proposed and compared with respect to their behavior and to their computational burden. We found that all time invariant and steady state filters have the same behavior using both of the proposed models. We found that (a) Kalman filter is faster than Lainiotis Filter, (b) Join Kalman Lainiotis Filter is slower than both Kalman filter and Lainiotis Filter, (c) steady state filters are faster than time invariant filters and (d) the filters using the model, which handles the movement in x-axis and y-axis separately, are faster than the same filters using the model, which handles the movement in x-axis and y-axis simultaneously. Finally, we proposed an FIR implementation of the Steady State Kalman and Lainiotis Filters, which does not require previous estimations but it requires a well-defined set of previous measurements.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] C. Drane, M. Macnaughtan, and C. Scott, "Positioning GSM telephones," *IEEE Communications Magazine*, vol. 36, no. 4, pp. 46–59, 1998.

[2] A. Küpper, *Location-Based Services*, John Wiley and Sons, 1st edition, 2005.

[3] M. Mouly and M. Pautet, *The GSM System for Mobile Communications*, Telecom Publishing, Washington, DC, USA, 1992.

[4] S. Rooney, P. Chippendale, R. Choony, C. Le Roux, and B. Honary, "Accurate vehicular positioning using a DAB-GSM hybrid system," in *Proceedings of the 51st Vehicular Technology Conference*, pp. 97–101, May 2000.

[5] J. P. Dubois, J. S. Daba, M. Nader, and C. El Ferkh, "GSM position tracking using a Kalman filter," *Wolrd Academy of Science, Engineering and Technology*, vol. 68, pp. 1610–1619, 2012.

[6] T. Kos, M. Grgic, and G. Sisul, "Mobile user positioning in GSM/UMTS cellular networks," in *Proceedings of the 48th International Symposium ELMAR-2006 on Multimedia Signal Processing and Communications*, pp. 185–188, June 2006.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Dover Publications, New York, NY, USA, 2005.

[8] N. Assimakis and M. Adam, "Discrete time Kalman and Lainiotis filters comparison," *International Journal of Mathematical Analysis*, vol. 1, no. 13, pp. 635–659, 2007.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering D*, vol. 82, no. 1, pp. 34–45, 1960.

[10] D. G. Lainiotis, "Partitioned linear estimation algorithms: discrete case," *IEEE Transactions on Automatic Control*, vol. 20, no. 2, pp. 255–257, 1975.

[11] D. R. Vaughan, "Nonrecursive algebraic solution for the disczte Riccati equation," *IEEE Transactions on Automatic Control*, vol. 15, no. 5, pp. 597–599, 1970.

[12] N. D. Assimakis, D. G. Lainiotis, S. K. Katsikas, and F. L. Sanida, "A survey of recursive algorithms for the solution of the discrete time riccati equation," *Nonlinear Analysis: Theory, Methods and Applications*, vol. 30, no. 4, pp. 2409–2420, 1997.

[13] N. Assimakis, S. Roulis, and D. Lainiotis, "Recursive solutions of the discrete time Riccati equation," *Neural, Parallel and Scientific Computations*, vol. 11, pp. 343–350, 2003.

[14] D. G. Lainiotis, N. D. Assimakis, and S. K. Katsikas, "A new computationally effective algorithm for solving the discrete Riccati equation," *Journal of Mathematical Analysis and Applications*, vol. 186, no. 3, pp. 868–895, 1994.

[15] D. G. Lainiotis, N. D. Assimakis, and S. K. Katsikas, "Fast and numerically robust recursive algorithms for solving the discrete time Riccati equation: the case of nonsingular plant noise covariance matrix," *Neural, Parallel and Scientific Computations*, vol. 3, no. 4, pp. 565–584, 1995.

[16] D. G. Lainiotis, "Discrete Riccati equation solutions: partitioned algorithms," *IEEE Transactions on Automatic Control*, vol. 20, no. 4, pp. 555–556, 1975.

[17] N. Assimakis and M. Adam, "FIR implementation of the steady state Kalman filter," *International Journal of Signal and Imaging Systems Engineering*, vol. 1, no. 3-4, pp. 279–286, 2008.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 487506, 6 pages
http://dx.doi.org/10.1155/2014/487506

Research Article

Through-Wall Image Enhancement Using Fuzzy and QR Decomposition

Muhammad Mohsin Riaz[1] and Abdul Ghafoor[2]

[1] *Centre for Advanced Studies in Telecommunication (CAST), Comsats, Islamabad, Pakistan*
[2] *Department of Electrical Engineering, College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan*

Correspondence should be addressed to Abdul Ghafoor; abdulghafoor-mcs@nust.edu.pk

Received 5 August 2013; Accepted 2 January 2014; Published 23 February 2014

Academic Editors: S. Bourennane and J. Marot

QR decomposition and fuzzy logic based scheme is proposed for through-wall image enhancement. QR decomposition is less complex compared to singular value decomposition. Fuzzy inference engine assigns weights to different overlapping subspaces. Quantitative measures and visual inspection are used to analyze existing and proposed techniques.

1. Introduction

Mapping of scenes behind obstacles (including building wall, rubbers, grass, etc.) using through-wall imaging (TWI) is an unfolded research domain. Different military and commercial applications (including antiterrorism, hostage rescue and surveillance, etc. [1]) can benefit from TWI. Beside other challenges, minimization of unwanted artifacts (clutters/noise) has enjoyed special importance over last few years [2–13]. These unwanted artifacts significantly decrease target detection and recognition capabilities.

Existing TWI image enhancement (clutter removal) techniques include background scene subtraction (only feasible if with and without target images are available) [2], spatial filtering (assuming wall homogeneity at low frequencies) [3], wall modeling and subtraction (requiring complex process for inhomogeneous walls) [4, 5] Doppler filtering (applicable for moving targets only) [6], image fusion (requiring multiple data of the same scene) [7], and statistical techniques [8–13].

In this paper, a TWI image enhancement (clutter reduction) technique using QR decomposition (QRD) and fuzzy logic is presented (preliminary results presented in [13]). Weights are assigned to different QRD subspaces using fuzzy inference engine. Simulation results evaluated using mean square error (MSE), peak signal to noise ratio (PSNR), improvement factor (IF), and visual inspection (based on miss detection (MD) and false detection (FD)) are used to verify the proposed scheme.

2. Proposed Image Enhancement Using QRD

Let the input image M (having dimensions $G \times H$) be decomposed into different subspaces (M_{cl}, M_{tar}, and M_{no}) using singular value decomposition (SVD) as

$$M = \underbrace{\sum_{g=1}^{l_1} s_g u_g v_g^T}_{M_{\text{cl}}} + \underbrace{\sum_{g=l_1+1}^{l_2} s_g u_g v_g^T}_{M_{\text{tar}}} + \underbrace{\sum_{g=l_2+1}^{G} s_g u_g v_g^T}_{M_{\text{no}}}, \quad (1)$$

where U and V are singular vector matrices and S contains singular values. As discussed in [13], conventional SVD for TWI image enhancement assumes that the target is limited to the second spectral component only; that is,

$$M_{\text{SVD}} = s_2 u_2 v_2^T. \quad (2)$$

Besides the high computational complexity of SVD which is $4G^2H + 8GH^2 + 9H^3$ [14], the statement of target containment in the second spectral component is not always true. To cater the above issues, QRD and fuzzy logic based scheme is proposed. The image M can be decomposed into an orthogonal unitary matrix Q (having dimensions $G \times H$ and

Table 1: Comparison analysis of QRD algorithms for TWI.

Algorithms	Accuracy				Complexity	Stability
	$M - QR$	$Q^TQ - I$	$Q^TM - R$	$MR^{-1} - Q$		
CGS QR	1.091×10^{-16}	1.310×10^{-10}	6.724×10^{-14}	9.913×10^{-14}	$2GH^2$	Unstable
MGS QR	1.075×10^{-16}	4.922×10^{-13}	4.842×10^{-14}	1.083×10^{-12}	$2GH^2$	Stable
HT QR	1.291×10^{-15}	3.795×10^{-15}	3.333×10^{-16}	1.263×10^{-12}	$4G^2H + 2GH^2 + \frac{2}{3}H^3$	Stable
Givens QR	7.532×10^{-16}	6.702×10^{-15}	2.711×10^{-16}	1.118×10^{-12}	$8G^2H + 2GH^2 + \frac{2}{3}H^3$	Stable

column vectors q_g) and an upper triangular matrix R (having dimensions $G \times H$ and row vectors r_g), that is,

$$M = QR. \tag{3}$$

Table 1 shows the accuracy, stability, and complexity analysis of different QRD algorithms (classical and modified Gram-Schmidt (CGS, MGS), Givens decomposition, Householder transformation (HT), etc. [14]) for TWI.

Identical to SVD, the first subspace $M_1 = q_1 r_1$ represents wall clutters and rest subspaces contain targets and noise. Note that due to overlapping boundaries of targets and noise, it is difficult to extract target subspaces accurately. Foregoing in view, a weighting QRD based scheme is proposed to enhance targets. The enhanced image M_{tar} is

$$M_{\text{tar}} = \sum_{g=2}^{G} w_g q_g r_g, \tag{4}$$

where w_g are weights applied to different subspaces. Fuzzy logic is used for the automatic weight assignment [15].

2.1. Input and Output MFs. Let $\xi_g = \|r_g\|$ and $\Delta\xi_g = \|r_g\| - \|r_{g+1}\|$ be norms and norm differences, respectively. Note that high value of ξ_g and $\Delta\xi_g$ the corresponding subspace $q_g r_g$ more likely contains target(s) and is therefore enhanced by applying heavy weights (and vise versa).

Three Gaussian membership functions (MFs) $\zeta_{X^x}(x_1) = \exp(-((c_1 - \bar{c}_1^{(x)})/\sigma_1^{(x)})^2)$ and $(x \in \{\text{High, Medium, Low}\})$ are defined for ξ_g. Similarly $\zeta_{Y^y}(x_2) = \exp(-((c_2 - \bar{c}_2^{(y)})/\sigma_2^{(y)})^2)$ and $(y \in \{\text{High, Medium, Low}\})$ are defined for $\Delta\xi_g$, where $\{c_1, c_2\} \in [0, 1]$, $\bar{c}_1^{(x)}$, $\bar{c}_2^{(y)}$ and $\sigma_1^{(x)}$, $\sigma_2^{(y)}$ are means and variances of fuzzy sets, respectively.

K-means algorithm [16] is used to adjust the fuzzy parameters. ξ_k and $\Delta\xi_h$ are first clustered into three groups based on respective histograms. The means and variances, respectively, of each group are used as centers $\bar{c}_1^{(x)}$, $\bar{c}_2^{(y)}$ and spreads $\sigma_1^{(x)}$, $\sigma_2^{(y)}$ of MFs. Five equally spaced output MFs $\zeta_{Z^z}(d) = -((d - \bar{d}^{(z)})/\varrho^{(z)})^2$ $(z \in \{\text{Very High, High, Medium, Low, Very Low}\})$, where mean $\bar{d}^{(z)}$ and variance $\varrho^{(z)}$ are used.

2.2. Product Inference Engine (PIE). Gaussian fuzzifier maps the input ξ_g and $\Delta\xi_g$ as

$$\zeta_{XY}(c_1, c_2) = \exp\left\{-\left(\frac{c_1 - \xi_g}{\nu_1}\right)^2\right\} \exp\left\{-\left(\frac{c_2 - \Delta\xi_h}{\nu_2}\right)^2\right\}, \tag{5}$$

where ν_1 and ν_2 are parameters used for input noise suppression and are chosen as $\nu_1 = 2\max_x \sigma_1^{(x)}$ and $\nu_2 = 2\max_y^3 \sigma_2^{(y)}$ [15].

Fuzzy IF-THEN rules for image enhancement are the following.

Rule 1: IF ξ_h is X^{High} and $\Delta\xi_h$ is Y^{High}, THEN w_h^{PIE} is $Z^{\text{Very High}}$.

Rule 2: IF ξ_h is X^{Med} and $\Delta\xi_h$ is Y^{High}, THEN w_h^{PIE} is Z^{High}.

Rule 3: IF ξ_h is X^{High} and $\Delta\xi_h$ is Y^{Med}, THEN w_h^{PIE} is Z^{High}.

Rule 4: IF ξ_h is X^{Med} and $\Delta\xi_h$ is Y^{Med}, THEN w_h^{PIE} is Z^{Med}.

Rule 5: IF ξ_h is X^{High} and $\Delta\xi_h$ is Y^{Low}, THEN w_h^{PIE} is Z^{Med}.

Rule 6: IF ξ_h is X^{Low} and $\Delta\xi_h$ is Y^{Med}, THEN w_h^{PIE} is Z^{Med}.

Rule 7: IF ξ_h is X^{Med} and $\Delta\xi_h$ is Y^{Low}, THEN w_h^{PIE} is Z^{Low}.

Rule 8: IF ξ_h is X^{Low} and $\Delta\xi_h$ is Y^{High}, THEN w_h^{PIE} is Z^{Low}.

Rule 9: IF ξ_h is X^{Low} and $\Delta\xi_h$ is Y^{Low}, THEN w_h^{PIE} is $Z^{\text{Very Low}}$.

The output of PIE using individual rule based inference, Mamdani implication, algebraic product for t-norm, and max operator for s-norm [15] is

$$\zeta_{Z'}(d_g) = \max_{\{x,y,z\}} \left[\sup_{\{c_1,c_2\}} \zeta_{XY}(c_1, c_2)\, \zeta_{X^x}(c_1)\, \zeta_{Y^y}(c_2)\, \zeta_{Z^z}(d_g) \right]. \tag{6}$$

The weights w_g^{PIE} are then computed as

$$w_g^{\mathrm{PIE}} = \frac{\sum_z \overline{d}^{(z)} \varpi_g^{(z)}}{\sum_z = 1 \varpi_h^{(z)}}, \quad (7)$$

where $\varpi_g^{(z)}$ is the height of $\zeta_{Z'}(d_g)$ in output MFs [15].

2.3. Takagi-Sugeno (TS) Inference. In contrast to PIE, TS inference engine adjusts the output MFs using adaptive and/or optimization techniques [17]. The TS rule-base (IF-THEN) for computing weights w_g^{TS} is

$$\text{IF } \xi_g \text{ is } X^{j_1} \text{ AND } \Delta\xi_g \text{ is } Y^{j_2}$$
$$\text{THEN } p^{(j_1+j_2-1)} = \left(\frac{1}{1 + \exp\left\{-\xi_g\right\} + \exp\left\{-\Delta\xi_g\right\}} \right)^{j_1+j_2-1}. \quad (8)$$

Note that the output reduces for large $j_1 + j_2$ (which is desirable). The aggregated weights w_g^{TS} are

$$w_g^{\mathrm{TS}} = \frac{\sum_{j_1=1}^{3} \sum_{j_2=1}^{3} p^{(j_1+j_2-1)} t\left\{\zeta_{X^{j_1}}(\xi_k), \zeta_{Y^{j_2}}\left(\Delta\xi_g\right)\right\}}{\sum_{j_1=1}^{3} \sum_{j_2=1}^{3} t\left\{\zeta_{X^{j_1}}\left(\xi_g\right), \zeta_{Y^{j_2}}\left(\Delta\xi_g\right)\right\}}, \quad (9)$$

where t represents algebraic product (intersection operator).

3. Simulation and Results

Experimental setup for TWI is constructed using Agilent's vector network analyzer (VNA) which generates stepped frequency waveforms between 2 GHz and 3 GHz (1 GHz band width (BW)) having step size of $\Delta f = 5$ MHz and step size $N_f = 201$. Maximum range is $R_{\max} = 30$ m and range resolution is $\Delta R = 0.15$ m.

Broadband horn antenna which is mounted on two-dimensional scanning frame (having dimensions 2.4 m × 3 m (width × height) and can slide along cross range and height) operates in monostatic mode with 12 dB gain. Thickness of the wall is 5 cm and relative permittivity and permeability are 2.3 and 1, respectively. The frame is placed 0.03 m away from wall and scanning is controlled by a microcontroller based mechanism. The scattering parameters are recorded at each step and transferred to a local computer for image reconstruction and processing. Received data is converted into time domain and beamforming algorithm is used for image reconstruction. Existing and proposed image enhancement algorithms are simulated in MATLAB and quantitative analysis is performed using MSE, PSNR, IF, FD, MD, and visual inspection:

$$\mathrm{MSE} = \frac{1}{G \times H} \sum_{g=1}^{G} \sum_{h=1}^{H} \left(M_{bs}(g,h) - M_{\mathrm{tar}}(g,h)\right)^2,$$
$$\mathrm{PSNR\,(dB)} = 10 \log_{10} \frac{1}{\mathrm{MSE}}, \quad (10)$$
$$\mathrm{IF\,(dB)} = 10 \log_{10} \left[\frac{P_{M_{\mathrm{tar}},t} \times P_{M,c}}{P_{M,t} \times P_{M_{\mathrm{tar}},c}} \right],$$

where M_{bs} is a reference image obtained by the difference of image, with and without target. $P_{M_{\mathrm{tar}},t}$ and $P_{M_{\mathrm{tar}},c}$ are average pixel values of target and clutter in enhanced image, respectively. $P_{M,t}$ and $P_{M,c}$ are average pixel values of target and clutter in the original image, respectively.

MD is defined as "target was present in the original image, but was not detected in enhanced image." FD is defined as "target was not present in the original image, but was detected in enhanced image." For calculating FD and MD, a threshold is calculated using global thresholding algorithm [18].

Figure 1 shows the original B-scan containing two targets, the background subtracted reference and enhanced images, using existing SVD and proposed QRD based schemes. It can be observed that proposed schemes detect both targets whereas SVD based scheme is unable to locate both targets accurately.

Figure 2 shows another example containing three targets. The proposed scheme detects all targets and provides a better target to background ratio compared to SVD based scheme. It is further noted that the proposed TS inference based scheme provides better results compared to PIE.

Table 2 shows that proposed fuzzy QRD schemes are better (as compared to the SVD image enhancement scheme) in terms of MSE, PSNR, IF, MD, and FD.

Table 2: MSE, PSNR, IF, MD, and FD comparison.

Scenario	Scheme	MSE	PSNR	IF	MD	FD
Two targets	SVD [11]	0.2726	5.6442	8.1258	1	0
	Fuzzy QRD (PIE)	0.1970	7.0553	11.2587	0	0
	Fuzzy QRD (TS)	0.1726	7.6296	11.5870	0	0
Three targets	SVD [11]	0.2814	5.5068	7.1265	1	1
	Fuzzy QRD (PIE)	0.1933	7.1377	10.8715	0	0
	Fuzzy QRD (TS)	0.1824	7.3898	11.0127	0	0

4. Conclusion

QRD and fuzzy logic based image enhancement scheme is proposed for TWI. Compared with SVD, QRD provides less computational complexity. PIE and TS inference engines are used to assign weights to different QRD subspaces. Simulation results compared on visual and quantitative analysis show the significance of the proposed scheme.

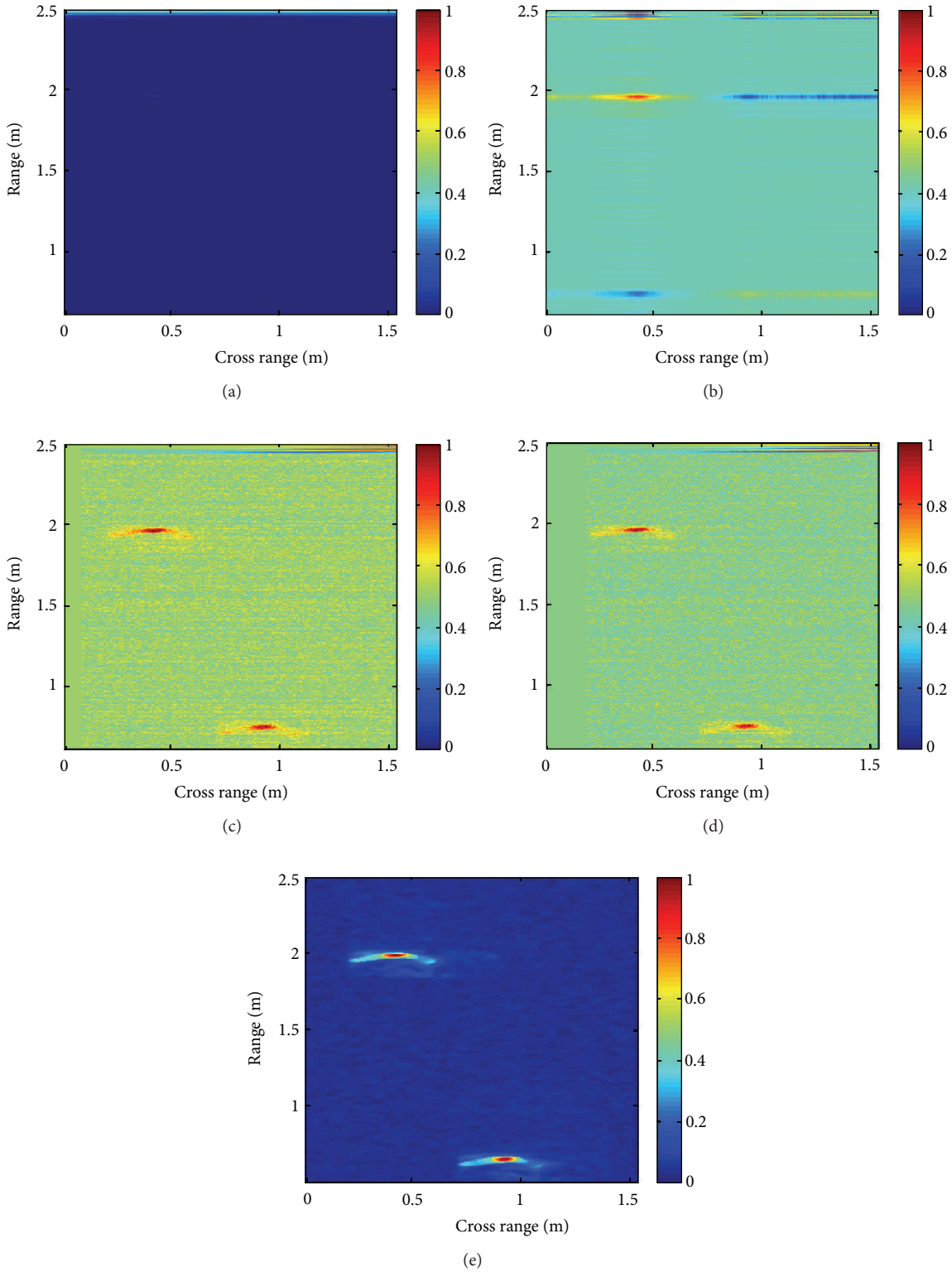


Figure 1: Image with two targets. (a) Original image. (b) SVD [11]. (c) Proposed fuzzy QRD (PIE). (d) Proposed fuzzy QRD (TS). (e) Background subtracted reference image.

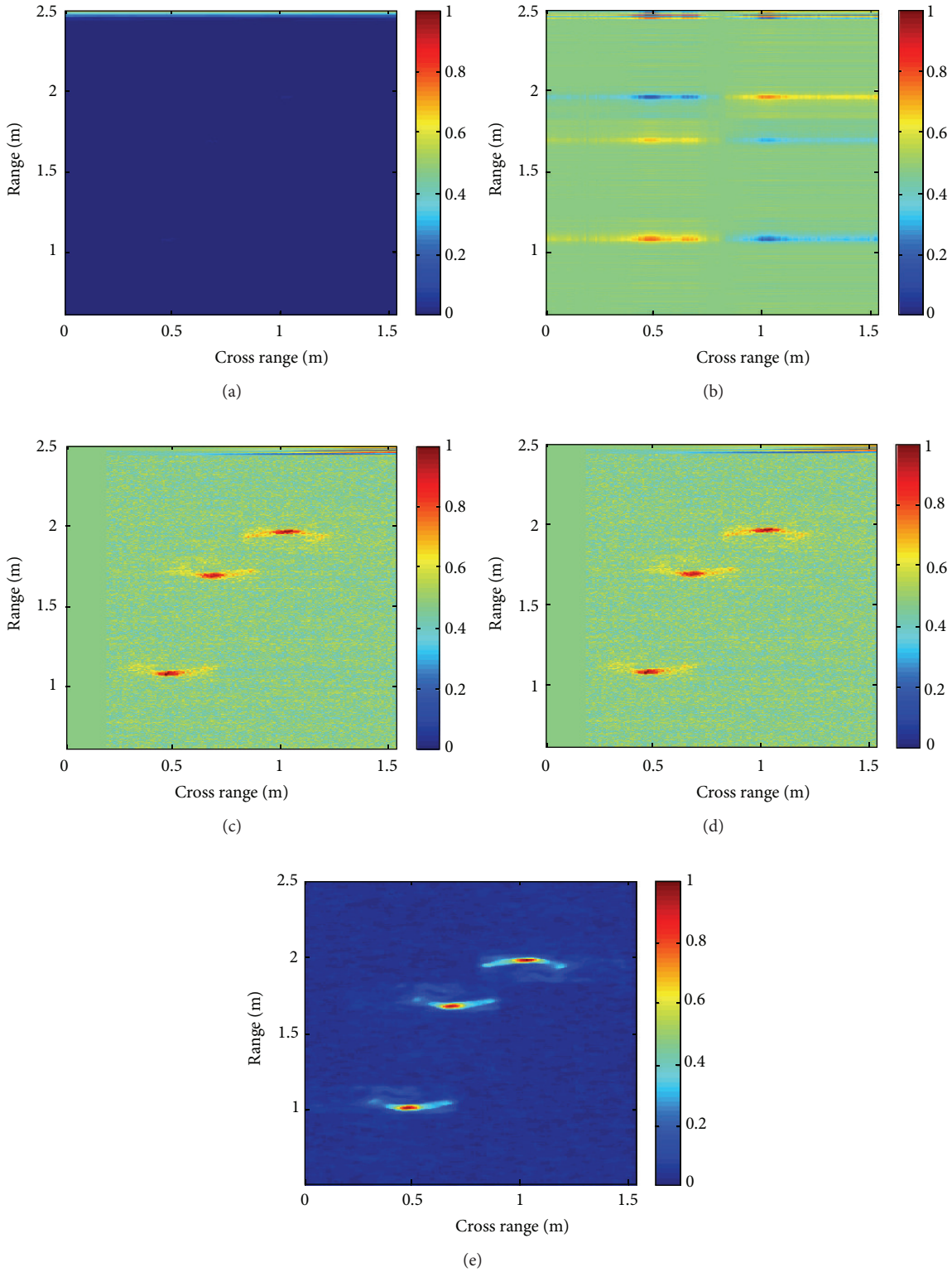


Figure 2: Image with three targets. (a) Original image. (b) SVD [11]. (c) Proposed fuzzy QRD (PIE). (d) Proposed fuzzy QRD (TS). (e) Background subtracted reference image.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. G. Amin, *Through the Wall Radar Imaging*, CRC Press, Boca Raton, Fla, USA, 2011.

[2] J. Moulton, S. Kassam, F. Ahmad, M. Amin, and K. Yemelyanov, "Target and change detection in synthetic aperture radar sensing of urban structures," in *Proceedings of the IEEE Radar Conference (RADAR '08)*, pp. 1–6, May 2008.

[3] Y. S. Yoon and M. G. Amin, "Spatial filtering for wall-clutter mitigation in through-the-wall radar imaging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 9, pp. 3192–3208, 2009.

[4] M. Dehmollaian and K. Sarabandi, "Refocusing through building walls using synthetic aperture radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1589–1599, 2008.

[5] G. E. Smith and B. G. Mobasseri, "Robust through-the-wall radar image classification using a target-model alignment procedure," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 754–767, 2012.

[6] S. S. Ram, C. Christianson, Y. Kim, and H. Ling, "Simulation and analysis of human micro-dopplers in through-wall environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 2015–2023, 2010.

[7] C. Debes, *Advances in detection and classification for through the wall radar imaging [Ph.D. Dissertation]*, Technische University Darmstadt, 2010.

[8] F. H. C. Tivive, M. G. Amin, and A. Bouzerdoum, "Wall clutter mitigation based on eigen-analysis in through-the-wall radar imaging," in *Proceedings of the 17th International Conference on Digital Signal Processing (DSP '11)*, pp. 1–8, July 2011.

[9] M. M. Riaz and A. Ghafoor, "Principle component analysis and fuzzy logic based through wall image enhancement," *Progress in Electromagnetic Research*, vol. 127, pp. 461–478, 2012.

[10] M. M. Riaz and A. Ghafoor, "Through wall image enhancement based on singular value decomposition," *International Journal of Antenna and Propagation*, vol. 2012, Article ID 961829, 20 pages, 2012.

[11] P. K. Verma, A. N. Gaikwad, D. Singh, and M. J. Nigam, "Analysis of clutter reduction techniques for through wall imaging in UWB range," *Progress in Electromagnetics Research B*, vol. 17, pp. 29–48, 2009.

[12] A. N. Gaikwad, D. Singh, and M. J. Nigam, "Application of clutter reduction techniques for detection of metallic and low dielectric target behind the brick wall by stepped frequency continuous wave radar in ultra-wideband range," *IET Radar, Sonar and Navigation*, vol. 5, no. 4, pp. 416–425, 2011.

[13] M. M. Riaz and A. Ghafoor, "QR decomposition based image enhancement for through wall imaging," in *Proceedings of the IEEE Radar Conference*, pp. 978–983, 2012.

[14] G. H. Golub and C. F. Loan, *Matrix Computation*, John Hopkins, Baltimore, Md, USA, 1998.

[15] L. X. Wang, *A Course in Fuzzy Systems and Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 1997.

[16] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithms: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[17] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[18] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 2007.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 582397, 13 pages
http://dx.doi.org/10.1155/2014/582397

Research Article

Incident Signal Power Comparison for Localization of Concurrent Multiple Acoustic Sources

Daniele Salvati[1] and Sergio Canazza[2]

[1] *Department of Mathematics and Computer Science, University of Udine, 33100 Udine, Italy*
[2] *Department of Information Engineering, University of Padova, 35131 Padova, Italy*

Correspondence should be addressed to Sergio Canazza; canazza@dei.unipd.it

Received 5 August 2013; Accepted 2 January 2014; Published 20 February 2014

Academic Editors: S. Bourennane and J. Marot

In this paper, a method to solve the localization of concurrent multiple acoustic sources in large open spaces is presented. The problem of the multisource localization in far-field conditions is to correctly associate the direction of arrival (DOA) estimated by a network array system to the same source. The use of systems implementing a Bayesian filter is a traditional approach to address the problem of localization in multisource acoustic scenario. However, in a real noisy open space the acoustic sources are often discontinuous with numerous short-duration events and thus the filtering methods may have difficulty to track the multiple sources. Incident signal power comparison (ISPC) is proposed to compute DOAs association. ISPC is based on identifying the incident signal power (ISP) of the sources on a microphone array using beamforming methods and comparing the ISP between different arrays using spectral distance (SD) measurement techniques. This method solves the ambiguities, due to the presence of simultaneous sources, by identifying sounds through a minimization of an error criterion on SD measures of DOA combinations. The experimental results were conducted in an outdoor real noisy environment and the ISPC performance is reported using different beamforming techniques and SD functions.

1. Introduction

The sensory capacity to analyze acoustic space is a very important function of an auditory system. The need for the development of an understanding of the sound environment has attracted many researchers over the past twenty years to build sensory systems that are capable of locating acoustic sources in space. Acoustic source localization (ASL) is an important task in a growing number of applications. Fields of application in which identification of the location of acoustic sources is desired include audio surveillance, teleconferencing systems, hands-free acquisition in car, system monitoring, human-machine interaction, musical control interfaces, videogames, virtual reality systems, voice recognition, fault analysis of machinery, autonomous robots, processors for digital hearing aids, high-quality recording, multiparty telecommunications, dictation systems, and acoustic scene analysis. The aim of an ASL system is to estimate the position of sound sources in space by analyzing the sound field with a microphone array, a set of microphones arranged to capture the spatial information of sound.

Several application areas that may potentially provide advantages in using the acoustic location have led to the development of many signal processing algorithms, which mostly consider the specific acoustic environment, the signal properties, and the localization goal.

ASL can be performed by two basic methods: indirect and direct. The indirect approach is used to estimate source positions by implementing the following two steps: in the first one, a set of time difference of arrivals (TDOAs) are estimated using measurements across various combinations of microphones, and in the second one, when the position of the sensors and the speed of sound are known, the source positions can be estimated using geometric considerations and approximate estimators: closed-formed estimators based on a least squares solution [1–7] (for an overview on closed-form estimators, see [8]) and iterative maximum likelihood estimators [9–15]. The direct approach involves the search

space by constructing a spatial energy map and estimating, for each possible point of interest, the values that maximize a specific likelihood function that provides a coherent value from the entire system of arrays. The position of the sources can be estimated directly and spatial likelihood functions can be defined [16–20].

In near-field conditions, since the sources radiate the sound in spherical waves, a hyperboloid describes all of the possible points of an acoustic source that generates the same TDOA to an array of two microphones. Indirect methods aim at estimating TDOAs for microphone pairs, typically using the generalized cross-correlation (GCC) [21] and the adaptive eigenvalue decomposition (AED) [22] based on the blind system identification, which focuses on the impulse responses between the source and the microphones. The extension of the AED in the case of multiple microphones was proposed in [23], and it is efficiently performed with a normalized multichannel frequency-domain least mean square algorithm [24, 25]. However, the steered response power (SRP) is a direct method based on maximizing the power output of a beamformer. Beamforming is a combination of the delayed signals from each microphone in a manner in which an expected pattern of radiation is preferentially observed. In general, the SRP is computed in frequency-domain using the fast Fourier transformer on a signal portion, calculating the response power on each frequency bin, and subsequently fusing these estimates to obtain the final result. The conventional SRP is performed with the delay and sum beamformer [26]; it consists of the synchronization of signals that steer the array in a certain direction, and it sums the signals to estimate the power of the spatial filter. The SRP phase transform (SRP-PHAT) [18] is a widely used filtered beamforming. PHAT filter [21] places equal importance on each frequency by dividing the spectrum by its magnitude. It normalizes the amplitude of the spectral density using only the phase information with the advantage to improve performance in case of moderate noise and reverberation. SRP-PHAT is deeply used due to the fact that it can be efficiently computed by coherent summing the GCC-PHAT from all of the microphone pairs for each possible point of interest. The high-resolution SRP has been developed to improve the performance of the spatial filter, and the adaptive beamformer is called the minimum variance distortionless response (MVDR) due to Capon [27]. The multiple signal classification (MUSIC) algorithm is based on an eigen subspace decomposition method [28, 29], and the estimation of signal parameters via rotational invariance techniques (ESPRIT) is based on subspace decomposition exploiting the rotational invariance [30–32].

In far-field conditions, we are no longer able to detect the spherical wavefront in relationship with the distance of source from an array and the size of the array, and the wavefront is approximate to a plane. In this condition, with an array of microphones, we are able to estimate only the direction of arrival (DOA) of the source but not its distance from the array. In the far-field case the hyperboloid, which is the locus of points that generates the same TDOA to a microphone pair, can be approximated with the cone whose vertex is located at the midpoint of the array. Thus, we need a network of arrays to perform the localization of a source (at least two arrays for two-dimensional space). Hence, the position estimation is computed by intersection of lines and by an approximated solution for overdetermined systems using the linear least squares method. In the case of an array containing M microphones ($M > 2$) the DOA estimation can be computed with the indirect method of multichannel cross-correlation coefficient (MCCC) [33, 34], which is based on TDOAs estimation using GCC, and on the use of the spatial prediction error to measure the correlation among multiple signals. It has the advantage of using the redundant information between microphones to estimate the DOA in a more robust manner under a reverberant and noisy condition. The family of SRP direct methods with an array is used to estimate the DOAs of sources by picking the values corresponding to the principal peaks of the steered response power of a beamforming.

Recently, more sophisticated algorithms have been proposed for time delay estimation that use minimum entropy [35, 36] and blind source separation [37–39]. In [39], the authors demonstrate that the broadband independent component analysis methods are more robust against high background noise levels compared with the conventional GCC-PHAT approach.

Both indirect and direct methods have been tested in many single source scenarios; however, in multiple sources cases, they require new considerations. Several works address the problem of multiple sources using a Bayesian approach based on the tracking of the sources and using Kalman filter [40–47] and Particle filter [19, 48–53]. Some studies consider an approach without tracking in reverberant environments [39, 54–57].

In a real open space, the traditional techniques based on Bayesian filters (Kalman and Particle filters) are difficult to apply for localization of concurrent multiple acoustic sources, because sources are often discontinuous with numerous short-duration events and the spatial resolution may be poor in some areas of analysis. Note that in practical applications the localization in a open space needs a reduced number of arrays, due to limited space for installing it and not to invade the monitoring spaces in an excessive way. Besides, methods based on movement tracking can fail in some specific situations: during the initialization phase of the filter, in the presence of sources with unpredictable trajectory (e.g., in the case of rapid changes of the velocity vector), and when two sources have intersecting trajectories.

As a solution to this problem, we present the approach based on the incident signal power comparison (ISPC). A preliminary work was proposed in [58, 59]. This paper describes a detailed step-by-step ISPC algorithm introducing a diagonal loading (DL) [60, 61] for MVDR beamforming, which gives more stable ISP estimation, and reporting new experimental results in a real scenario. ISPC is designed for a distributed array system, and it is based on source extraction and on a verification of similarity among sound sources. The first step consists of source extraction using beamforming techniques and estimation of the incident signal power (ISP) of every source captured on the array. The second step involves the comparison of the ISP spectrum from different arrays using a spectral distance (SD) measure. The ISP

spectrum permits identification of sounds so that the spectrum power distance minimizes an error criterion. Therefore, the identification of the correct combination of DOAs is estimated by identifying the minor value of SD measures.

The location in a free-field outdoor environment can be employed for audio surveillance, sound monitoring, and analysis of acoustic scenes. In particular, Section 5 describes a prototype system for multiple source localization in a public space for monitoring a large area with a joint audio-video system, in which the positional estimates by acoustic analysis are used to steer a video-camera consequently.

The paper is organized as follows. After presenting the signal model in Section 2, the multiple sources localization problem is described in Section 3. In Section 4 the ISPC algorithm is presented. Finally, Section 5 illustrates experimental results obtained in a real-world scenario.

2. Signal Model

We assume N acoustic sources and R arrays, each composed of M microphones, and consider the omnidirectional characteristics of both the sources and the microphones. We will refer to the model of discrete-time obtained by performing a sampling operation on the continuous-time signal $x(t)$ with a uniform sampling period T_s. A discrete-time signal is expressed by

$$x\left(kT_S\right) = x\left(\frac{k}{f_s}\right) \quad k = 0, 1, \ldots, \tag{1}$$

where k is the sample time index and f_s is the sampling frequency. As usual, we will allow the sample period T_s to remain implicit and refer to it simply as $x(k)$.

The free-field discrete-time signal received by the mth microphone of the rth array can be modeled as

$$x_{rm}(k) = \sum_{n=1}^{N} \alpha_{rnm} s_n\left(k - k_{rn} - \tau_{rnm}\right) + v_{rm}(k), \tag{2}$$

where α_{rnm} is the attenuation of the sound propagation (inversely proportional to the distance from source n to microphone m of array r), $s_n(k)$ are the unknown uncorrelated source signals, k_{rn} is the propagation time from the unknown source n to the reference sensor of array r, τ_{rnm} is the TDOA of the signal n between the mth microphone and the reference of the rth array, and $v_{rm}(k)$ is the additive noise signal at the sensor m of array r, assumed to be uncorrelated with not only all of the source signals but also with the noise observed at the other sensors.

In far-field case the relationship between TDOA and DOA can be solved easily with geometrical considerations. Therefore, for a generic pair of microphones with TDOA τ_{rn}, DOA estimate is obtained as

$$\theta_{rn} = \arcsin\left(\frac{\tau_{rn} c}{d}\right), \tag{3}$$

where c is the speed of sound and d the distance between microphones.

The vector $\boldsymbol{\Theta}_n$ for each source n, considering the signal model (2), is defined by

$$\boldsymbol{\Theta}_n = \left[\theta_{1n}, \theta_{2n}, \ldots, \theta_{Rn}\right]^T \tag{4}$$

which contains the DOAs of the acoustic source n by each array. In the case of N sources and R arrays, we can write the matrix $R \times N$, which contains all DOAs of distributed array network as

$$\boldsymbol{\Theta} = \left[\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_N\right] = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1N} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{R1} & \theta_{R2} & \cdots & \theta_{RN} \end{bmatrix}. \tag{5}$$

The estimated DOAs angles, obtained for each array r, are written with the following vector:

$$\underline{\widehat{\boldsymbol{\Theta}}}_r = \left[\widehat{\theta}_{r1}, \widehat{\theta}_{r2}, \ldots, \widehat{\theta}_{rN}\right], \tag{6}$$

where we consider the DOA values in ascending order ($\widehat{\theta}_{r1} < \widehat{\theta}_{r2} < \widehat{\theta}_{r3}$, etc.). Next, the estimated sorted DOAs matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ is defined as

$$\underline{\widehat{\boldsymbol{\Theta}}} = \begin{bmatrix} \underline{\widehat{\boldsymbol{\Theta}}}_1 \\ \underline{\widehat{\boldsymbol{\Theta}}}_2 \\ \vdots \\ \underline{\widehat{\boldsymbol{\Theta}}}_R \end{bmatrix} = \begin{bmatrix} \widehat{\theta}_{11} & \widehat{\theta}_{12} & \cdots & \widehat{\theta}_{1N} \\ \widehat{\theta}_{21} & \widehat{\theta}_{22} & \cdots & \widehat{\theta}_{2N} \\ \vdots & \vdots & \ddots & \ddots \\ \widehat{\theta}_{R1} & \widehat{\theta}_{R2} & \cdots & \widehat{\theta}_{RN} \end{bmatrix}. \tag{7}$$

The position of the source n can be calculated by combining the DOAs estimated by the R arrays for that source.

3. Multiple Sources Localization

The multiple sources localization problem is to correctly assign the R DOAs values to the source n. In some applications, situations arise for which we cannot assign unambiguously TDOAs or DOAs to the same source. The example in Figure 1 shows the case of two sources with a configuration of two arrays for the 2D location. As we can see, the combination of incorrect DOAs leads to an incorrect position estimation. The two DOAs calculated by the two arrays can be combined following two different configurations: (1) $\widehat{\theta}_{11} - \widehat{\theta}_{21}, \widehat{\theta}_{12} - \widehat{\theta}_{22}$; (2) $\widehat{\theta}_{12} - \widehat{\theta}_{21}, \widehat{\theta}_{11} - \widehat{\theta}_{22}$. The first configuration implies the correct localization of the sound sources, whereas the second leads to an incorrect localization of both the sources.

In general, the goal is to get the matrix $\boldsymbol{\Theta}$ to properly order the values of (7). Considering θ_{rn} as the nth DOA of array r, the assignment of the correct value of the DOA for the unknown sources can be ambiguous; namely the exact position of the elements in the matrix of (6) cannot be uniquely determined:

$$\widehat{\theta}_{rn} \longrightarrow \theta_{rn}. \tag{8}$$

The possible combinations of the DOAs of matrix (7) are $O = (N!)^{(R-1)}$.

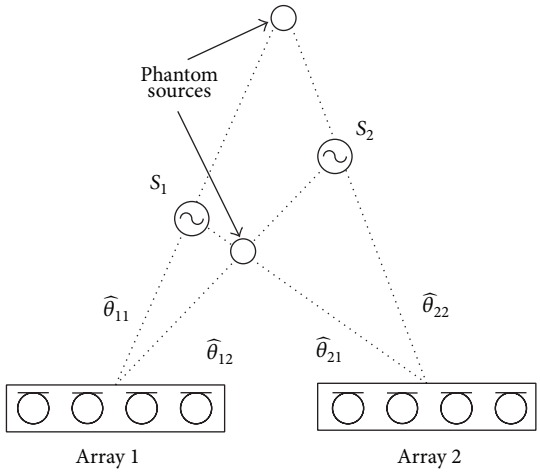


Figure 1: The problem of multiple sources localization.

4. Incident Signal Power Comparison (ISPC)

Incident signal power comparison (ISPC) combines the DOAs from different arrays by considering the similarity criterion among acoustic sources. To check for this similarity, we can estimate for each array the ISP referring to all estimated DOAs using beamforming techniques. Once the ISPs are obtained, we can define an efficient error criterion for comparing the different possible combinations of the DOAs using a SD measure of ISPs pair between different arrays.

DOA estimation is a crucial step of ASL systems. In a free-field environment for far-field cases, it can be calculated by means of MCCC and SRP methods. After obtaining an estimation of the sorted DOAs matrix (the matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ of (7)), the steps of the ISPC algorithm are (1) source extraction using beamforming techniques and estimation of ISPs for each DOA, (2) ISPC using SD measurement between ISPs of different array, (3) calculation of all DOAs combinations, and (4) verification of the most consistent target combinations minimizing an error criterion on SD measurements. Finally, the localization of multiple sources can be computed by considering the estimated DOAs combination. Figure 2 illustrates the ISPC steps.

4.1. Incident Signal Power Estimation. The ISP is the power spectral density of the beamformer output that is steered to a specified direction. The SRP is based on maximizing the power output of a beamformer. Beamforming is a multichannel signal processing techniques that enhance the acoustic signals coming from a specific steered position, while reducing the signals coming from other directions. In the frequency domain, the output of a generic beamformer of rth array in matrix notation can be written as

$$Y_r(f) = \mathbf{W}^H(f,\theta_{rn})\,\mathbf{X}_r(f), \tag{9}$$

where $\mathbf{X}_r(f) = [X_{r1}(f), X_{r2}(f), \ldots, X_{rM}(f)]^T$, $Y_r(f)$ and $X_{rm}(f)$ are the discrete Fourier transform of the signals, $\mathbf{W}(f,\theta_{rn}) = [W_1(f,\theta_{rn}), W_2(f,\theta_{rn}), \ldots, W_M(f,\theta_{rn})]^T$ is the vector of the beamformer weights for steering and filtering the data on the direction θ_{rn}, f is the frequency bin index, and the superscript H represents the Hermitian (complex conjugate) transpose.

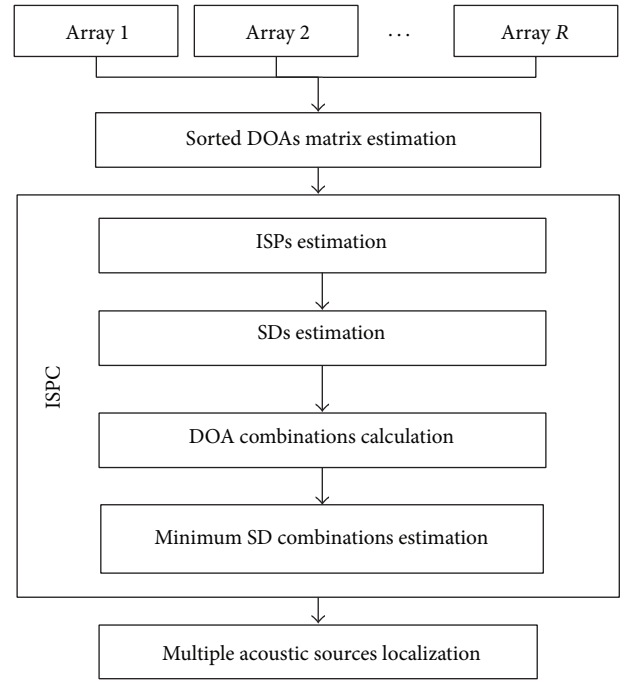


Figure 2: The steps for the ISPC algorithm.

The ISP of the beamformer output for a generic frequency f is given by

$$\begin{aligned}\mathrm{ISP}(f) &= E\left\{\left|Y_r(f)\right|^2\right\}\\ &= \mathbf{W}^H(f,\theta_{rn})\,E\left\{\mathbf{X}_r(f)\,\mathbf{X}_r^H(f)\right\}\mathbf{W}(f,\theta_{rn})\\ &= \mathbf{W}^H(f,\theta_{rn})\,\boldsymbol{\Phi}_r(f)\,\mathbf{W}(f,\theta_{rn}),\end{aligned} \tag{10}$$

where $\boldsymbol{\Phi}_r(f)$ is the cross-spectral density matrix, which is square $M\times M$ and symmetric, and $E\{\cdot\}$ denotes mathematical expectation. We consider a power spectrum calculated with $f = F_{\min}, F_{\min}+1, \ldots, F_{\max}$, where $F_{\min}$ and $F_{\max}$ are the index values of a specific frequency range (FR), which defines the range interesting for the optimal performance of the ISPC algorithm. Note that beamformer pattern function is frequency dependent; then the main lobe narrows with increasing frequency and spatial aliasing can occur (for a comprehensive dissertation, refer to [62]).

Several beamforming techniques exist (a review can be found in [63]); however, the spatial filter methods that are used for comparing ISPC experimental results are the SRP based on delay and sum beamforming, the SRP with the Dolph-Chebyshev window (SRP-DC), and the MVDR with DL.

Hence, the ISP corresponding to delay and sum SRP can be written from (10) as

$$\mathrm{ISP}_{rn}^{\mathrm{SRP}}(f) = \mathbf{A}^H(f,\theta_{rn})\,\boldsymbol{\Phi}_r(f)\,\mathbf{A}(f,\theta_{rn}), \tag{11}$$

where $\mathbf{A}(f,\theta_{rn})$ is the steering vector corresponding to direction θ_{rn}. For a uniform linear array with microphone distance d, the steering vector takes the form

$$\mathbf{A}(f,\theta_{rn}) = \left[1, e^{j2\pi(f-1)df_s\sin\theta_{rn}/c}, \ldots, e^{(j2\pi(f-1)df_s\sin\theta_{rn}/c)(M-1)}\right]^T. \tag{12}$$

The SRP-DC is obtained from (11) introducing the Dolph-Chebyshev window $\mathbf{w}$:

$$\mathrm{ISP}_{rn}^{\text{SRP-DC}}(f) = \left[\mathbf{w}\odot\mathbf{A}(f,\theta_{rn})\right]^H \mathbf{\Phi}_r(f)\left[\mathbf{w}\odot\mathbf{A}(f,\theta_{rn})\right], \tag{13}$$

where $\odot$ denotes element-wise multiplication.

The adaptive MVDR beamforming [27] is based on minimization problem of the following equation

$$\underset{\mathbf{W}(f,\theta_{rn})}{\operatorname{argmin}} \mathbf{W}^H(f,\theta_{rn})\,\mathbf{\Phi}_r(f)\,\mathbf{W}(f,\theta_{rn}) \quad \text{subject to} \quad \mathbf{W}^H(f,\theta_{rn})\,\mathbf{A}(f,\theta_{rn}) = 1. \tag{14}$$

The aim of the MVDR filter is to minimize the noise and sources coming from different directions, while keeping a fixed gain on the desired direction. Solving (14) using the method of Lagrange multipliers, we can write

$$\mathbf{W}_{\text{MVDR}}(f,\theta_{rn}) = \frac{\mathbf{\Phi}_r^{-1}(f)\,\mathbf{A}(f,\theta_{rn})}{\mathbf{A}^H(f,\theta_{rn})\,\mathbf{\Phi}_r^{-1}(f)\,\mathbf{A}(f,\theta_{rn})}. \tag{15}$$

In practical applications, the inverse of the cross-spectral density matrix can be calculated using the Moore-Penrose pseudoinverse, defined as

$$\mathbf{\Gamma}^{+} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^H, \tag{16}$$

where $\mathbf{\Gamma} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ is the singular value decomposition of the matrix $\mathbf{\Gamma}$. Moreover, if the cross-spectral density matrix is ill-conditioned, the spatial spectrum may not exist. Therefore, a DL [60, 61] method is adopted to calculate the inverse matrix in a stable way. The ISP with MVDR filter and DL becomes

$$\mathrm{ISP}_{rn}^{\text{MVDR}}(f) = \frac{1}{\mathbf{A}^H(f,\theta_{rn})\left(\mathbf{\Phi}_r(f)+\mu\mathbf{I}\right)^{+}\mathbf{A}(f,\theta_{rn})}, \tag{17}$$

where $\mathbf{I}$ is the identity matrix and μ is the loading level:

$$\mu = \frac{1}{L}\,\mathrm{tr}\left\{\mathbf{\Phi}_r(f)\right\}\Delta, \tag{18}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of the squared matrix and Δ is the normalized loading constant. Typically, the values are $\Delta = 0.1$, $\Delta = 1$, $\Delta = 10$ [64].

Therefore, we can define the matrix $\mathbf{P}$ containing all the ISPs related to the matrix (7):

$$\mathbf{P} = \left[\mathbf{P}_{11}, \mathbf{P}_{12}, \ldots, \mathbf{P}_{1N}, \mathbf{P}_{21}, \mathbf{P}_{22}, \ldots, \mathbf{P}_{2N}, \mathbf{P}_{R1}, \mathbf{P}_{R2}, \ldots, \mathbf{P}_{RN}\right] \tag{19}$$

which has a dimension of $(F_{\max} - F_{\min}) \times RN$, where the total number of ISPs is $I = NR$ and $\mathbf{P}_{rn} = [\mathrm{ISP}_{rn}(F_{\min}), \mathrm{ISP}_{rn}(F_{\min}+1), \ldots, \mathrm{ISP}_{rn}(F_{\max})]^T$.

4.2. Spectral Distance Estimation. To compare the ISPs of different arrays, spectral distance (SD) functions are used. Distance measures produce measurements of the dissimilarity of two sound spectra. We define the SD estimation between the ISP_{rn} and the ISP_{ij} of two DOAs of different arrays as

$$E_{rnij} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\mathcal{S}\left\{\mathrm{ISP}_{rn}(f), \mathrm{ISP}_{ij}(f)\right\}\right|, \tag{20}$$

where $L = (F_{\max} - F_{\min} + 1)$, r and i are the index labels of the array, $r \neq i$, n and j are the index labels for the sorted DOAs of array, and $\mathcal{S}\{\mathrm{ISP}_{rn}(f), \mathrm{ISP}_{ij}(f)\}$ is the SD function to measure the dissimilarity of spectra. We consider the four most common SD functions to verify how our system performance varies as a function of different equations. A classic spectral estimation method is linear prediction (LP) [65], for which we insert a negative one to standardize the minimum to zero as all functions

$$E_{rnij}^{\text{LP}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - 1\right|. \tag{21}$$

The other functions are the Itakura-Saito (IS) distance measure [66]

$$E_{rnij}^{\text{IS}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - \log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - 1\right|, \tag{22}$$

the root mean square (RMS) log [67]

$$E_{rnij}^{\text{RMS}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left(\log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)}\right)^2, \tag{23}$$

and the COSH measure [68]

$$E_{rnij}^{\text{COSH}} = \frac{1}{L}\sum_{f=F_{\min}}^{F_{\max}} \left|\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} - \log\frac{\mathrm{ISP}_{rn}(f)}{\mathrm{ISP}_{ij}(f)} + \frac{\mathrm{ISP}_{ij}(f)}{\mathrm{ISP}_{rn}(f)} - \log\frac{\mathrm{ISP}_{ij}(f)}{\mathrm{ISP}_{rn}(f)} - 2\right|. \tag{24}$$

The total number of SD measures between all the ISPs pair of different arrays is $Q = N^2R(R-1)/2$.

4.3. DOA Combinations Calculation. Let us represent the sorted matrix of the DOAs using the graph theory to better understand the DOAs combinations calculation and the verification of the most consistent target combination minimizing an error criterion. Then, we can express the matrix (7) and all of its combinations as being composed of nodes and edges, connecting pairs of vertices. An example of three arrays and three sources is shown in Figure 3. Each row of the graph contains the sorted DOAs of an array: $\underline{\widehat{\mathbf{\Theta}}}_1 = [\widehat{\theta}_{11}, \widehat{\theta}_{12}, \widehat{\theta}_{13}]^T$, $\underline{\widehat{\mathbf{\Theta}}}_2 = [\widehat{\theta}_{21}, \widehat{\theta}_{22}, \widehat{\theta}_{23}]^T$, and $\underline{\widehat{\mathbf{\Theta}}}_3 = [\widehat{\theta}_{31}, \widehat{\theta}_{32}, \ldots, \widehat{\theta}_{3N}]^T$. Each DOA is a node of graph and the edges represent the possible connections between nodes with the values E_{rnij}, which is the estimated SD between the ISPs on array r of DOA i and on array n of DOA j. The combination of incorrect

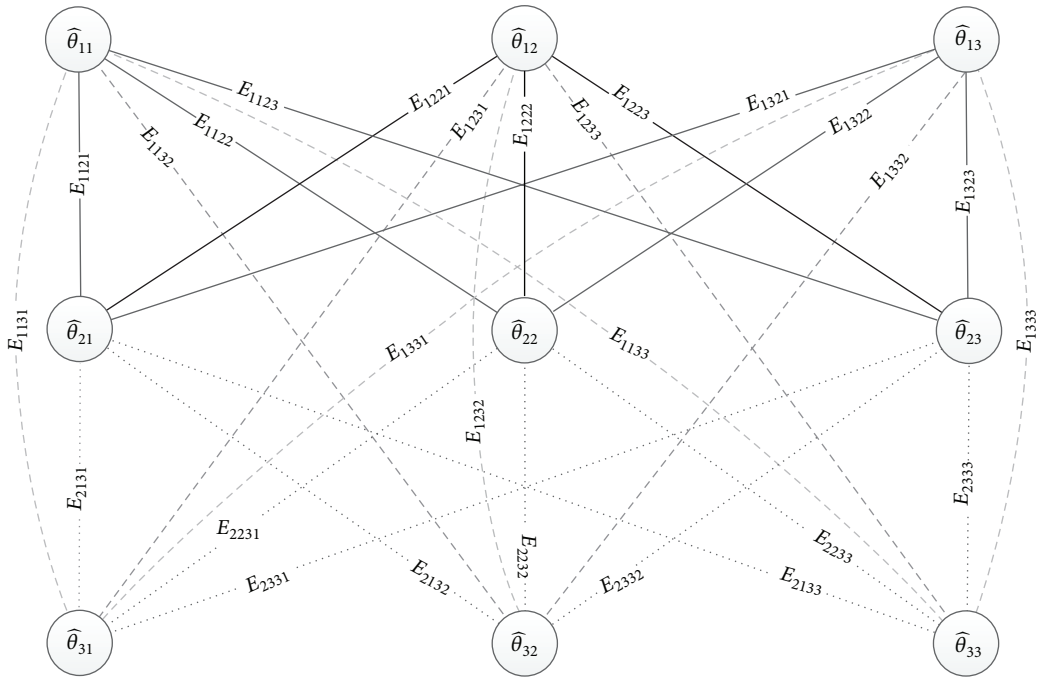


Figure 3: Graphic representation of DOAs and SDs estimations.

DOAs leads to an incorrect position estimation (see Figure 1). Thus, if we represent a combination of DOAs as a sum of values of the edges that connect the nodes, we expect that the minimum value of different sums corresponds to the correct combination. To calculate the possible combinations of DOAs between the arrays, it is helpful to introduce a matrix labeling of DOAs (7):

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1N} \\ B_{21} & B_{22} & \dots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{R1} & B_{R2} & \dots & B_{RN} \end{bmatrix} \tag{25}$$

in which the generic element is expressed as

$$B_{rn} = (r-1)\,N + n \tag{26}$$

with $r = 1, 2, \dots, R$ and $n = 1, 2, \dots, N$. The matrix label $\mathbf{B}$ associates the position of the DOAs referring to the sorted matrix $\underline{\widehat{\mathbf{\Theta}}}$. Estimating the minimum error of an SD combination, we can obtain the matrix $\widehat{\mathbf{\Theta}}$ with the correct position of the DOAs, in which each column contains the DOAs of the source n.

Furthermore, we can represent the graph representation of DOAs and the SDs as the adjacency matrix $\mathbf{\Lambda}$, which is an $RN \times RN$ matrix of SD values. The entry in row ($B_{rn} = 1, \dots, RN$) and column ($B_{ij} = 1, \dots, RN$) is defined as an SD estimation E_{rnij} if there is an edge connecting vertex B_{rn} and vertex B_{ij} in the graph, or it is defined as zero otherwise. The relationships between DOAs and SDs can be expressed by the following equation of the adjacency matrix element:

$$\Lambda_{B_{rn}B_{ij}} = E_{rnij}. \tag{27}$$

The symmetric adjacency matrix results in the following equation:

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & \dots & 0 & E_{1121} & \dots & E_{112N} & \dots & E_{11R1} & \dots & E_{11RN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & E_{1N21} & \dots & E_{1N2N} & \dots & E_{1NR1} & \dots & E_{1NRN} \\ E_{2111} & \dots & E_{211N} & 0 & \dots & 0 & \dots & E_{21R1} & \dots & E_{21RN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{2N11} & \dots & E_{2N1N} & 0 & \dots & 0 & \dots & E_{2NR1} & \dots & E_{2NRN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{R111} & \dots & E_{R11N} & E_{R121} & \dots & E_{1121} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_{RN11} & \dots & E_{RN1N} & E_{RN21} & \dots & E_{RN2N} & \dots & 0 & \dots & 0 \end{bmatrix}. \tag{28}$$

These SD values are weights of the edges of the graph. An example of three arrays and three sources is presented in Figure 3; in this example, we have 27 total SD comparisons (3 for each source).

To calculate all possible combinations of DOAs, we can work on the label matrix **B**. Considering that the first row of **B** related to the first array remains unchanged, we can compute the combinations in two steps. In the first step, the permutations of the N labels of **B** for each $R-1$ row (for each array) are calculated. The number of permutations for each row is $U = N!$. Thus, we define the permutation matrix T as

$$\begin{aligned}\mathbf{T} = \big[&\mathbf{T}_1, \mathbf{T}_{21}, \mathbf{T}_{22}, \ldots, \mathbf{T}_{2U}, \mathbf{T}_{31}, \mathbf{T}_{32}, \ldots, \mathbf{T}_{3U}, \ldots,\\ &\mathbf{T}_{R1}, \mathbf{T}_{R2}, \ldots, \mathbf{T}_{RU}\big],\end{aligned} \tag{29}$$

where

$$\mathbf{T}_1 = \left[B_{11}, B_{12}, \ldots, B_{1N}\right]^T, \tag{30}$$

$$\mathbf{T}_{ru} = \mathscr{P}_u\left\{\left[B_{r1}, B_{r2}, \ldots, B_{rN}\right]^T\right\}, \tag{31}$$

where $\mathbf{T}_{ru}$ is the vector of uth permutation $\mathscr{P}_u$ ($u = 1, \ldots, U$) of rth array ($r = 2, 3, \ldots, R$), which contains the N DOAs label of row r. The matrix **T** has a dimension of $N \times U(R-1)+1$. In the second step, the combinations of column indices of matrix **T** with value from 2 to RU give the $U^{(R-1)} = N!^{(R-1)} = O$ possible combinations. We consider the combinations of $R-1$ groups, each one composed by U elements (the permutation), assuming that one member (the index column of matrix **T**) from each of the $R-1$ groups is used in each combination and assuming that the order is not a distinguishing factor. We define a matrix **Z** of dimension $O \times (R-1)$, which stores the combinations of groups of column indices of matrix **T**:

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1(R-1)} \\ Z_{21} & Z_{22} & \cdots & Z_{2(R-1)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{O1} & Z_{O2} & \cdots & Z_{O(R-1)} \end{bmatrix}. \tag{32}$$

The generic element Z_{or} with $o = 1, 2, \ldots, O$ and $r = 1, 2, \ldots, R-1$ can be calculated with the following equations:

$$\begin{aligned} Z_{(o+i-1)r} &= U(r-1)+2, \quad i = 1, 2, \ldots, U_1,\\ Z_{(o+U_1+1)r} &= \begin{cases} Z_{(o+U_1+1)r}, & \text{if } Z_{(U_1+1)r} > U_2,\\ Z_{(o+U_1+1)r}+1, & \text{otherwise,}\end{cases}\end{aligned} \tag{33}$$

where $U_1 = U^{(r-1)}$ and $U_2 = U(r-1)+U+1$.

Hence, a combination label matrix **C** of $I \times O$ dimension is used to store the DOA label of all combinations:

$$\mathbf{C} = \left[\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_O\right], \tag{34}$$

where $\mathbf{C}_o$ is the vector, which contains the I DOA labels of combination o:

$$\begin{aligned}\mathbf{C}_o = \big[&B_{11}, T_{Z_{o1}1}, T_{Z_{o2}1} \ldots, T_{Z_{o(R-1)}1},\\ &B_{12}, T_{Z_{o1}2}, T_{Z_{o2}2}, \ldots, T_{Z_{o(R-1)}2}, \ldots,\\ &B_{1N}, T_{Z_{o1}N}, T_{Z_{o2}N}, \ldots, T_{Z_{o(R-1)}N}\big]^T\\ = &\left[C_{1o}, C_{2o}, \ldots, C_{Io}\right]^T.\end{aligned} \tag{35}$$

4.4. Minimum SD Measure Estimator. For each source, identified by R nodes (the arrays), we have $R(R-1)/2$ edges; then, the number of edges for a combination of DOAs is $G = NR(R-1)/2 = Q/N$. The values of matrix **C** are used to calculate the SD estimation of all combinations. Thus, we can define the SD estimation of the generic combination o as the sum of the weights of all the edges:

$$D_o = \sum_{n=1}^{N}\sum_{r=1}^{R-1}\sum_{i=1}^{R-r}\Lambda_{C_{o_1 o}C_{o_2 o}}, \tag{36}$$

where $o = 1, 2, \ldots, O$, $o_1 = (n-1)R+r$ and $o_2 = (n-1)R+r+i$. Accordingly, we define the SD vector of all combinations

$$\mathbf{D} = \left[D_1, D_2, \ldots, D_O\right]^T. \tag{37}$$

Finally, the index of the minimum value of the vector **D** identifies the target combination as

$$\hat{o} = \underset{o}{\operatorname{argmin}}\, D_o, \tag{38}$$

and the DOAs matrix $\widehat{\boldsymbol{\Theta}}$ is estimated by ordering the label matrix **B** with the combination $\mathbf{C}_{\hat{o}}$.

4.5. Overall Procedure. The processing steps of the full ISPC algorithm are summarized in Algorithm 1. After the DOAs estimation and creation of the matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ defined by (7) the ISPC algorithm is applied if multiple sources are detected. In practice, the matrix does not always present all the DOA values. In these cases, the missing value of array r can be represented with a zero value in the label matrix **B** (25). The overall procedure is composed by the following steps: (1) building of the label matrix **B** (25) and calculation of ISPs and the matrix **P** (19); (2) estimation of the SD measurements between all ISP pairs of arrays and creation of the adjacency matrix $\boldsymbol{\Lambda}$ (28); (3) calculation of the permutations matrix **T** (29) and the all DOA combination matrix **C** (34); (4) calculation of the vector **D** (37) that contains the SD estimation for each DOAs combination and finding the minimum value of **D** (38), for using the index value $\hat{o}$ in the matrix **C** to properly order the matrix $\underline{\widehat{\boldsymbol{\Theta}}}$ and estimate the matrix $\widehat{\boldsymbol{\Theta}}$.

5. Experimental Results

The experimental results were conducted in an outdoor real noisy environment and the ISPC performance is reported using different beamforming techniques (SRP, SRP-DC, MVDR) and SD estimations (LP, IS, RMS, COSH). A prototype system for two-dimensional localization has been

Require: $N > 1$
  $I = NR$ {ISPs}
  $O = N!^{(R-1)}$ {DOA combinations}
  $Q = N^2R(R-1)/2$ {SDs}
  $U = N!$ {DOA label permutations for an array}
  for $r = 1$ to R **do**
    for $n = 1$ to N **do**
      Calculate (26)
      Calculate (10) with (11) (13) (17)
    end for
  end for
  Calculate (25)
  Calculate (19)
  for $r = 1$ to $R - 1$ **do**
    for $n = r + 1$ to R **do**
      for $i = 1$ to N **do**
        for $j = 1$ to N **do**
          Calculate (20) with (21) (22) (23) (24)
          Calculate (27)
        end for
      end for
    end for
  end for
  Calculate (28)
  for $r = 2$ to R **do**
    for $u = 1$ to U **do**
      Calculate (31)
    end for
  end for
  Calculate (29)
  for $r = 1$ to $R - 1$ **do**
    while $o < O + 1$ **do**
      Calculate (33)
    end while
  end for
Calculate (32)
Calculate (34)
Calculate (36)
Calculate (38)

Algorithm 1: ISPC.

installed on the roof of the building that houses the Computer Science Department in Udine University (Figure 4).

5.1. System Setup. The acoustic localization prototype includes two linear arrays, each composed of four omnidirectional microphones. Very small sized arrays are used because a real application of such systems would require that the public spaces are not invaded in an excessive way; therefore, there might not be enough space to install the arrays. The arrays are located 11.4 m apart at a height of 12.1 m above the plane. The sample rate of the digital system is 48 kHz, and the microphone distance is 25 cm. The system consists of two parallel processing lines, corresponding to the Array 1 and Array 2 (Figure 5).

The first processing step is the DOAs estimation. SRP-PHAT is used for the DOAs estimation. The values corresponding to the principal N peaks of the SRP-PHAT function (in practice, those peaks which are above a given threshold)

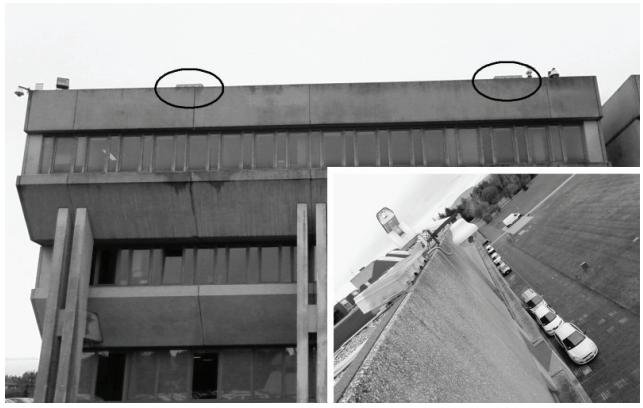


Figure 4: The prototype installed on the roof of the University building. The two arrays are encircled.

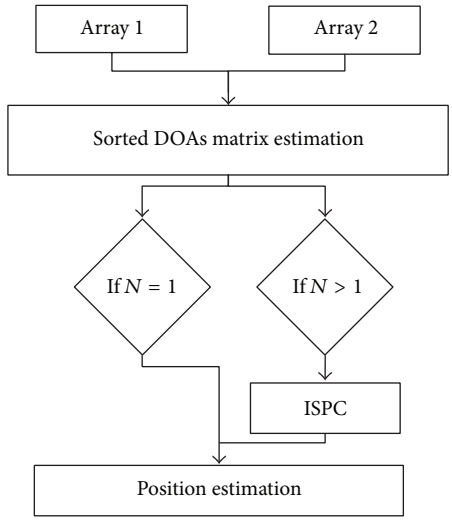


Figure 5: The block diagram of the processor showing the data flow of all of the tasks of the experimental prototype.

allow the DOAs estimation of the N acoustic sources. The assumed DOA range is -90° $+90^\circ$, where zero is in front of the array and the microphone reference is the first from left.

In the second step, the two-dimensional coordinates of the source can be estimated by combining the DOAs at the arrays. If more than one source is identified, a beamformer and an SD comparison provide a guide to solve the problem of associating the DOAs of the Array 1 with those of the Array 2. The calculation of the two-dimensional position of the source is a simple triangulation problem. However, we must consider that the two arrays are not coincidental with the plane of interest but are placed at a certain height. We must consider that the possible points identified by the DOA are located on a cone surface whose vertex is placed in the array and whose axis is the straight line joining the two arrays. Every array represents a cone: the intersection of the two cones is represented by a circumference. The intersection point between the circumference and the plane of interest is the estimation of the source distance from arrays (see Figure 6). Hence, we consider d_a to be the distance of the arrays, h to be the height

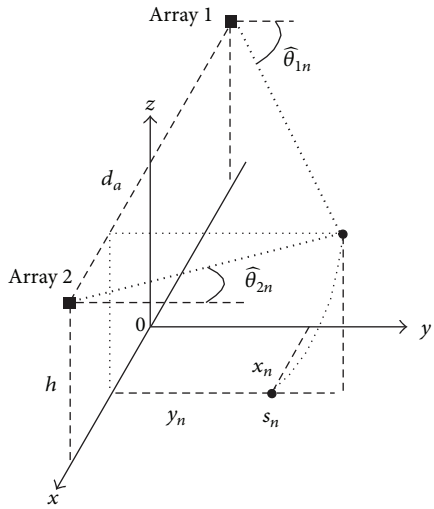


Figure 6: The two-dimensional position of the source of the experimental prototype.

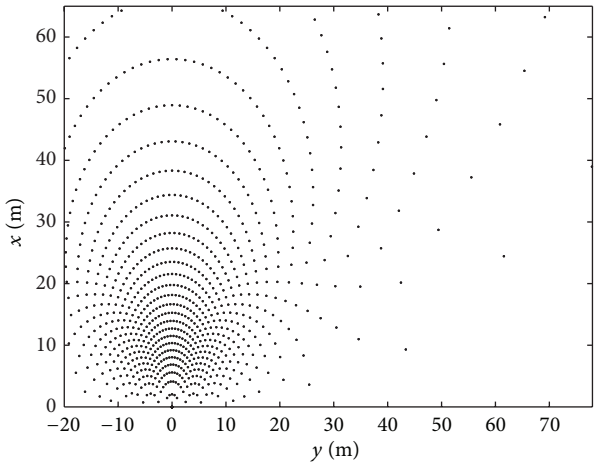


Figure 7: The x-y sample space position of the plane of interest.

of arrays above the plane of interest, and $\widehat{\theta}_{1n}$ and $\widehat{\theta}_{2n}$ to be the DOA estimated on the Array 1 and Array 2, and we obtain

$$x_n = \frac{d_a}{2}\left(\frac{\tan\widehat{\theta}_{2n} + \tan\widehat{\theta}_{1n}}{\tan\widehat{\theta}_{2n} - \tan\widehat{\theta}_{1n}}\right),$$
$$y_n = \sqrt{\left(\frac{d_a}{\tan\widehat{\theta}_{2n} - \tan\widehat{\theta}_{1n}}\right)^2 - h^2}. \quad (39)$$

The spatial resolution of the system depends on the distance between the microphones, the distance between the arrays, and the sample frequency of digital system. Figure 7 shows the possible xy coordinates of the considered area of analysis. The zero of the xy axes reference is located in the middle of the distance between the two arrays. The spatial resolution tends to decrease with an increasing distance from the arrays and an increasing angle from the axis perpendicular to the array.

Table 1: Position reporting the source label for each test and the SPL of the sources.

Test label	s_1 (Voice) 70 dB(A)	s_2 (Hammer) 100 dB(A)	s_3 (Motor car) 68 dB(A)	s_4 (Honk) 88 dB(A)
p_1	1	2	3	—
p_2	1	3	4	—
p_3	1	4	5	—
p_4	1	5	6	—
p_5	1	7	8	—
p_6	1	8	9	—
p_7	1	9	10	—
p_8	1	10	11	—
p_9	2	3	1	—
p_{10}	3	4	1	—
p_{11}	4	5	1	—
p_{12}	5	6	1	—
p_{13}	7	7	1	—
p_{14}	8	8	1	—
p_{15}	9	10	1	—
p_{16}	10	11	1	—
p_{17}	10	12	—	13
p_{18}	10	13	—	14
p_{19}	10	14	—	15
p_{20}	10	15	—	19
p_{21}	6	16	—	17
p_{22}	6	17	—	18
p_{23}	6	18	—	19
p_{24}	6	19	—	20
p_{25}	12	13	—	10
p_{26}	13	14	—	10
p_{27}	14	15	—	10
p_{28}	15	18	—	10
p_{29}	16	17	—	6
p_{30}	17	18	—	6
p_{31}	18	19	—	6
p_{32}	19	20	—	6

5.2. Experiment Setup. Experiments were conducted that consider the area of analysis of 60 × 90 m shown in Figure 8, that is, the parking lots of the University. Twenty zones of acoustic source positioning are considered. They are labeled with a number as we can see in Figure 8. The sources used are a human voice (s_1), a hammer striking an iron bar (s_2), a motor car (s_3), and a honk car (s_4). The hammer striking an iron bar and the honk car are short-duration event sounds.

Two types of experiments were performed. The first type used sounds with different spectral content, named Test 1. The second type, instead, used sounds with similar spectral content, named Test 2. Test 1 is composed of thirty-two parts ($p_1, p_2, \ldots, p_{32}$), each one with three sources placed in different positions (see Table 1). In various parts of Test 1, the sources were positioned at increasing distances along the y axis and the x axis. Table 1 also reported the sound pressure level (SPL) of each source. The environmental noise was in

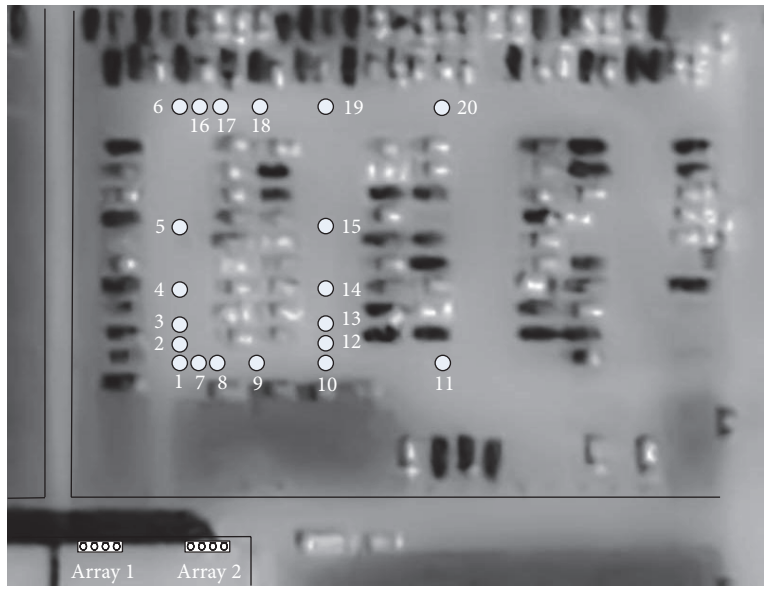


Figure 8: Map of the study area indicating the position of arrays (Array 1 and Array 2) and sources (the twenty labeling numbers: 1, 2, . . . , 20).

a range of 40–50 dB(A). In Test 2, two car sounds were used. The test was performed by placing two car sources in 1 and 7, as shown in Figure 8.

5.3. System Localization Evaluation. An evaluation of the system localization using a single source for each position was computed. Table 2 shows the real xy coordinates of the source points and the root mean square (RMS) error of the estimation using SRP-PHAT method. We can see that the estimation error increases in distant areas and when the angle of incidence on the array is large.

5.4. ISPC Evaluation. Tables 3 and 4 summarize the results for Test 1 and Test 2, respectively, comparing the localization success rate (as a percentage) with different beamforming algorithms (SRP, SRP-DC, and MVDR) and SD functions (LP, IS, RMS, and COSH).

The localization success rate is the ratio between the number of correct combinations and the number of ambiguities (NOA). NOA is the number of frames in which we have ambiguity to properly associate the DOAs to the sources; that is, the associations are incorrect in practice. The audio signal frame was divided into 17.5 ms overlapping and a Hann-windowed with a length of 140 ms. The parking area, where the tests were conducted, is a public area. Thus, we must consider that there are other sources in the acoustic scene: other sounds of cars that are moving in the parking area and in the nearby streets.

Table 3 summarizes the results of all thirty-two tests (Test 1). The number of NOA is 750, and the three frequency ranges (FR) for the SD estimation are 20–675 Hz, 20–2000 Hz, and 20–8000 Hz. The frequency value of 675 Hz takes into account the spatial aliasing limit, which, in our case, is $f = c/(2d) = 337/(2 \cdot 0.25) = 675$ Hz. The phenomena of spatial aliasing implies that the main lobe of the beamformer has a set of identical copies, called grating lobes. The appearance of grating lobes is a function of both microphone spacing and incident frequency. When fully visible, a grating lobe is equal in amplitude to the main lobe of the array. This fact reduces the array response, and, therefore, by defining the spatial sampling requirement and removing the grating lobes, we obtain a greater efficiency in the ISPC.

Table 4 depicts the results of Test 2 with an FR of 20–675 Hz and a NAM of 100. We can note that the accuracy decreases, especially with regard to the RMS and COSH

Table 2: Position referring to Figure 8 and the RMS errors of the localization estimation using SRP-PHAT method.

Source label	x (m)	y (m)	RMS error
1	1.5	20	1.1
2	1.5	23	1.7
3	1.5	26	2.2
4	1.5	32	1.9
5	1.5	38	2.5
6	1.5	52	4.1
7	4.5	20	0.9
8	7.5	20	4.2
9	10.5	20	8.6
10	20	20	18.2
11	30	20	8.8
12	20	23	9.2
13	20	26	15.8
14	20	32	15.7
15	20	38	4.8
16	4.5	52	4.2
17	7.5	52	8.3
18	10.5	52	17.5
19	20	52	20.2
20	30	52	23.6

Table 3: Results of Test 1 reporting the summary of the thirty-two tests ($p_1, p_2, \ldots, p_{32}$).

(Hz)	Localization success rate (%)											
FR	SRP-LP	SRP-IS	SRP-RMS	SRP-COSH	SRP-DC-LP	SRP-DC-IS	SRP-DC-RMS	SRP-DC-COSH	MVDR-LP	MVDR-IS	MVDR-RMS	MVDR-COSH
20–675	28.4	38.2	85.2	79.1	23.4	37.8	86.4	73.7	43.6	68.6	90.1	83.4
20–2000	42.1	45.2	72.3	63.5	43.8	42.4	71.2	61.1	45.4	63.1	77.6	72.1
20–8000	53.2	52.4	69.5	58.5	51.9	49.4	68.3	58.7	48.4	56.0	65.5	61.8

Table 4: Results of Test 2 using two car sounds.

(Hz)	Localization success rate (%)											
FR	SRP-LP	SRP-IS	SRP-RMS	SRP-COSH	SRP-DC-LP	SRP-DC-IS	SRP-DC-RMS	SRP-DC-COSH	MVDR-LP	MVDR-IS	MVDR-RMS	MVDR-COSH
20–675	50.2	53.8	57.5	47.1	52.4	51.3	59.4	55.7	45.4	62.0	52.0	53.5

functions, and this result highlights the limitation of the proposed approach in the case of spectrally similar sources.

The best results for Test 1 were obtained with the RMS log SD function and FR = [20–675] Hz. MVDR has the greatest capacity for location with 90.1% of successful DOAs association.

6. Conclusions

The novel incident signal power comparison algorithm is used to solve the ambiguous problem of correctly linking the DOAs from different arrays to the same source in a far-field condition with concurrent sources. Experimental results have shown that this approach can be a solution for a multisource localization that requires a frame-to-frame analysis, that is, in those cases in which the traditional filtering approach can not be applied. An evaluation of the system in a real scenario is reported, installing a hardware/software prototype on the roof of the University building and analyzing the results comparing three types of beamforming and four functions for the SD estimation. The interest in locating in a far-field outdoor context may be attractive for audio surveillance, sound monitoring, and the analysis of acoustic scenes. The ISPC is successfully used in a joint audio-video system for monitoring a large area. The best performances are obtained with RMS SD measure on frequency range between 20 Hz and the spatial aliasing frequency limit. We achieved a success rate of 90.1% using MVDR beamforming. We showed the limitation of the proposed algorithm in case of sources that have a similar spectral content.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors would like to thank Rudolf Rabenstein (Professor, University of Erlangen-Nurnberg, Germany) for his valuable advice and helpful comments.

References

[1] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 8, no. 6, pp. 821–835, 1972.

[2] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 8, pp. 1223–1225, 1987.

[3] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.

[4] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, 1994.

[5] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.

[6] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.

[7] M. D. Gillette and H. F. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, 2008.

[8] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 63–66, 2006.

[9] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.

[10] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 5, pp. 1210–1217, 1983.

[11] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and cramer-rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, 1990.

[12] M. Segal, E. Weinstein, and B. R. Musicus, "Estimate-maximize algorithms for multichannel time delay and signal estimation,"

IEEE Transactions on Signal Processing, vol. 39, no. 1, pp. 1–16, 1991.

[13] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[14] P. G. Georgiou and C. Kyriakakis, "Robust maximum likelihood source localization: the case for sub-gaussian versus gaussian," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1470–1480, 2006.

[15] G. Destino and G. Abreu, "On the maximum likelihood approach for source and network localization," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4954–4970, 2011.

[16] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.

[17] M. Omologo and P. S. R. DeMori, "Acoustic transduction," in *Spoken Dialogue with Computers*, Academic Press, London, UK, 1998.

[18] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing, pp. 157–180, Springer, Berlin, Germany, 2001.

[19] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.

[20] P. Pertilä, T. Korhonen, and A. Visa, "Measurement combination for acoustic source localization in a room environment," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, Article ID 278185, pp. 1–14, 2008.

[21] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[22] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *The Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–391, 2000.

[23] Y. Huang and J. Benesty, "Adaptive multichannel time delay estimation based on blind system identification for acoustic source localization," in *Adaptive Signal Processing: Applications to Real-World Problems*, Signals and Communication Technology, pp. 227–247, Springer, Berlin, Germany, 2003.

[24] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, 2003.

[25] D. Salvati and S. Canazza, "Adaptive time delay estimation using filter length constraints for source localization in reverberant acoustic environments," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 507–510, 2013.

[26] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, no. 4096, pp. 686–687, 1948.

[27] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proceedings of the RADC Spectrum Estimation Workshop*, pp. 243–258, Rome, NY, USA, October 1979.

[29] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[30] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.

[31] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1340–1342, 1986.

[32] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[33] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 549–557, 2003.

[34] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 509–519, 2004.

[35] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Processing Letters*, vol. 14, no. 3, pp. 157–160, 2007.

[36] F. Wen and Q. Wan, "Robust time delay estimation for speech signals using information theory: a comparison study," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2011, article 3, pp. 1–10, 2011.

[37] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 411–414, Paris, France, July 2003.

[38] B. Loesch, S. Uhlich, and B. Yang, "Multidimensional localization of multiple sound sources using frequency domain ICA and an extended state coherence transform," in *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing (SSP '09)*, pp. 677–680, Cardiff, UK, September 2009.

[39] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1490–1503, 2011.

[40] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking: a presentation general methodology," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 22–31, 2001.

[41] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video signal processing for object localization and tracking," in *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing, pp. 203–225, Springer, Berlin, Germany, 2001.

[42] D. Bechler, M. Grimm, and K. Kroschel, "Speaker tracking with a microphone array using kalman filtering," *Advances in Radio Science*, vol. 1, pp. 113–117, 2003.

[43] I. Potamitis, H. Chen, and G. Tremoulis, "Tracking of multiple moving speakers with multiple microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 520–529, 2004.

[44] U. Klee, T. Gehrig, and J. McDonough, "Kalman filters for time delay of arrival-based source localization," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 012378, pp. 1–15, 2006.

[45] S. Gannot and T. G. Dvorkind, "Microphone array speaker localizers using spatial-temporal information," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 059625, pp. 1–17, 2006.

[46] Z. Liang, X. Ma, and X. Dai, "Robust tracking of moving sound source using multiple model Kalman filter," *Applied Acoustics*, vol. 69, no. 12, pp. 1350–1355, 2008.

[47] C. Seguraa, A. Abad, J. Hernando, and C. Nadeu, "Multispeaker localization and tracking in intelligent environments," in *Multimodal Technologies for Perception of Humans*, vol. 4625 of *Lecture Notes in Computer Science*, pp. 82–90, Springer, Berlin, Germany, 2008.

[48] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Joint audio-visual tracking using particle filters," *EURASIP Journal on Applied Signal Processing*, vol. 2002, Article ID 162620, pp. 1154–1164, 2002.

[49] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources using particle filtering," in *Proceedings of the 14th European Signal Processing Conference*, pp. 1–4, Florence, Italy, September 2006.

[50] J.-M. Valin, V. F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," *Robotics and Autonomous Systems*, vol. 55, no. 3, pp. 216–228, 2007.

[51] F. Talantzis, A. Pnevmatikakis, and A. G. Constantinides, "Audio-visual active speaker tracking in cluttered indoors environments," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 38, no. 3, pp. 799–807, 2008.

[52] A. Quinlan, M. Kawamoto, Y. Matsusaka, H. Asoh, and F. Asano, "Tracking intermittently speaking multiple speakers using a particle filter," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2009, Article ID 673202, pp. 1–11, 2009.

[53] A. Levy, S. Gannot, and E. A. P. Habets, "Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, pp. 1540–1555, 2011.

[54] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1053–1056, Istanbul, Turkey, June 2000.

[55] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1479–1489, 2008.

[56] J.-S. Hu and C.-H. Yang, "Estimation of sound source number and directions under a multisource reverberant environment," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 870756, pp. 1–14, 2010.

[57] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, Article ID 147495, pp. 1–17, 2010.

[58] D. Salvati, A. Rodà, S. Canazza, and G. L. Foresti, "A real-time system for multiple acoustic sources localization based on ISP comparison," in *Proceedings of the 13th International Conference on Digital Audio Effects*, pp. 201–208, Graz, Austria, September 2010.

[59] D. Salvati, A. Rodà, S. Canazza, and G. L. Foresti, "Multiple acoustic sources localization using incident signal power comparison," in *Proceedings of the 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '11)*, pp. 77–82, Klagenfurt, Austria, September 2011.

[60] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[61] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.

[62] J. Dmochowski, J. Benesty, and S. Affès, "On spatial aliasing in microphone arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1383–1395, 2009.

[63] B. V. Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[64] Q.-H. Huang, Q. Zhong, and Q.-L. Zhuang, "Source localization with minimum variance distortionless response for spherical microphone arrays," *Journal of Shanghai University*, vol. 15, no. 1, pp. 21–25, 2011.

[65] J. Makhoul, "Linear prediction: a tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[66] R. J. McAulay, "Maximum likelihood spectral estimation and its application to narrow-band speech coding," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 243–251, 1984.

[67] L. L. Pfeifer, "Inverse filter for speaker identification ," RADC TR-74-214, Speech Communications Research Lab Inc, Santa Barbara, Calif, USA, 1974.

[68] J. A. H. Gray Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 380–391, 1976.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 695752, 7 pages
http://dx.doi.org/10.1155/2014/695752

Research Article

Improved Guided Image Fusion for Magnetic Resonance and Computed Tomography Imaging

Amina Jameel,[1] Abdul Ghafoor,[1] and Muhammad Mohsin Riaz[2]

[1] *Department of Electrical Engineering, College of Signals, National University of Sciences and Technology (NUST), Islamabad 46000, Pakistan*
[2] *Center for Advanced Studies in Telecommunication (CAST), COMSATS Institute of Information Technology, Islamabad, Pakistan*

Correspondence should be addressed to Abdul Ghafoor; abdulghafoor-mcs@nust.edu.pk

Received 6 August 2013; Accepted 15 December 2013; Published 13 February 2014

Academic Editors: S. Bourennane and J. Marot

Improved guided image fusion for magnetic resonance and computed tomography imaging is proposed. Existing guided filtering scheme uses Gaussian filter and two-level weight maps due to which the scheme has limited performance for images having noise. Different modifications in filter (based on linear minimum mean square error estimator) and weight maps (with different levels) are proposed to overcome these limitations. Simulation results based on visual and quantitative analysis show the significance of proposed scheme.

1. Introduction

Medical images from different modalities reflect different levels of information (tissues, bones, etc.). A single modality cannot provide comprehensive and accurate information [1, 2]. For instance, structural images obtained from magnetic resonance (MR) imaging, computed tomography (CT), and ultrasonography, and so forth, provide high-resolution and anatomical information [1, 3]. On the other hand, functional images obtained from position emission tomography (PET), single-photon emission computed tomography (SPECT), and functional MR imaging, and so forth, provide low-spatial resolution and functional information [3, 4]. More precisely, CT imaging provides better information on denser tissue with less distortion. MR images have more distortion but can provide information on soft tissue [5, 6]. For blood flow and flood activity analysis, PET is used which provide low space resolution. Therefore, combining anatomical and functional medical images through image fusion to extract much more useful information is desirable [5, 6]. Fusion of CT/MR images combines anatomical and physiological characteristics of human body. Similarly fusion of PET/CT is helpful for tumor activity analysis [7].

Image fusion is performed on pixels, features, and decision levels [8–10]. Pixel-level methods fuse at each pixel and hence reserve most of the information [11]. Feature-level methods extract features from source images (such as edges or regions) and combine them into a single concatenated feature vector [12, 13]. Decision-level fusion [11, 14] comprises sensor information fusion, after the image has been processed by each sensor and some useful information has been extracted out of it.

Pixel-level methods include addition, subtraction, division, multiplication, minimum, maximum, median, and rank as well as more complicated operators like Markov random field and expectation-maximization algorithm [15]. Besides these, pixel level also includes statistical methods (principal component analysis (PCA), linear discriminant analysis, independent component analysis, canonical correlation analysis, and nonnegative matrix factorization). Multiscale transforms like pyramids and wavelets are also types of pixel-level fusion [11, 14]. Feature-level methods include feature based PCA [12, 13], segment fusion [13], edge fusion [13], and contour fusion [16]. They are usually robust to noise and misregistration. Weighted decision methods (voting techniques) [17], classical inference [17], Bayesian inference [17], and

Dempster-Shafer method [17] are examples of decision-level fusion methods. These methods are application dependent; hence, they cannot be used generally [18].

Multiscale decomposition based medical image fusion decompose the input images into different levels. These include pyramid decomposition (Laplacian [19], morphological [20], and gradient [21]); discrete wavelet transform [22]; stationary wavelet transform [23]; redundant wavelet transform [24]; and dual-tree complex wavelet transform [25]. These schemes produce blocking effects because the decomposition process is not accompanied by any spatial orientation selectivity.

To overcome the limitations, multiscale geometric analysis methods were introduced for medical image fusion. Curvelet transform based fusion of CT and MR images [26] does not provide a proper multiresolution representation of the geometry (as curvelet transform is not built directly in the discrete domain) [27]. Contourlet transform based fusion improves the contrast, but shift-invariance is lost due to subsampling [27, 28]. Nonsubsampled contourlet transform with a variable weight for fusion of MR and SPECT images has large computational time and complexity [27, 29].

Recently, guided filter fusion (GFF) [30] is used to preserve edges and avoid blurring effects in the fused image. Guided filter is an edge-preserving filter and its computational time is also independent of filter size. However, the method provides limited performance for noisy images due to the use of Gaussian filter and two-level weight maps. An improved guided image fusion for MR and CT imaging is proposed to overcome these limitations. Simulation results based on visual and quantitative analysis show the significance of proposed scheme.

2. Preliminaries

In this section, we briefly discuss the methodology of GFF [30]. The main steps of the GFF method are filtering (to obtain the two-scale representation), weight maps construction, and fusion of base and detail layers (using guided filtering and weighted average method).

Let F be the fused image obtained by combining input images A and B of same sizes ($M \times N$). The base (I_{11} and I_{12}) and detail (I_{21} and I_{22}) layers of source images are

$$\left[\begin{array}{c|c} I_{11} & I_{12} \\ \hline I_{21} & I_{22} \end{array}\right] = \left[\begin{array}{c|c} A * f & B * f \\ \hline A - A * f & B - B * f \end{array}\right], \quad (1)$$

where f is the average filter. The base and detail layers contain large- and small-scale variations, respectively. The saliency images are obtained by convolving A and B and with a Laplacian filter h followed by a Gaussian filter g; that is,

$$\left[\begin{array}{c} S_1 \\ \hline S_2 \end{array}\right] = \left[\begin{array}{c} |A * h| * g \\ \hline |B * h| * g \end{array}\right]. \quad (2)$$

The weight maps P_1 and P_2 are

$$\left[\begin{array}{c} P_1 \\ \hline P_2 \end{array}\right] = \left[\begin{array}{c} \xi(S_1, S_2) \\ \hline \xi(S_2, S_1) \end{array}\right], \quad (3)$$

where $\xi(S_1, S_2)$ is a function with value 1 for $S_1(m,n) \geq S_2(m,n)$ and value 0 for $S_1(m,n) < S_2(m,n)$ (similarly for $\xi(S_2, S_1)$). $S_1(m,n)$ and $S_2(m,n)$ are the saliency values for (m,n) pixel in A and B, respectively.

Guided image filtering is performed to obtain the refined weights W_{11}, W_{12}, W_{21}, and W_{22} as

$$\left[\begin{array}{c|c} W_{11} & W_{12} \\ \hline W_{21} & W_{22} \end{array}\right] = \left[\begin{array}{c|c} G_{r_1,\epsilon_1}(P_1, A) & G_{r_2,\epsilon_2}(P_1, A) \\ \hline G_{r_1,\epsilon_1}(P_2, B) & G_{r_2,\epsilon_2}(P_2, B) \end{array}\right], \quad (4)$$

where r_1, ϵ_1, r_2, and ϵ_2 are the parameters of the guided filter W_{11} and W_{21} and W_{12} and W_{22} are the base layer and the detail layer weight maps.

The fused image F is obtained by weighted averaging of the corresponding layers; that is,

$$F = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} W_{i_1 i_2} I_{i_2 i_1}. \quad (5)$$

The major limitations of GFF [30] scheme are summarized as follows.

(1) The Gaussian filter from (2) is not suitable for Rician noise removal. Thus, the algorithm has limited performance for noisy images. Hence filter of (2) needs to be modified to incorporate noise effects.

(2) The weight maps P_1 and P_2 from (3) can be improved by defining more levels. The main issue with binary assignment (0 and 1) is that when the saliency values are approximately equal, the effect of one value is totally discarded, which results in degraded fused image.

3. Proposed Methodology

The proposed scheme follows the methodology of GFF [30] with necessary modifications to incorporate the above listed limitations. This section first discusses the modification proposed due to noise artifacts and then the improved weight maps are presented.

3.1. Improved Saliency Maps. The acquired medical images are usually of low quality (due to artifacts), which degrade the performance (both in terms of human visualization and quantitative analysis).

Beside other artifacts, MR images often contain Rician Noise (RN) which causes random fluctuations in the data and reduces image contrast [31]. RN is generated when real and imaginary parts of MR data are corrupted with zero-mean, equal variance uncorrelated Gaussian noise [32]. RN is a nonzero mean noise. Note that the noise distribution tends to Rayleigh distribution in low intensity regions and to a Gaussian distribution in regions of high intensity of the magnitude image [31, 32].

Let $\dot{A} = A + N_R$ be image obtained using MR imaging containing Rician noise N_R. The CT image B has higher spatial resolution and negligible noise level [33, 34].

The source images are first decomposed into base $\dot{I}_{11}$ and I_{12} and detail $\dot{I}_{21}$ and I_{22} layers following (1):

$$\left[\begin{array}{c|c} \dot{I}_{11} & I_{12} \\ \hline \dot{I}_{21} & I_{21} \end{array}\right] = \left[\begin{array}{c|c} \dot{A} * f & B * f \\ \hline \dot{A} - \dot{A} * f & B - B * f \end{array}\right]. \tag{6}$$

$\dot{I}_{11}$ and $\dot{I}_{21}$ have an added noise term compared to (1). Linear minimum mean square error estimator (LMMSE) is used instead of Gaussian filter for minimizing RN, consequently improving fused image quality.

The saliency maps $\dot{S}_1$ and $\dot{S}_2$ are thus computed by applying the LMMSE based filter q and following (2):

$$\left[\begin{array}{c} \dot{S}_1 \\ \hline \dot{S}_2 \end{array}\right] = \left[\begin{array}{c} |A * h| * q + |N_R * h| * q \\ \hline |B * h| * q \end{array}\right]. \tag{7}$$

The main purpose of q is to make the extra term $N_R * h$ as in $\dot{S}_1$ small as possible while enhancing the image details.

3.2. Improved Weight Maps. The saliency maps are linked with detail information in the image. The main issue with 0 and 1 weight assignments arises in GFF [30] when different images have approximately equal saliency values. In such cases, one value is totally discarded. For noisy MR images, the saliency value may be higher at a pixel due to noise; in that case it will assign value 1 (which is not desirable). An appropriate solution is to define a range of values for weight maps construction.

Let $\Delta_{AB} = S_1(m,n) - S_2(m,n)$, and $S_1(m,n) \geq S_2(m,n)$,

$$\dot{P}_1(m,n) = \begin{cases} 1 & \text{if } \Delta_{AB} \geq 0.3 \\ 0.8 & \text{if } \Delta_{AB} \geq 0.25 \\ 0.7 & \text{if } \Delta_{AB} \geq 0.2 \\ 0.6 & \text{if } \Delta_{AB} \geq 0.15 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Let $\Delta_{BA} = S_2(m,n) - S_1(m,n)$, and $S_2(m,n) \geq S_1(m,n)$,

$$\dot{P}_2(m,n) = \begin{cases} 1 & \text{if } \Delta_{BA} \geq 0.3 \\ 0.8 & \text{if } \Delta_{BA} \geq 0.25 \\ 0.7 & \text{if } \Delta_{BA} \geq 0.2 \\ 0.6 & \text{if } \Delta_{BA} \geq 0.15 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

These values are selected empirically and may be further adjusted to improve results. Figures 1(a) and 1(b) show CT and noisy MR images, respectively. Figures 1(c)–1(f) show the results of applying different weights. The information in the upper portion of the fused image increases as more levels are added to the weight maps.

The weight maps are passed through guided filter to obtain $\dot{W}_{11}$, $\dot{W}_{21}$, $\dot{W}_{12}$, and $\dot{W}_{22}$. Finally the fused image $\dot{F}$ is

$$\dot{F} = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \dot{W}_{i_1 i_2} \dot{I}_{i_2 i_1}. \tag{10}$$

LMMSE based filter reduces the Rician noise and the more levels of weight maps ensure that more information is transferred to the fused image. The incorporation of the LMMSE based filter and a range of weight map values makes the proposed method suitable for noisy images.

4. Results and Analysis

The proposed method is tested on several pairs of source (MR and CT) images. For quantitative evaluation, different measures including mutual information (MI) [35] measure ζ_{MI}, structural similarity (SSIM) [36] measure ζ_{SSIM}, Xydeas and Petrović's [37] measure ζ_{XP}, Zhao et al.'s [38] measure ζ_{Z}, Piella and Heijmans's [39] measures ζ_{PH_1} and ζ_{PH_2}, and visual information fidelity fusion (VIFF) [40] metric ζ_{VIFF} are considered.

4.1. MI Measure. MI is a statistical measure which provides the degree of dependencies in different images. Large value of MI implies better quality and vice versa [11, 33, 35]:

$$\zeta_{\text{MI}} = 2\left[\frac{1}{(H_A + H_F)} \sum_{a,f} P_{AF}(a,f) \log \frac{P_{AF}(a,f)}{P_A(a)\, P_F(f)} + \frac{1}{(H_B + H_F)} \sum_{b,f} P_{BF}(b,f) \log \frac{P_{BF}(b,f)}{P_B(b)\, P_F(f)}\right], \tag{11}$$

where H_A, H_B, and H_F are the entropies of A, B, and F images, respectively. P_{AF} is the jointly normalized histogram of A and F, P_{BF} is the jointly normalized histogram of B and F, and P_A, P_B, and P_F are the normalized histograms of A, B, and F, respectively.

4.2. SSIM [36] Measure. SSIM [36] measure is defined as

$$\zeta_{\text{SSIM}}(A,B,F) = \begin{cases} \lambda_w \zeta_{\text{SSIM}}(A_w, F_w) + (1-\lambda_w)\, \zeta_{\text{SSIM}}(B_w, F_w) \\ \quad \text{if } \zeta_{\text{SSIM}}(A_w, B_w \mid w) \geq 0.75 \\ \max\left(\zeta_{\text{SSIM}}(A_w, F_w), \zeta_{\text{SSIM}}(B_w, F_w)\right) \\ \quad \text{if } \zeta_{\text{SSIM}}(A_w, B_w \mid w) < 0.75, \end{cases} \tag{12}$$

where w is a sliding window and $\lambda(w)$ is

$$\lambda_w = \frac{\sigma_{A_w}}{\sigma_{A_w} + \sigma_{B_w}}, \tag{13}$$

where σ_{A_w} and σ_{B_w} are the variance of images A and B, respectively.

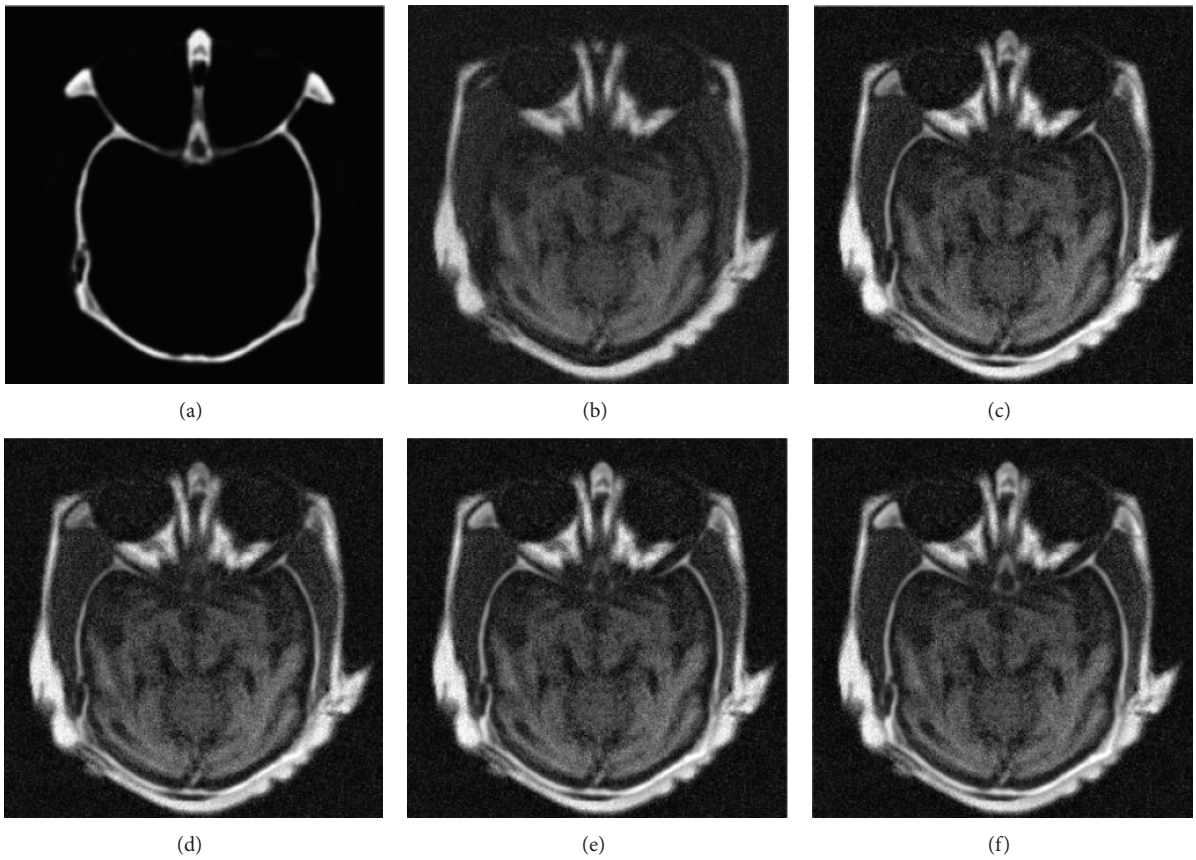


Figure 1: Weight maps comparison: (a) CT image, (b) noisy MR image, (c) fused image with 3 weight maps, (d) fused image with 4 weight maps, (e) fused image with 5 weight maps, and (f) fused image with 6 weight maps.

4.3. Xydeas and Petrović's [37] Measure. Xydeas and Petrović [37] proposed a metric to evaluate the amount of edge information, transferred from input images to fused image. It is calculated as

$$\zeta_{\mathrm{XP}} = \sum_{m=1}^{N}\sum_{n=1}^{M}\left(Q^{AF}(m,n)\,\tau^{A}(m,n) + Q^{BF}(m,n)\,\tau^{B}(m,n)\right) \times \left(\sum_{m=1}^{N}\sum_{n=1}^{M}\left(\tau^{A}(m,n) + \tau^{B}(m,n)\right)\right)^{-1}, \tag{14}$$

where Q^{AF} and Q^{BF} are the product of edge strength and orientation preservation values at location (m,n), respectively. The weights $\tau^{A}(m,n)$ and $\tau^{B}(m,n)$ reflect the importance of $Q^{AF}(m,n)$ and $Q^{BF}(m,n)$, respectively.

4.4. Zhao et al.'s [38] Metric. Zhao et al. [38] used the phase congruency (provides an absolute measure of image feature) to define an evaluation metric. The larger value of the metric describes a better fusion result. The metric ζ_{Z} is defined as the geometric product of phase congruency, maximum and minimum moments, respectively.

4.5. Piella and Heijmans's [39] Metric. Piella and Heijmans's [39] metrics ζ_{P_1} and ζ_{P_2} are defined as

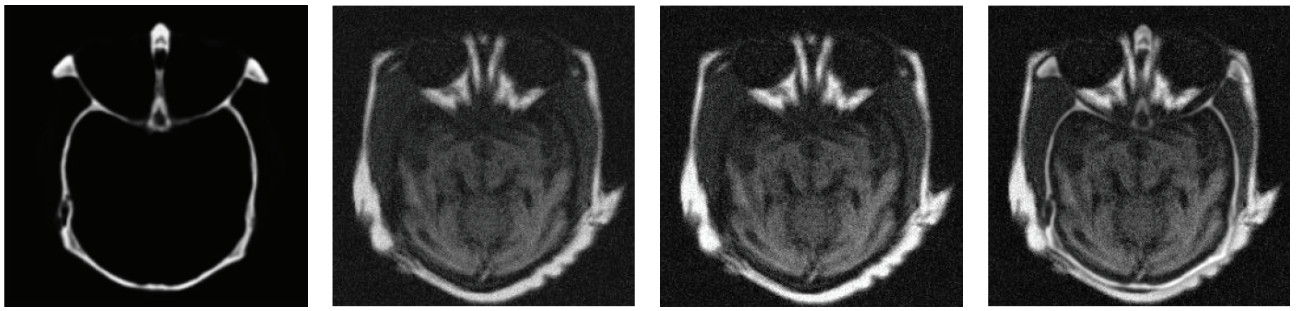
$$\begin{aligned}\zeta_{P_1} &= \frac{1}{|W|}\sum_{w\in W}\left[\lambda(w)\,Q_o(A,F\mid w) + (1-\lambda(w))\,Q_o(B,F\mid w)\right]\\ \zeta_{P_2} &= \sum_{w\in W} c(w)\left[\lambda(w)\,Q_o(A,F\mid w) + (1-\lambda(w))\,Q_o(B,F\mid w)\right],\end{aligned} \tag{15}$$

where $Q_o(A,F \mid w)$ and $Q_o(B,F \mid w)$ are the local quality indexes calculated in a sliding window w and $\lambda(w)$ is defined as in (13). Consider

$$Q_o(A,F\mid w) = \frac{4\sigma_{AF}\overline{A}\,\overline{F}}{\left(\overline{A}^2+\overline{F}^2\right)\left(\sigma_A^2+\sigma_F^2\right)}, \tag{16}$$

where $\overline{A}$ is the mean of A and σ_A^2 and σ_{AF} are the variance of A and the covariance of A, B, respectively. Consider

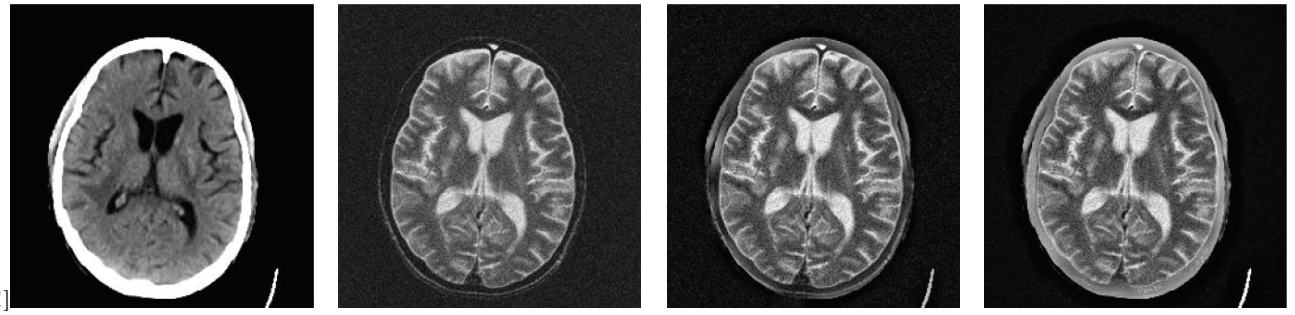
$$c_w = \frac{\max\left[\sigma_{A_w},\sigma_{B_w}\right]}{\sum_{w'\in W}\left[\sigma_{A'_w},\sigma_{A'_w}\right]}, \tag{17}$$



(a) (b) (c) (d)

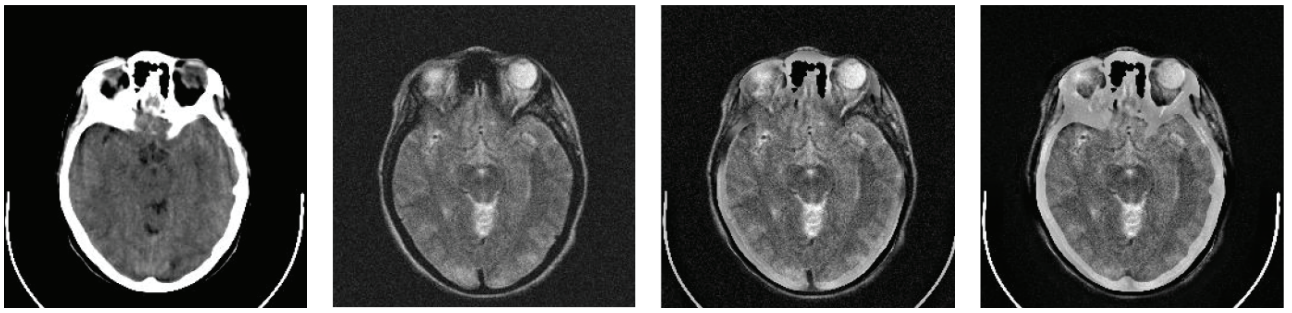
Figure 2: (a) CT image, (b) noisy MR image, (c) GFF [30] fused image, and (d) proposed fused image.

[t!]



(a) (b) (c) (d)

Figure 3: (a) CT image, (b) noisy MR image, (c) GFF [30] fused image, and (d) proposed fused image.



(a) (b) (c) (d)

Figure 4: (a) CT image, (b) noisy MR image, (c) GFF [30] fused image, and (d) proposed fused image.

where σ_{A_w} and σ_{B_w} are the variance of images A and B within the window w, respectively.

4.6. VIFF [40] Metric. VIFF [40] is a multiresolution image fusion metric used to assess fusion performance objectively. It has four stages. (1) Source and fused images are filtered and divided into blocks. (2) Visual information is evaluated with and without distortion information in each block. (3) The VIFF of each subband is calculated and the overall quality measure is determined by weighing (of VIFF at different subbands).

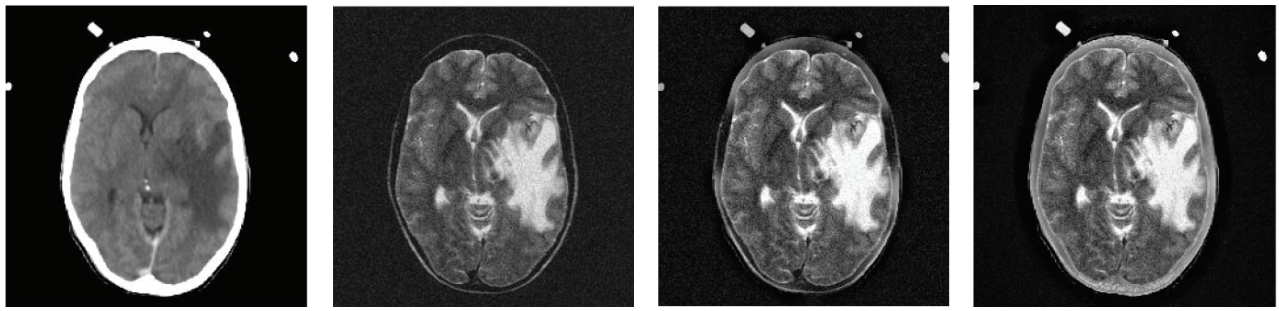
Figure 2 shows a pair of CT and MR images. It can be seen that the CT image (Figure 2(a)) provides clear bones information but no soft tissues information, while in contrast to CT image the MR image in Figure 2(b) provides soft tissues information. The fused image must contain both the information of bones and soft tissues. The fused image obtained using proposed scheme in Figure 2(d) shows better results as compared to fused image obtained by GFF [30] in Figure 2(c).

Figure 3 shows the images of a patient suffering from cerebral toxoplasmosis [41]. A more comprehensive information consisting of both the CT and MR images is the requirement in clinical diagnosis. The improvement in fused image using proposed scheme can be observed in Figure 3(d) compared to image obtained by GFF [30] in Figure 3(c).

Figure 4 shows a pair of CT and MR images of a woman suffering from hypertensive encephalopathy [41]. The improvement in fused image using proposed scheme can be observed in Figure 4(d) compared to image obtained by GFF [30] in Figure 4(c).

Table 1: Quantitative analysis of GFF [30] and proposed schemes.

Quantitative measures	Example 1		Example 2		Example 3		Example 4	
	GFF [30]	Proposed	GFF [30]	Proposed	GFF [30]	Proposed	GFF [30]	Proposed
ζ_{MI}	0.2958	0.2965	0.4803	0.5198	0.4164	0.4759	0.4994	0.5526
ζ_{SSIM}	0.3288	0.3540	0.3474	0.3519	0.3130	0.3139	0.2920	0.2940
ζ_{XP}	0.4034	0.5055	0.4638	0.4678	0.4473	0.4901	0.4498	0.4653
ζ_{Z}	0.1600	0.1617	0.3489	0.3091	0.2061	0.2193	0.3002	0.2855
ζ_{P_1}	0.4139	0.4864	0.2730	0.3431	0.2643	0.3247	0.2729	0.3339
ζ_{P_2}	0.4539	0.7469	0.5188	0.6387	0.6098	0.7453	0.5268	0.6717
ζ_{VIFF}	0.2561	0.3985	0.1553	0.2968	0.1852	0.3009	0.1842	0.3487



(a) (b) (c) (d)

Figure 5: (a) CT image, (b) noisy MR image, (c) GFF [30] fused image, and (d) proposed fused image.

Figure 5 shows a pair of images containing acute stroke disease [41]. The improvement in quality of fused image obtained using proposed scheme can be observed in Figure 5(d) compared to Figure 5(c) (image obtained by GFF [30]).

Table 1 shows that proposed scheme provides better quantitative results in terms of ζ_{MI}, ζ_{SSIM}, ζ_{XP}, ζ_{Z}, ζ_{P_1}, ζ_{P_2}, and ζ_{VIFF} as compared to GFF [30] scheme.

5. Conclusions

An improved guided image fusion for MR and CT imaging is proposed. Different modifications in filter (LMMSE based) and weights maps (with different levels) are proposed to overcome the limitations of GFF scheme. Simulation results based on visual and quantitative analysis show the significance of proposed scheme.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] Y. Yang, D. S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 579341, 13 pages, 2010.

[2] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proceedings of the IEEE*, vol. 91, no. 10, pp. 1699–1721, 2003.

[3] S. Das, M. Chowdhury, and M. K. Kundu, "Medical image fusion based on ripplet transform type-I," *Progress In Electromagnetics Research B*, no. 30, pp. 355–370, 2011.

[4] S. Daneshvar and H. Ghassemian, "MRI and PET image fusion by combining IHS and retina-inspired models," *Information Fusion*, vol. 11, no. 2, pp. 114–123, 2010.

[5] V. Barra and J.-Y. Boire, "A general framework for the fusion of anatomical and functional medical images," *NeuroImage*, vol. 13, no. 3, pp. 410–424, 2001.

[6] A. Polo, F. Cattani, A. Vavassori et al., "MR and CT image fusion for postimplant analysis in permanent prostate seed implants," *International Journal of Radiation Oncology Biology Physics*, vol. 60, no. 5, pp. 1572–1579, 2004.

[7] A. L. Grosu, W. A. Weber, M. Franz et al., "Reirradiation of recurrent high-grade gliomas using amino acid PET (SPECT)/CT/MRI image fusion to determine gross tumor volume for stereotactic fractionated radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 63, no. 2, pp. 511–519, 2005.

[8] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.

[9] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.

[10] A. Jameel, A. Ghafoor, and M. M. Riaz, "Entropy dependent compressive sensing based image fusion," in *Intelligent Signal Processing and Communication Systems*, November 2013.

[11] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Information Fusion*, vol. 13, no. 1, pp. 10–19, 2012.

[12] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 1, pp. 49–61, 2000.

[13] F. Al-Wassai, N. Kalyankar, and A. Al-Zaky, "Multisensor images fusion fased on feature-level," *International Journal of Latest Technology in Engineering, Management and Applied Science*, vol. 1, no. 5, pp. 124–138, 2012.

[14] M. Ding, L. Wei, and B. Wanga, "Research on fusion method for infrared and visible images via compressive sensing," *Infrared Physics and Technology*, vol. 57, pp. 56–67, 2013.

[15] H. B. Mitchell, *Image Fusion Theories, Techniques and Applications*, Springer, 2010.

[16] V. Sharma and W. Davis, "Feature-level fusion for object segmentation using mutual information," in *Augmented Vision Perception in Infrared*, pp. 295–319, 2008.

[17] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[18] A. Ardeshir Goshtasby and S. Nikolov, "Image fusion: advances in the state of the art," *Information Fusion*, vol. 8, no. 2, pp. 114–118, 2007.

[19] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[20] A. Toet, "A morphological pyramidal image decomposition," *Pattern Recognition Letters*, vol. 9, no. 4, pp. 255–261, 1989.

[21] V. S. Petrović and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Transactions on Image Processing*, vol. 13, no. 2, pp. 228–237, 2004.

[22] A. Z. Chitade and S. K. Katiyar, "Multiresolution and multispectral data fusion using discrete wavelet transform with IRS images: cartosat-1, IRS LISS III and LISS IV," *Journal of the Indian Society of Remote Sensing*, vol. 40, no. 1, pp. 121–128, 2012.

[23] S. Li, J. T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images," *Information Fusion*, vol. 3, no. 1, pp. 17–23, 2002.

[24] R. Singh, M. Vatsa, and A. Noore, "Multimodal medical image fusion using redundant discrete wavelet transform," in *Proceedings of the 7th International Conference on Advances in Pattern Recognition (ICAPR '09)*, pp. 232–235, February 2009.

[25] G. Chen and Y. Gao, "Multisource image fusion based on double density dualtree complex wavelet transform," in *Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 9, pp. 1864–1868, 2012.

[26] F. Alia, I. El-Dokanya, A. Saada, and F. Abd El-Samiea, "A curvelet transform approach for the fusion of MR and CT images," *Journal of Modern Optics*, vol. 57, no. 4, pp. 273–286, 2010.

[27] L. Wang, B. Li, and L.-F. Tian, "Multi-modal medical image fusion using the inter-scale and intra-scale dependencies between image shift-invariant shearlet coefficients," *Information Fusion*, 2012.

[28] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.

[29] T. Li and Y. Wang, "Biological image fusion using a NSCT based variable-weight method," *Information Fusion*, vol. 12, no. 2, pp. 85–92, 2011.

[30] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, 2013.

[31] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408–1419, 1999.

[32] C. S. Anand and J. S. Sahambi, "Wavelet domain non-linear filtering for MRI denoising," *Magnetic Resonance Imaging*, vol. 28, no. 6, pp. 842–861, 2010.

[33] Y. Nakamoto, M. Osman, C. Cohade et al., "PET/CT: comparison of quantitative tracer uptake between germanium and CT transmission attenuation-corrected images," *Journal of Nuclear Medicine*, vol. 43, no. 9, pp. 1137–1143, 2002.

[34] J. E. Bowsher, H. Yuan, L. W. Hedlund et al., "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, vol. 4, pp. 2488–2492, October 2004.

[35] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'Information measure for performance of image fusion'," *Electronics Letters*, vol. 44, no. 18, pp. 1066–1067, 2008.

[36] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Information Fusion*, vol. 9, no. 2, pp. 156–160, 2008.

[37] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[38] J. Zhao, R. Laganière, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *International Journal of Innovative Computing, Information and Control*, vol. 3, no. 6, pp. 1433–1447, 2007.

[39] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, pp. 173–176, September 2003.

[40] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127–135, 2013.

[41] Harvard Image Database, https://www.med.harvard.edu/.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 903160, 9 pages
http://dx.doi.org/10.1155/2014/903160

Research Article

Alignment-Free and High-Frequency Compensation in Face Hallucination

Yen-Wei Chen,[1,2] So Sasatani,[2] and Xian-Hua Han[2]

[1] *College of Computer Science and Information Technology, Central South University of Forestry and Technology, Hunan 410004, China*
[2] *College of Information Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan*

Correspondence should be addressed to Yen-Wei Chen; chen@is.ritsumei.ac.jp

Received 25 August 2013; Accepted 21 November 2013; Published 12 February 2014

Academic Editors: S. Bourennane and J. Marot

Face hallucination is one of learning-based super resolution techniques, which is focused on resolution enhancement of facial images. Though face hallucination is a powerful and useful technique, some detailed high-frequency components cannot be recovered. It also needs accurate alignment between training samples. In this paper, we propose a high-frequency compensation framework based on residual images for face hallucination method in order to improve the reconstruction performance. The basic idea of proposed framework is to reconstruct or estimate a residual image, which can be used to compensate the high-frequency components of the reconstructed high-resolution image. Three approaches based on our proposed framework are proposed. We also propose a patch-based alignment-free face hallucination. In the patch-based face hallucination, we first segment facial images into overlapping patches and construct training patch pairs. For an input low-resolution (LR) image, the overlapping patches are also used to obtain the corresponding high-resolution (HR) patches by face hallucination. The whole HR image can then be reconstructed by combining all of the HR patches. Experimental results show that the high-resolution images obtained using our proposed approaches can improve the quality of those obtained by conventional face hallucination method even if the training data set is unaligned.

1. Introduction

There is a high demand for high-resolution (HR) images such as video surveillance, remote sensing, and medical imaging because high-resolution images can reveal more information than low-resolution images. However, it is hard to improve the image resolution by replacing sensors because of the high cost, hardware physical limits. Super resolution image reconstruction (SR) is one promising technique to solve the problem [1, 2]. SR can be broadly classified into two families of methods: (1) the classical multiframe super resolution [2] and (2) the single-frame super resolution, which is also known as example-based or learning-based super resolution [3–5]. In the classical multiimage SR, the HR image is reconstructed by combining subpixel-aligned multiimages (LR images). In the learning-based SR, the HR image is reconstructed by learning correspondence between low and high-resolution image patches from a database.

Face hallucination is one of learning-based SR techniques proposed by Baker and Kanade [1, 6], which is focused on resolution enhancement of facial images. To date, a lot of algorithms of face hallucination methods have been proposed [7–12]. Though face hallucination is a powerful and useful technique, some detailed high-frequency components cannot be recovered. In this paper, we propose a high-frequency compensation framework based on residual images for face hallucination method in order to improve the reconstruction performance. The basic idea of proposed framework is to reconstruct or estimate a residual image, which can be used to compensate the high-frequency components of the reconstructed high-resolution image. Three approaches based on our proposed framework are proposed. We also propose a patch-based alignment-free face hallucination method. In the patch-based face hallucination, we first segment facial images into overlapping patches and construct training patch pairs. For an input LR image, the overlapping patches are also used to obtain the corresponding HR patches by face hallucination. The whole HR image can then be reconstructed by combining all of the HR patches.

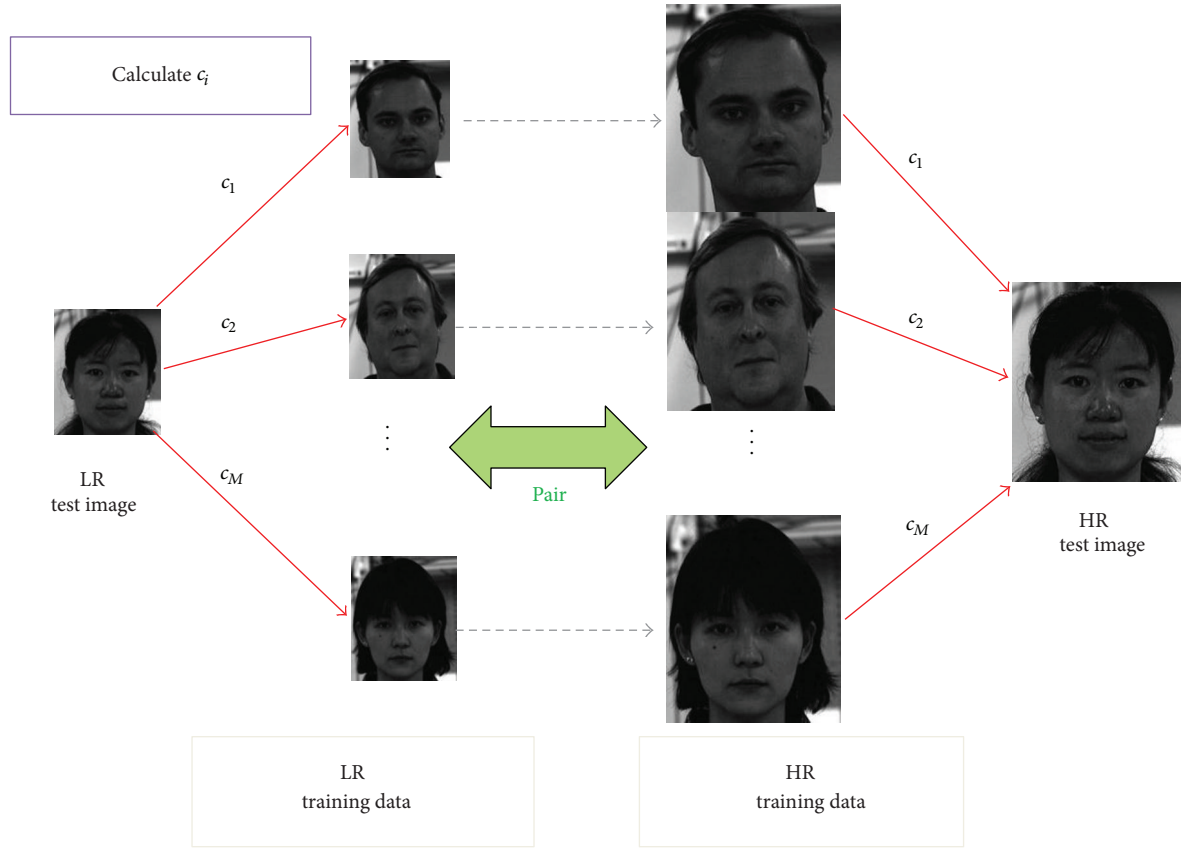


Figure 1: Schematic diagram of face hallucination.

Figure 2: Framework for recovering high-frequency components in face hallucination.

The paper is organized as follows. In Section 2, we describe the conventional face hallucination method. Our proposed residual image compensation methods are presented in Section 3. Our proposed patch-based alignment-free method is presented in Section 4. Section 5 presents experimental results and quantitative evaluation. Section 6 summarizes our conclusions.

2. Face Hallucination

The face hallucination method is one of learning-based SR methods, which is proposed for resolution enhancement of facial images [6–12]. In this section, we briefly introduce the basic concept of face hallucination, which is shown in Figure 1.

The basic idea of face hallucination is that a face image can be reconstructed from other face images by linear combination because all facial images have a similar structure. In face hallucination, an input LR image can be represented as a linear sum of the LR training images along with some learned coefficients. Due to the correlation between the LR and HR images in the training dataset, the output HR image can also be calculated by finding the linear sum of the corresponding HR images using the same coefficients.

We represent a two-dimensional face image using a column vector of all pixel values, and $\mathbf{X}_I$ represents the input LR face image. HR training images are denoted by $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_M]$, and the corresponding LR training images are $\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_M]$, where M is the number of training image pairs. First, we interpolate the LR training images and the input (test) LR image to the resolution space of the

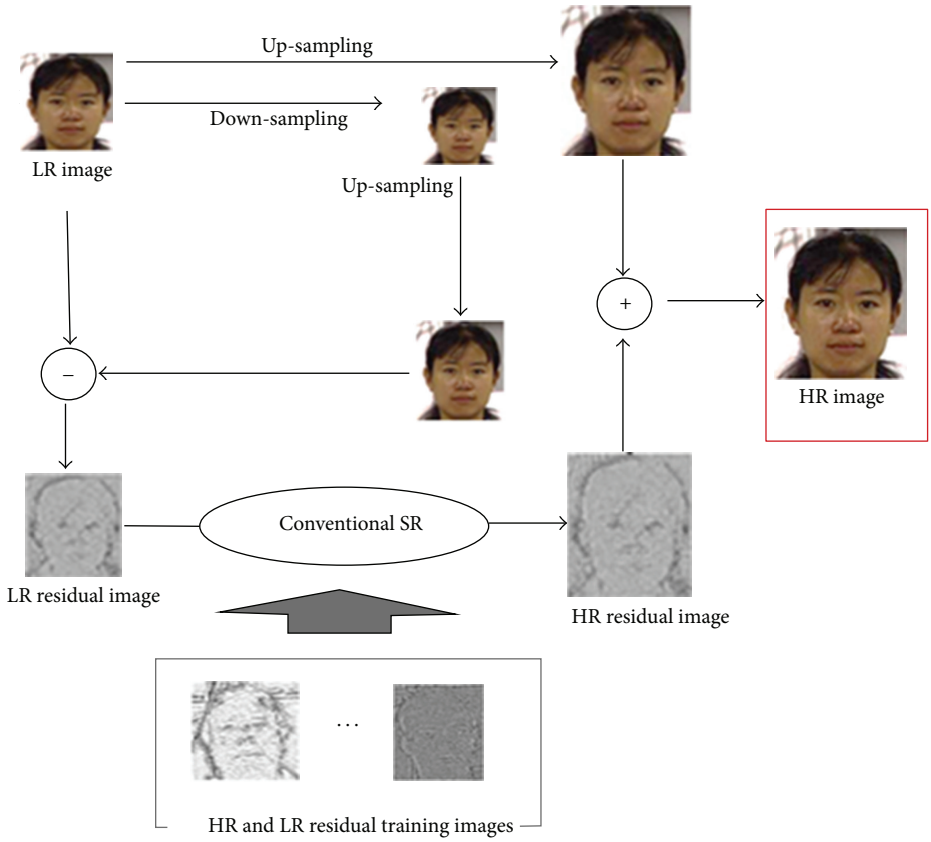


Figure 3: Proposed super resolution method 1.

HR training images, denoted by $\widetilde{\mathbf{L}} = [\widetilde{\mathbf{L}}_1, \widetilde{\mathbf{L}}_2, \ldots, \widetilde{\mathbf{L}}_M]$ and $\widetilde{\mathbf{X}}_l$, respectively. $\widetilde{\mathbf{X}}_l$ may be represented by a linear sum of interpolated training LR images using

$$\widetilde{\mathbf{X}}_l = c_1\widetilde{\mathbf{L}}_1 + c_2\widetilde{\mathbf{L}}_2 + \cdots + c_M\widetilde{\mathbf{L}}_M, \tag{1}$$

where $\mathbf{C} = [c_1, c_2, \ldots, c_M]$ are the weight coefficients, satisfying the following constraint:

$$c_1 + c_2 + \cdots + c_M = 1. \tag{2}$$

The optimal weights can be calculated by minimizing the error in reconstructing the input LR image $\widetilde{\mathbf{X}}_l$ from training LR images. This error is defined in (3). After substitution of the constraints in (2) into (3), the weight vector may be obtained using covariance matrix $\mathbf{S}$ in (4). So,

$$\varepsilon(\mathbf{C}) = \left|\mathbf{X}_h - \widetilde{\mathbf{X}}\right|^2 = \left|\mathbf{X}_h - \sum_{i=1}^{M} c_i\widetilde{\mathbf{L}}_i\right|^2, \tag{3}$$

$$\mathbf{C} = \frac{\mathbf{S}^{-1}1}{1^T\mathbf{S}^{-1}1}, \quad \left(\mathbf{S} = \left(\widetilde{\mathbf{X}}_l 1^T - \widetilde{\mathbf{L}}\right)^T\left(\widetilde{\mathbf{X}}_l 1^T - \widetilde{\mathbf{L}}\right)\right). \tag{4}$$

After obtaining the coefficients for reconstructing the input LR image with LR training images as given in (1), we replace $\widetilde{\mathbf{L}}$ with $\mathbf{H}$ using the same coefficients $\mathbf{C}$. Subsequently, the HR image $\mathbf{X}_h$ can be obtained using

$$\mathbf{X}_h = c_1\mathbf{H}_1 + c_2\mathbf{H}_2 + \cdots + c_M\mathbf{H}_M. \tag{5}$$

3. High Frequency Compensations Using Residual Images

Though face hallucination is a powerful and useful technique, some detailed high-frequency components cannot be recovered. In this paper, we propose a high-frequency compensation framework based on residual images for face hallucination method in order to improve the reconstruction performance. The basic idea of proposed framework is to reconstruct or estimate a residual image, which can be used to compensate the high-frequency components of the reconstructed high-resolution image as shown in Figure 2.

In order to estimate the residual image, we construct a new residual image database (pairs of LR and HR residual images) in addition to existing image database (pairs of LR and HR images) for training. The construction of the residual image database is shown in Figure 3. The HR and LR training image pairs $[\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_M]$ and $[\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_M]$ already exist in conventional face hallucination. With conventional face hallucination, for each LR training image $\mathbf{L}$, the other $M - 1$ training image pairs are used to obtain

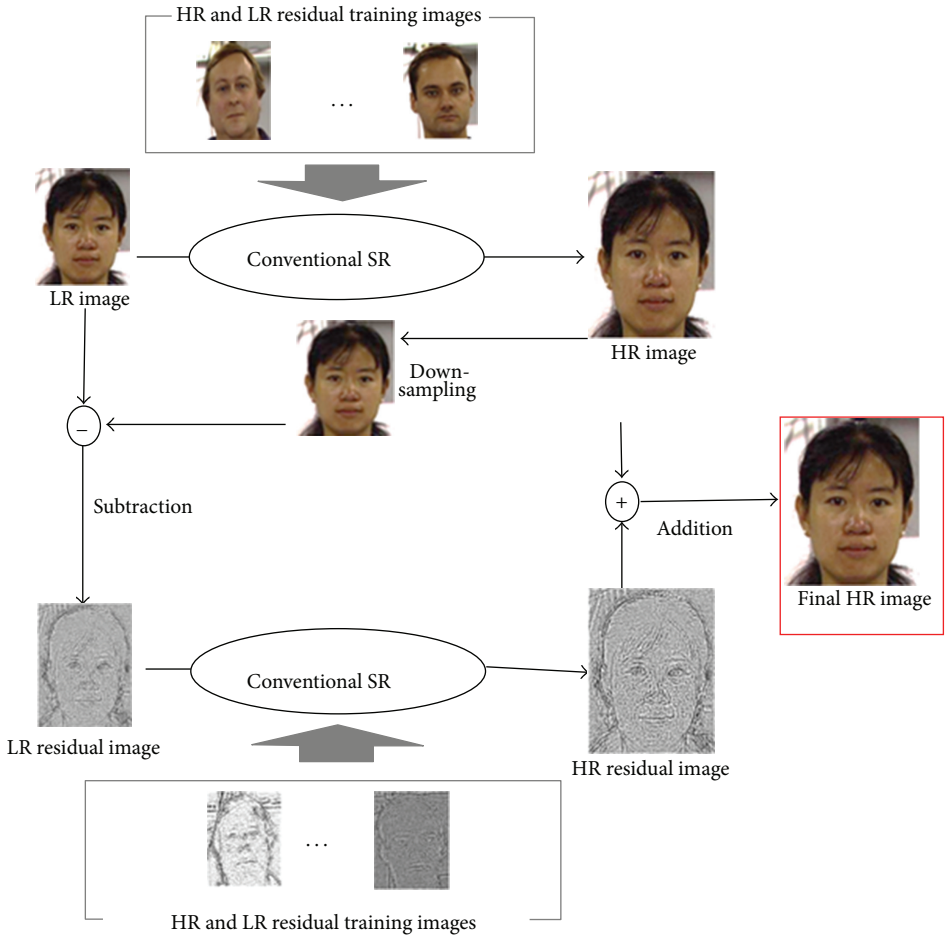


Figure 4: Proposed super resolution method 2.

the approximated HR image $\widehat{\mathbf{H}}$. The HR training residual image is the difference between the original HR image and the reconstructed HR image $\widehat{\mathbf{H}}$, while the LR residual image is the difference between the original LR image and the downsampled version of the reconstructed HR image.

With the two training pair databases, three approaches are proposed for high frequency compensation.

3.1. Proposed Method 1. The first approach is shown in Figure 3. We first use the conventional interpolation method to obtain an HR image and calculate the LR residual image between the input LR image and the downsampled reconstructed HR image. Then we reconstruct the HR residual image from the LR residual image using training residual image pairs. Finally we merge the HR residual and the interpolated HR images.

3.2. Proposed Method 2. The second approach is shown in Figure 4. We first use the conventional face hallucination method to obtain an HR image and calculate the LR residual image between the input LR image and the downsampled reconstructed HR image. Then we reconstruct the HR residual image from the LR residual image using training residual image pairs. Finally we merge the HR residual and the reconstructed HR images.

3.3. Proposed Method 3. The third approach is shown in Figure 5. We first use our proposed SR method 1 to obtain an HR image and calculate the LR residual image between the input LR image and the downsampled reconstructed HR image. Then we reconstruct the HR residual image from the LR residual image using training residual image pairs. Finally we merge the HR residual and the reconstructed HR images.

4. Alignment-Free Patch-Based Face Hallucination

In the conventional face hallucination (Figure 6(a)), whole face image is used in training and SR reconstruction. Each pixel is reconstructed by the use of the corresponding pixel pairs at the same position. So the conventional face hallucination needs an accurate alignment of facial images in order to obtain good reconstructed HR images. With some imperfectly aligned facial datasets, it is difficult to reconstruct sharp HR images using face hallucination. In this paper, we propose a patch-based face hallucination. The basic

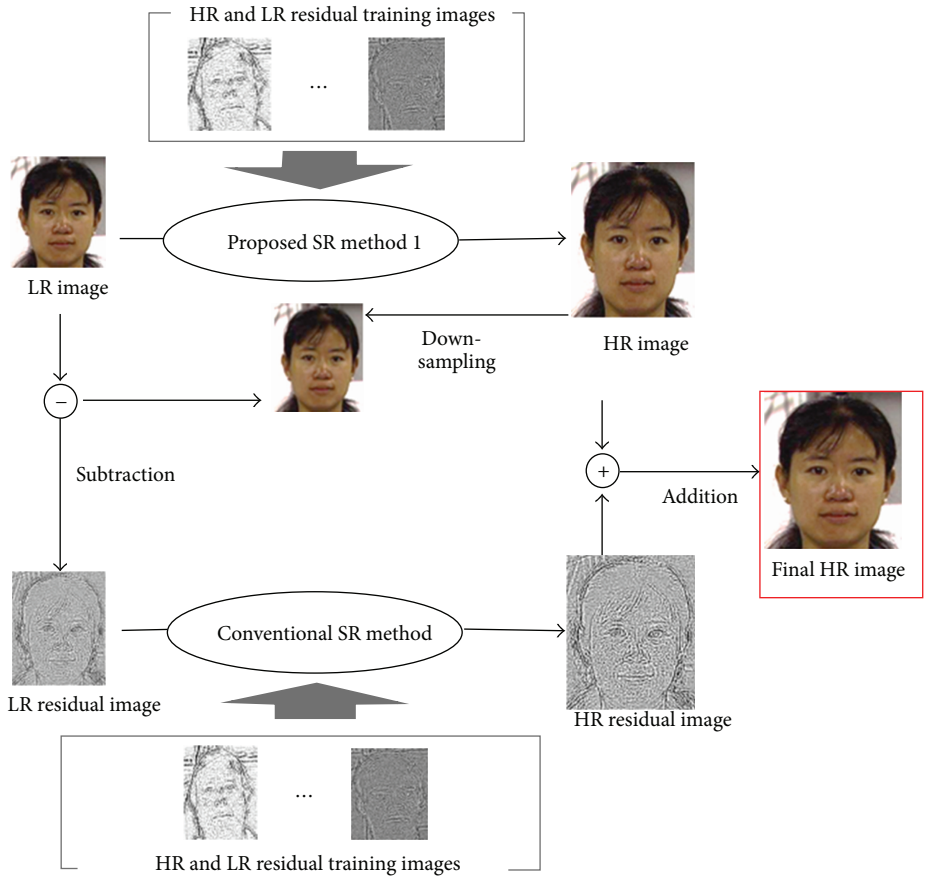


Figure 5: Proposed super resolution method 3.

idea of our proposed method is shown in Figure 6(b). In our proposed alignment-free patch-based face hallucination, we first segment facial images into overlapping patches and construct training patch pairs. For reconstruction of HR image, each overlapping patch in the input LR image is used as a template and its corresponding patches in LR training images are found by the use of an SSD (sum of squared difference) based template matching. Then its HR patches can be obtained by face hallucination. The whole HR image can be reconstructed by combining all of the HR patches.

5. Experimental Results

In order to validate the effectiveness of our proposed methods, we apply our proposed methods to two face databases. The first one is our developed MaVIC database (multiangle View, illumination and cosmetic facial image database) [13], which contains 99 aligned images of different persons and the size of each image is 320 × 400. The second one is C&P database provided by Kanade et al. [14] and Pie [11], which contains 165 imperfectly aligned frontal face images, and each image size is 264 × 320. We first generate the LR and HR image pairs by downsampling the original images. The size of LR images is 50 × 61, while the size of HR images is 200 × 244. The leave-one-out method is used in our experiments. In each database, we select one LR image randomly as a test image and its HR image is used as a ground truth image for quantitative evaluation. Other image pairs are used for training. Our proposed three methods are used for HR reconstruction of the LR test image. In order to make a comparison, the conventional face hallucination method and the bicubic interpolation method are also used for reconstructions. For each method, a total of 20 experiments with a different test image are performed. The peak signal-to-noise ratio (PSNR) [dB] is used as a quantitative measure for evaluation of the HR reconstruction performance. For C&P's imperfectly aligned facial datasets, our proposed patch face hallucination method is used with a patch size of 3 × 3 and a 1 × 1 patch that overlaps with adjacent patches.

Firstly, we show experimental results with the aligned face database MaVIC. A typical example is shown in Figure 7. The test LR image, which is not included in the training samples, is shown in the upper left of Figure 7. The grand truth HR image is shown in the lower right of Figure 7. Others are reconstructed HR images with their PSNR by different methods. It can be seen that the reconstructed high-resolution images obtained using our proposed approaches are much better than those obtained by conventional face hallucination method and bicubic interpolation method and the proposed method 3 shows the best performance among three proposed methods. Similar results have also been obtained with other test images. PSNR evaluation results for

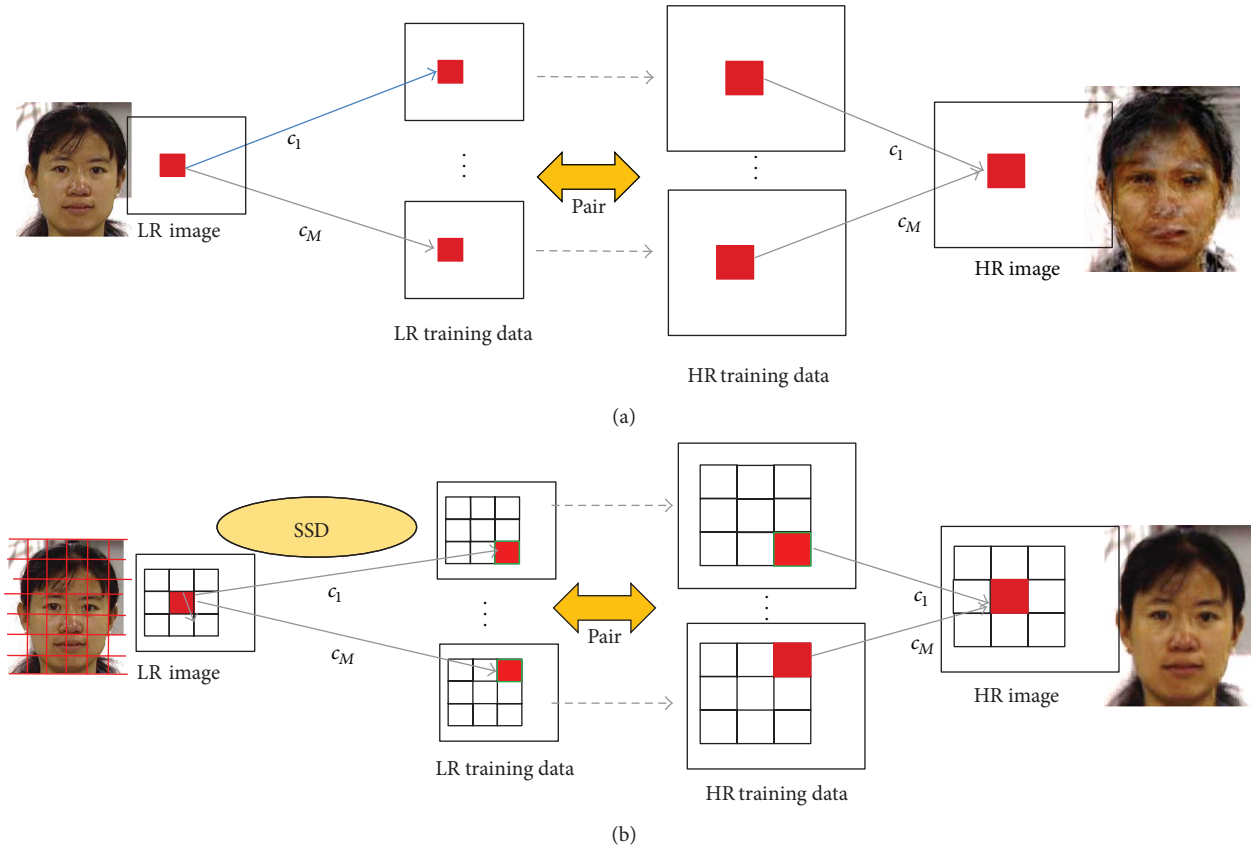


Figure 6: (a) Conventional face hallucination, (b) our proposed alignment-free patch-based face hallucination.

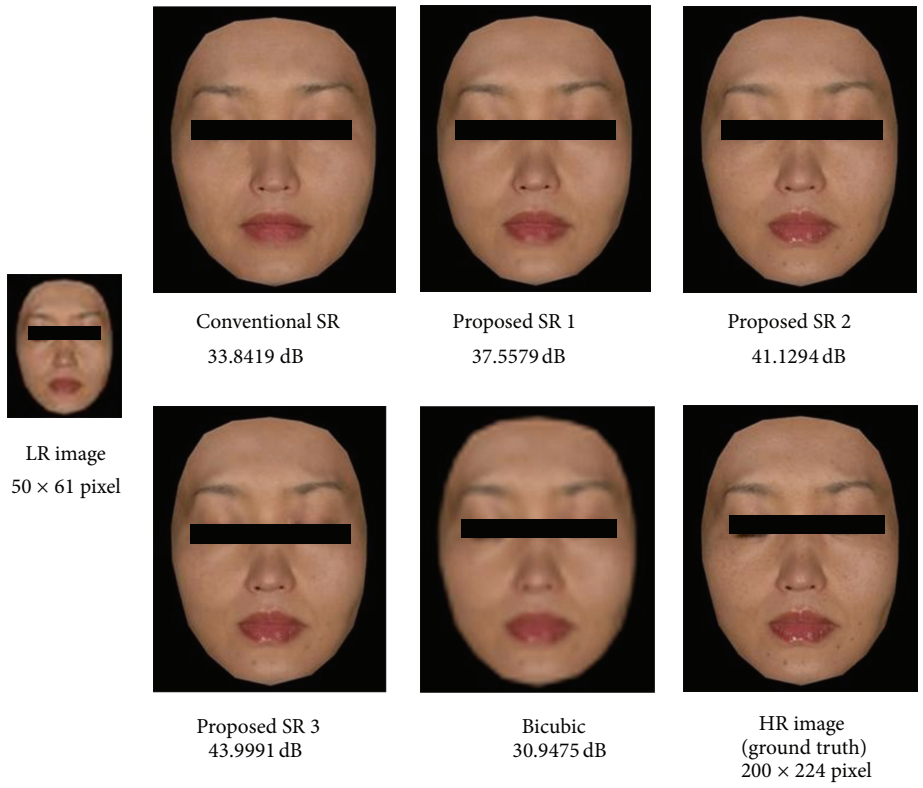


Figure 7: Typical reconstructed images for an aligned face image database (MaVIC).

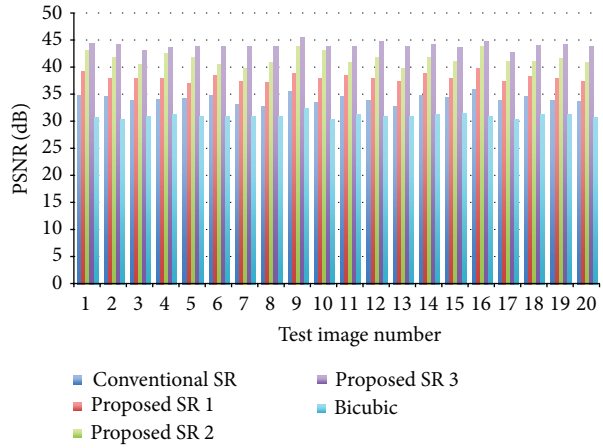


Figure 8: PNSR of 20 test images (MaVIC) by different methods.

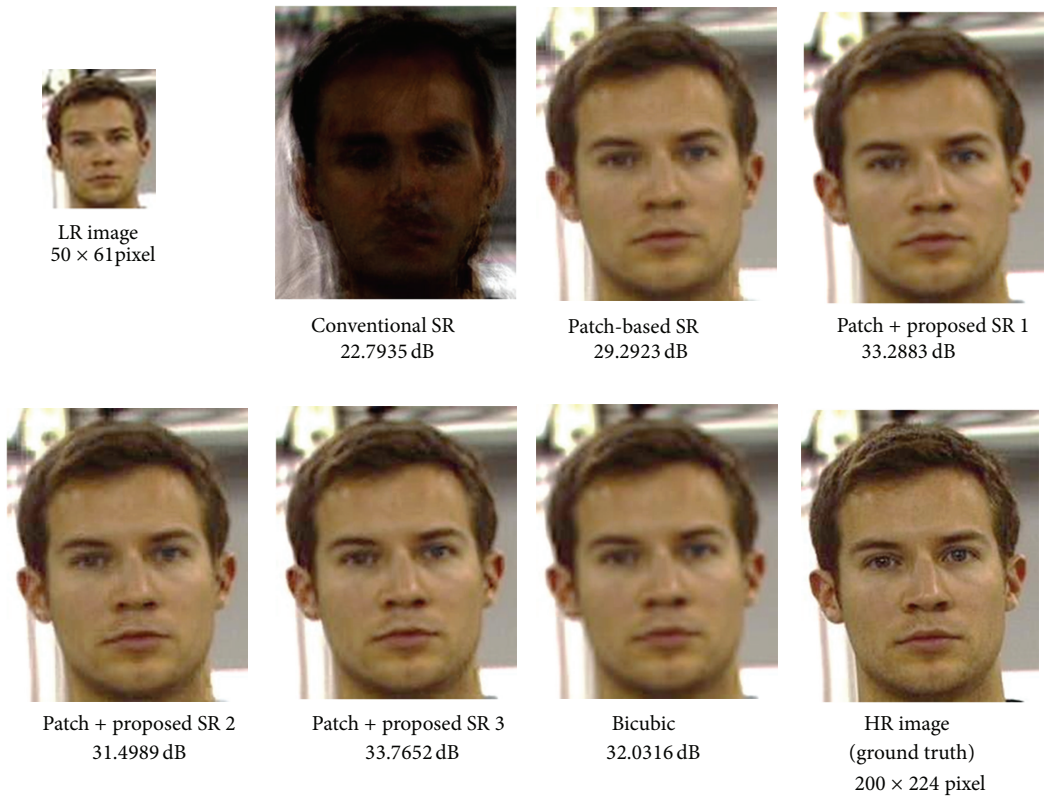


Figure 9: Typical reconstructed images for a nonaligned face image database (C&P).

all test images are shown in Figure 8. The image shown in Figure 7 is corresponding to image no. 20. The mean and standard deviation over 20 experiments for each method are summarized in Table 1. It can be seen that the averaged PSNR with bicubic interpolation is about 31.16. The averaged PNSR can be improved to 34.35 [dB] by the use of conventional face hallucination. Significant improvements can be achieved by our proposed high-frequency compensation methods. The proposed method 3 shows the best performance among three proposed methods. The averaged PSNR is improved to 44.17 [dB].

Table 1: Comparison of the averaged PSNR for an aligned face database (MaVIC).

Method	Conventional SR [10]	Proposed SR 1	Proposed SR 2	Proposed SR 3	Bicubic
Mean	34.35	38.18	41.74	44.17	31.16
Stv.	0.81	0.72	1.18	0.60	0.45

Next, we show experimental results with the unaligned face database (C&P). Unlike experiments with MaVIC, our proposed patch-based method is used for reconstruction of

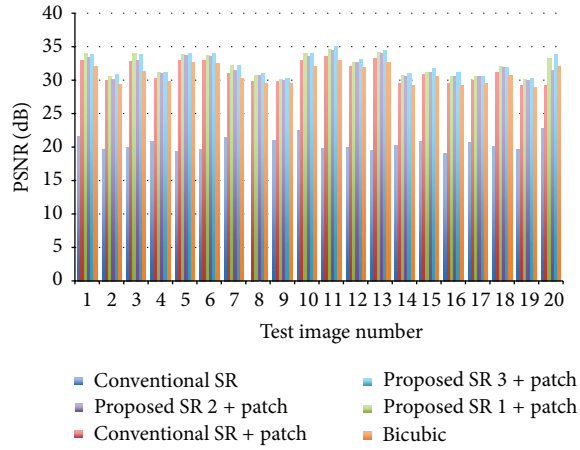


Figure 10: PNSR of 20 test images (C&P) by different methods.

Table 2: Comparison of the averaged PSNR for a nonaligned face image database (C&P).

Methods	Conventional SR [10]	Conventional SR + patch	Proposed SR 1 + patch	Proposed SR 2 + patch	Proposed SR 3 + patch	Bicubic
Mean	20.42	31.21	32.49	32.19	32.96	30.86
Stv.	1.03	1.54	1.60	1.50	1.58	1.37

HR images. A typical example is shown in Figure 9. The test LR image, which is not included in the training samples, is shown in the upper left of Figure 9. The grand truth HR image is shown in the lower right of Figure 9. Others are reconstructed HR images with their PSNR by different methods. PSNR evaluation results for all test images are shown in Figure 10. The image shown in Figure 9 is corresponding to image no. 20. The mean and standard deviation over 20 experiments for each method are summarized in Table 2. As we discussed in previous section, conventional face hallucination needs an accurate alignment of facial images in order to obtain good reconstructed HR images. For the unaligned face database C&P (usually existing face databases are not aligned), the conventional SR method, which uses whole face image for super resolution, gives a very poor result, while if we use our proposed patch-based framework, we can significantly improve the reconstructed HR image even with the conventional face hallucination method. The averaged PSNR can be improved from 20.42 [dB] to 31.21 [dB]. It means that our proposed method is an alignment-free SR method. Furthermore, the HR image with higher quality can be achieved by our proposed high-frequency compensation methods. As well as the results with MaVIC, the proposed method 3 shows the best performance among three proposed methods. The averaged PSNR is improved to 32.96 [dB]. Though smaller patch size may improve the quality of reconstructed HR images, it will increase the computation cost. Optimum patch size depends on the purpose and applications.

6. Conclusions

We proposed a residual image compensation framework together with a patch-based alignment-free method to improve the reconstruction quality for face hallucination. The basic idea of our proposed residual image compensation framework was to reconstruct or estimate a residual image, which can be used to compensate the high-frequency components of the reconstructed high-resolution image. Three approaches based on our proposed framework were proposed. In the patch-based alignment-free face hallucination, we first segmented facial images into overlapping patches and constructed training patch pairs. For an input LR image, the overlapping patches are also used to obtain the corresponding HR patches by face hallucination. The whole HR image can then be reconstructed by combining all of the HR patches. The effectiveness of our proposed methods has been demonstrated on both the aligned face database (MaVIC) and the unaligned face database (C&P). The reconstructed high-resolution images obtained using our proposed approaches are much better than those obtained by conventional face hallucination method and bicubic interpolation method. The averaged PSNR of reconstructed HR images was improved from 34.35 [db] to 44.17 [dB] for the aligned face database (MaVIC) and from 20.42 [dB] to 32.40 [dB] for the unaligned face database (C&P). The proposed high-frequency compensation method 3 shows the best performance among three proposed approaches.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported in part by the MEXT Support Program for the Strategic Research Foundation at Private Universities (2013–2017) and in part by the R-GIRO Research Fund from Ritsumeikan University. Images in MaVIC database are provided by Kao Co., Ltd. Authors would like to thank Mr T. Igarashi of Kao Co., Ltd., Mr. Y. Iwamoto, K. Taniguchi, and S. Nojima for their useful discussions and assistance in this paper and experiments.

References

[1] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[3] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[4] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. I275–I282, July 2004.

[5] J. Qiao, J. Liu, and Y.-W. Chen, "Joint blind super-resolution shadow removing," *IEICE Transactions on Information and Systems*, vol. E90-D, no. 12, pp. 2060–2069, 2007.

[6] S. Baker and T. Kanade, "Hallucinating faces," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, March 2000.

[7] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: theory and practice," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[8] X. Ma, J. Zhang, and C. Qi, "An example-based two-step face Hallucination method through coefficient learning," in *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR '09)*, vol. 5627 of *Lecture Notes in Computer Science*, pp. 471–480, 2009.

[9] X. Wang and X. Tang, "Face hallucination and recognition," in *Proceedings of the Audio- and Video-Based Biometric Person Authentication (AVBPA' 03)*, vol. 2688 of *Lecture Notes in Computer Science*, p. 1054, 2003.

[10] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 35, no. 3, pp. 425–434, 2005.

[11] S. Sasatani, X.-H. Han, T. Igarashi, M. Ohashi, Y. Iwamoto, and Y.-W. Chen, "High frequency compensated face hallucination," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 1561–1564, September 2011.

[12] Y. Liang, J.-H. Lai, W.-S. Zheng, and Z. Cai, "A survey of face hallucination," in *Proceedings of the 7th Chinese Conference on Biometric Recognition (CCBR '12)*, vol. 7701 of *Lecture Notes in Computer Science*, pp. 83–93, Springer.

[13] Y. W. Chen, T. Fukui, X. Qiao, T. Igarashi, K. Nakao, and A. Kashimoto, "Multi-angle view, illumination and cosmetic facial image database (MaVIC) and quantitative analysis of facial appearance," in *Structural, Syntactic, and Statistical Pattern Recognition*, vol. 5342 of *Lecture Notes in Computer Science*, pp. 411–420, Springer, 2008.

[14] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 46–53, 2000.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 140930, 17 pages
http://dx.doi.org/10.1155/2014/140930

Research Article

Analysis of Infrared Signature Variation and Robust Filter-Based Supersonic Target Detection

Sungho Kim,[1] Sun-Gu Sun,[2] and Kyung-Tae Kim[3]

[1] *Yeungnam University, Gyeongsan-si, Gyeongsangbuk-do 712-749, Republic of Korea*
[2] *Agency for Defense Development, 111 Sunam-dong, Daejeon 305-152, Republic of Korea*
[3] *Electromagnetic Technology Laboratory, POSTECH, Pohang 790-784, Republic of Korea*

Correspondence should be addressed to Sungho Kim; sunghokim@ynu.ac.kr

Received 22 August 2013; Accepted 17 November 2013; Published 11 February 2014

Academic Editors: S. Bourennane and J. Marot

The difficulty of small infrared target detection originates from the variations of infrared signatures. This paper presents the fundamental physics of infrared target variations and reports the results of variation analysis of infrared images acquired using a long wave infrared camera over a 24-hour period for different types of backgrounds. The detection parameters, such as signal-to-clutter ratio were compared according to the recording time, temperature and humidity. Through variation analysis, robust target detection methodologies are derived by controlling thresholds and designing a temporal contrast filter to achieve high detection rate and low false alarm rate. Experimental results validate the robustness of the proposed scheme by applying it to the synthetic and real infrared sequences.

1. Introduction

Small infrared target detection is important for a range of military applications, such as infrared search and track (IRST) and active protection system (APS). In particular, an APS is a system designed to protect tanks from guided missile or rocket attack by a physical counterattack, as shown in Figure 1. An antitank missile, such as high explosive antitank (HEAT), should be detected at a distance of at least 1 km and tracked for active protection using RADAR and infrared (IR) cameras. Although RADAR and IR can complement each other, this study focused on the IR sensor-based approach because it can provide a high resolution angle of arrival (AOA) and detect high temperature targets.

IR sensors are inherently passive systems and do not have all weather capability. In addition, IR images show severe variations according to background, time, temperature, and humidity, which makes the target detection difficult. The use of adaptive IR sensor management techniques can enhance the target detection performance. On the other hand, few studies have analyzed the IR variations in terms of small target detection using the data collected over a 24-hour period. In 2006, Jacobs summarized the basics of radiation, atmospheric parameters, and infrared signature characterization [1]. He measured the thermal variations in various environments. Recently, the TNO research team characterized small surface targets in coastal environments [2–4]. In 2007, the TNO team introduced the measurement environment and examined the target contrast and contrast features of a number of small sea surface vessels [2]. The analysis revealed a variation in contrast due to changes in solar illumination, temperature cooling by wind and sun-glint. In 2008, the team analyzed the variations in the vertical profile radiance and clutter in a coastal environment [3]. Based on the analysis, a Maritime Infrared Background Simulator (MIBS) background simulation was performed under these measurement conditions. They can predict clutter in coastal background accurately. In 2009, they extended the IR-based analysis to visible cameras, hyperspectral cameras, and polarization filters to validate the contrast and clutter model [4].

The first contribution of this paper is IR variation analysis in terms of small target detection. The second contribution is the acquisition of 24-hour IR data in winter and spring.

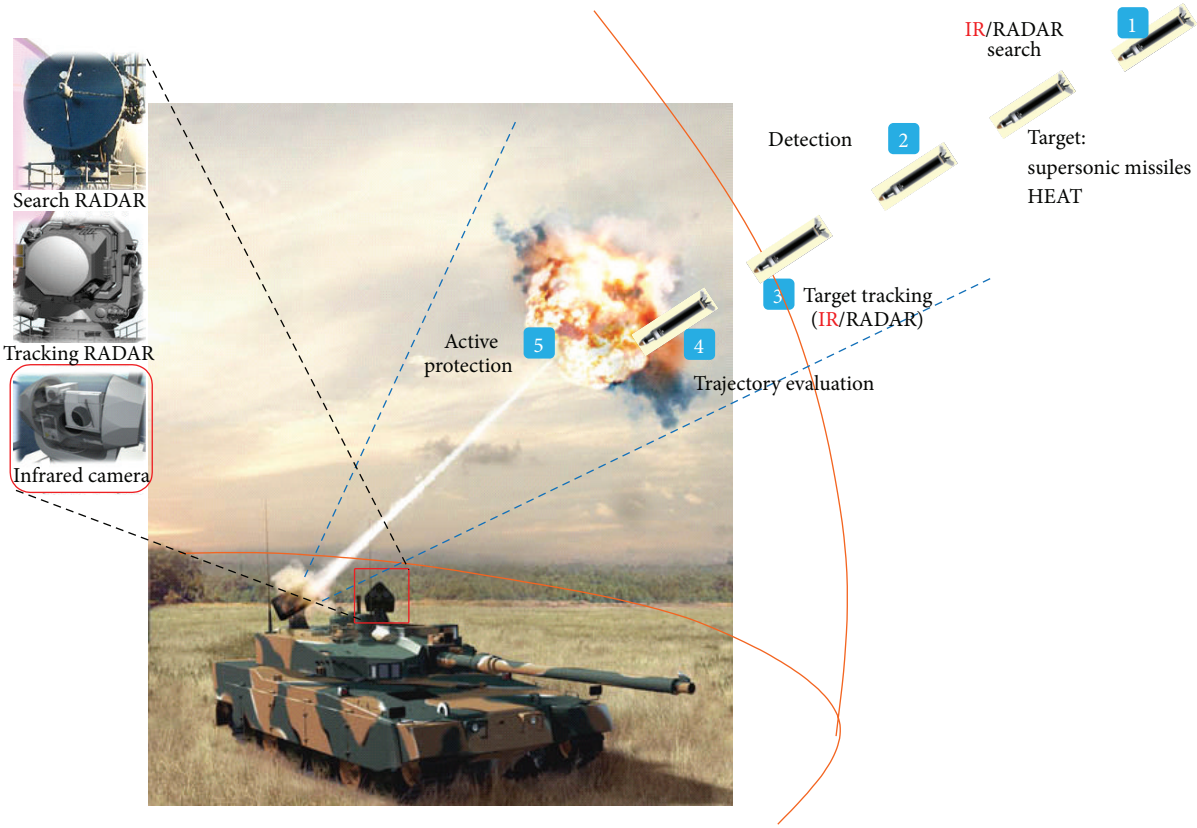


Figure 1: IR-based target detection and tracking in active protection system (APS).

The third contribution is variation analysis for different backgrounds. The fourth contribution is the proposition of robust small target detection based on the variation analysis. Section 2 characterize the infrared signature and presents the sources of infrared variations. Section 3 explains the target detection parameters and variation measurement results including the details of the IR variability in different background, time, temperature, and humidity. Section 4 proposes small target detection methods to overcome the target signature variations. Section 5 presents the experimental results for synthetic and real test sequences. Section 6 discusses the analysis results and concludes the paper.

2. Characterization of Infrared Target Signature and Variation

2.1. Physical Modeling of Infrared Target Signature. In general, IR target images are obtained by the process of IR radiation contrast, attenuation by atmospheric transmittance, photon to voltage conversion in the IR sensor, and analog-to-digital conversion (ADC), as shown in Figure 2. The temperature contrast between the target and background is radiated and propagated in the air. The radiation contrast is attenuated by atmospheric effects, such as absorption, emission, and scattering of H_2O and CO_2.

Objects with temperatures higher than 0 K radiate thermal energy, and such energy differences create voltage differences in thermal detectors. Most target detection algorithms use the difference information between target and background energy. Based on such basic radiation theory, we can predict relative target digital signal levels given the information of thermal radiation intensity difference between the target and background by carefully modeling the energy transformation processes. Modeling procedures are as follows. First, we input a thermal radiant intensity difference between the target and background. Second, we calculate atmospheric transmissivity. Third, we calculate the number of photons in front of the sensors. Fourth, we calculate the voltage levels at a detector. Finally, we obtain gray levels or digital counts after analog-to-digital (A/D) converter.

Since the objective of APS is to detect distant targets as early as possible, targets can be regarded as point sources. Therefore, we use the output voltage model in the thermal detector as the following equation [5]:

$$\Delta V_{\text{pnt}} = G \int_{\lambda_1}^{\lambda_2} R_d(\lambda) \frac{I(\lambda, \Delta T) A_0}{R^2} \cdot \text{PVF} \cdot \tau_O(\lambda) \cdot T_{\text{atm}}(\lambda) \, d\lambda, \quad (1)$$

where $I(\lambda, \Delta T)$ [watt/sr μm] denotes radiant intensity at wavelength (λ) when the temperature difference is ΔT. R [m] represents the distance between the IR camera and a target. These two parameters should be provided by users. ΔV_{pnt} means voltage difference produced by target region and background region in a thermal detector. $T_{\text{atm}}(\lambda)$ denotes atmospheric transmissivity that is defined as a ratio (the fraction or percent of a particular frequency or wavelength

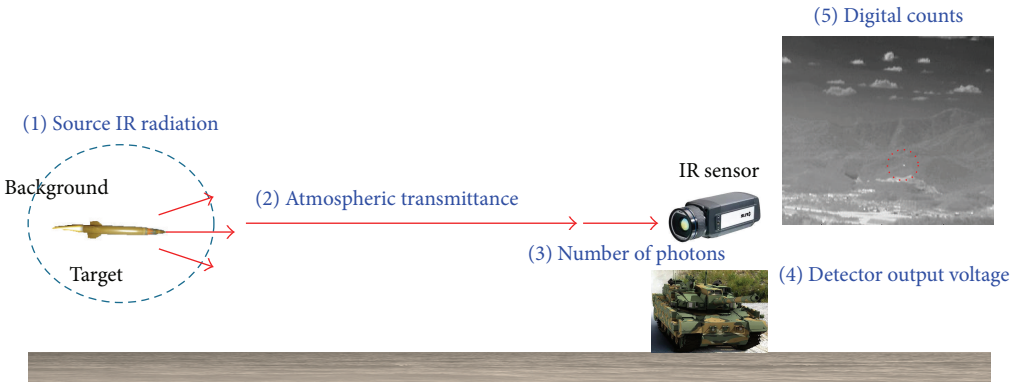


Figure 2: IR target imaging process and sources of IR variations.

of electromagnetic radiation that passes through a medium without being absorbed or reflected). A_0 [m^2] represents the area of entrance aperture of the IR camera. $\tau_O(\lambda)$ represents the optical transmissivity of the IR camera, which considers the mirrors and lenses. Ideally, the thermal energy of a distant point target should be gathered in a pixel of a detector. However, thermal energy of a point target is dispersed (blurred) due to the diffraction and aberration of the optical system. PVF (point visibility factor) can model such phenomenon quantitatively. So, PVF is defined as the ratio of center pixel energy over total target energy. λ_1, λ_2 represent operation ranges, lower limit, and upper limit of thermal detector. $R_d(\lambda)$ denotes detector responsivity versus wavelength. If the responsivity is available, then we can estimate the output voltage given infrared intensity. G denotes the detector gain. We usually use radiant intensity ΔI [watt/ sr] by integrating $I(\lambda, \Delta T)$ with $\lambda_1 \sim \lambda_2$. Furthermore, if we use average transmissivity, optics transmissivity, and detector responsivity within wavelength, then (1) is simplified as the following equation:

$$\Delta V_{\text{pnt}} = \Delta I \times T_{\text{atm}} \times \frac{A_0}{R^2} \times \tau_O \times \text{PVF} \times R_d \times G. \quad (2)$$

Given radiant intensity (ΔI) of target-background, the total number of photons is calculated by dividing the radiant intensity by energy per photon as in the following equation:

$$\Delta E = \frac{\Delta I}{hc/\lambda_{\text{center}}}, \quad (3)$$

where $h = 6.626 \times 10^{-34}$ [Js] denotes Plank's constant, $c = 3 \times 10^8$ [m/s] denotes the speed of light, and $\lambda_{\text{center}} = (\lambda_1 + \lambda_2)/2$ means center wavelength.

The responsivity and gain are determined by quantum efficiency (η_Q), electron charge (q [C]), integration time (T_i [s]), dark current (I_{dark}), and equivalent capacitance (C_{eq}) in the readout integrated circuit. If we use this information, then (2) is rewritten as the following equation:

$$\Delta V_{\text{pnt}} = \left(\Delta E \times T_{\text{atm}} \times \frac{A_0}{R^2} \times \tau_O \times \eta_Q \times q + I_{\text{dark}}\right) \times \text{PVF} \times \frac{1}{C_{\text{eq}}}. \quad (4)$$

Since the dark current in a cooled thermal detector is so small, it can be removed in the above. In addition, the estimated PVF is reflected by Gaussian filtering in the image domain. So, the final form for point source is simplified as the following equation:

$$\Delta V_{\text{pnt}} = \left(\Delta E \times T_{\text{atm}} \times \frac{A_0}{R^2} \times \tau_O \times \eta_Q \times q\right) \times \frac{1}{C_{\text{eq}}}. \quad (5)$$

The atmospheric transmissivity is calculated using MODerate resolution atmospheric TRANsmission (MODTRAN) and Beer's law [6] according to the target distance. In Beer's law, atmospheric transmittance is defined as $T_{\text{atm}} = e^{-\gamma R}$, where τ denotes attenuation coefficient (km^{-1}). If the target distance (R) is larger than 20 km then Beer's law is used. Otherwise, we use MODTRAN to estimate atmospheric transmissivity.

Let us assume that a target of ΔI [watt/sr] is at distance R [m]. If the projected target size is smaller than 1 pixel, then we use (5) to calculate the difference voltage output. Digital value of difference voltage is obtained as (6) by considering the bit resolution (m [bit]) in the A/D converter and voltage dynamic range ($\Delta V_{\text{dynamic}}$):

$$\Delta D = \frac{2^m}{\Delta V_{\text{dynamic}}} \cdot \Delta V. \quad (6)$$

2.2. Sources of Signature Variations. The radiation contrast between the target and background can be used to target detection. However, it is challenging problem due to the dynamic behavior of the radiation contrast (ΔE) and atmospheric transmissivity (T_{atm}). According to Jacobs's analysis [1], IR signature variations are generated by the target conditions, environmental variations, and material properties. The target conditions included exhaust grid/gases, crew compartment heating/cooling, power generator, material properties, camouflage, target location, and orientation. Because the targets (HEAT) in APS are small (length: 1 m, diameter: 0.1 m) and only incoming targets are considered, the variations caused by the target conditions were not considered in the present study. The environmental variations include the induced weather (sun, clouds, rain, and snow), atmospheric

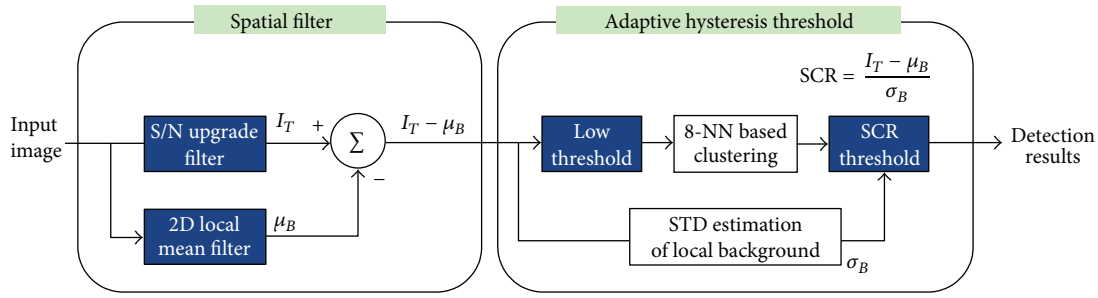


Figure 3: Spatial filter and adaptive hysteresis threshold-based small infrared target detection system.

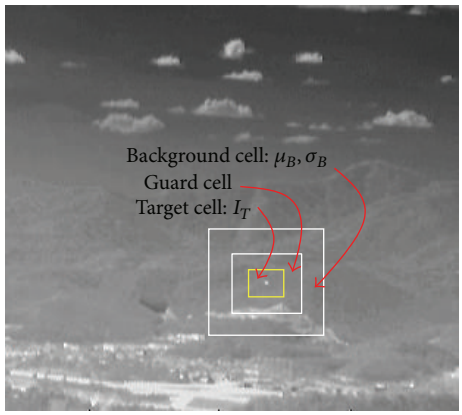


Figure 4: SCR estimation, target cell, guard cell, and background cell are selected using the clustering information.

effects on transmission, and the geographical location. In this study, IR variations caused by 24-hour weather in winter and spring were considered. The material properties can be another source of IR variations. Different materials in different backgrounds exhibit different IR signatures. For example, there are trees in remote mountains, concrete buildings in urban areas, soil and grass in near fields, and air/cloud in the sky. The material properties are related to the radiation contrast between target and background. Environmental weather conditions are related to the atmospheric transmissivity.

3. Variation Analysis of Infrared Target Signature

3.1. Basic Parameters of Small Target Detection. In the infrared small target detection community, background subtraction-based approaches are well established and embedded in military systems. In 2011, Kim proposed a modified mean subtraction filter (M-MSF) and a hysteresis threshold-based target detection method [7]. As shown in Figure 3, an input image ($I(x, y)$) is pre filtered using Gaussian coefficients ($G_{3\times3}(x, y)$) to enhance the target signal and reduce the level of thermal noise according to the following equation:

$$I_T(x, y) = I(x, y) * G_{3\times3}(x, y). \quad (7)$$

The signal-to-clutter ratio (SCR) is defined as (max target signal − background intensity)/(standard deviation of background). Simultaneously, the background image ($\mu_B(x, y)$) is estimated using a 11×11 moving average kernel ($\mathrm{MA}_{11\times11}(x, y)$) as the following equation:

$$\mu_B(x, y) = I(x, y) * \mathrm{MA}_{11\times11}(x, y). \quad (8)$$

The pre-filtered image is subtracted from the background image, which produces a target-background contrast image ($I_{\mathrm{contrast}}(x, y)$), as expressed in (9). The modified mean subtraction filter (M-MSF) can upgrade the conventional mean subtraction filter (MSF) in terms of false detection:

$$I_{\mathrm{contrast}}(x, y) = I_T(x, y) - \mu_B(x, y). \quad (9)$$

The last step of small target detection is how to determine which pixels correspond to the target pixels. Kim proposed an adaptive hysteresis thresholding method, as shown in Figure 3. A contrast threshold (k_1) is selected to be as low as possible to locate the candidate target region. The 8-nearest neighbor (8-NN) based clustering method is then used to group the detected pixels. The SCR threshold (k_2) is selected properly to meet the detection probability and false alarm rate, as expressed in (10). A probing region is declared as a target if

$$\mathrm{SCR}(x, y) = \frac{I_T(x, y) - \mu_B(x, y)}{\sigma_B(x, y)} > k_2, \quad (10)$$

where μ_B and σ_B represent the average and standard deviation of the background region, respectively. k_2 denotes the user defined parameter used in the control detection rate and false alarm rate. A probing region is divided into the target cell, guard cell, and background cell, according to the results of contrast thresholding and clustering, as shown in Figure 4. Therefore, the key parameters of small target detection are the SCR-related terms, such as the average background intensity (μ_B), target intensity (I_T), and standard deviation of the background (σ_B). In the SCR computation, the target-background contrast parameter ($I_T - \mu_B$) can be derived from the key parameters, which should be analyzed according to the IR variations.

3.2. Acquisition of Infrared Images

Measurement Devices. The objective of this study was to evaluate the small target detection parameters ($I_T, \mu_B, I_T -$

Figure 5: Overview of the target and sensor positions.

μ_B, σ_B) over a 24-hour period for a range of backgrounds, such as the remote mountain, building, near field, and sky. As shown in Table 1, a LWIR camera, CCD camera, DSLR camera, thermal target, and IR thermometer were used to record the target and environmental information. The FLIR Tau302 can record digital IR data with a 14-bit resolution. The SONY NEX-VG20 can record visible data with HD video resolution. The 5D Mark II was used to record the overall experimental status. The BMH-30 is a thermal target to simulate the plume of an antitank missile. The normal temperature of the target is approximately 450°C. The DT-8865 can measure the temperature of the targets and backgrounds with the range of −50 to 1000°C.

Location and Meteorological Data. In APS, the incoming antitank missiles should be detected at a distance of at least 1 km and then tracked for the following hard killing process. Because the focus of this study was to analyze the effects of the day/night changes on the SCR parameters for different backgrounds, a small region was selected within the campus, as shown in Figure 5. The distance between sensors and a target is around 300 m. The sensing point is selected carefully to include a variety of backgrounds, such as remote mountains, buildings, near field, and sky. During the recording, the meteorological measurement data from Korea Meteorological Administration web site (KMA, http://web.kma.go.kr) was also checked, as shown in Table 2 (winter) and Table 3 (spring). The tables consist of the recording time, overall weather, visibility, cloud, temperature, and relative humidity. In winter, the overall weather was clear with a temperature and humidity range of −3.0°C~4.8°C and the humidity range of 14%~66%. In spring, the overall weather was cloudy with a temperature and humidity range of 5.6°C~15.4°C and the humidity range of 27%~75%.

Based on the measured weather, the atmospheric transmittance can be simulated using the MODTRAN (MODerate resolution atmospheric TRANsmission, http://modtran5.com/) program designed to model the atmospheric propagation of electromagnetic radiation for the 0.2 to 100 μm spectral range. The simulation parameters were selected, as shown in Figure 6(a). Figure 6(b) shows the corresponding transmittance according to distance. The transmittance was evaluated from 100 m to 1200 m at 100 m intervals because small targets should be detected and tracked at 1200 m. The transmittance at 1200 m was 0.76 and increases to 0.95 at 100 m. Note that the transmittance was quite high, so the signal attenuation could be negligible due to the atmosphere.

Examples of Acquired Images. The recording area was selected to include sky, mountain, building, and field backgrounds, as shown in Figure 7, where the locations of a target and cameras are indicated. Over a 24-hour period, a pair of LWIR and CCD images was recorded at 1-hour intervals. In LWIR, both digital 14-bit data and contrast enhanced image were acquired for variability analysis. Figures 8 and 9 give partial examples of a 24-hour recording in winter and spring, respectively. The scene temperatures are indicated in the LWIR images and the recording times are displayed on the CCD images.

3.3. Variability Analysis of Infrared Images

Analysis Factors. The key parameters in small target detection are the pre-filtered input image (I_T), estimated background image (μ_B), standard deviation map (μ_B). The contrast data

Table 1: Measurement sensors used during 24-hour recording.

Device	Model	Specification	Comment
LWIR camera	FLIR Tau320 	Spectral band: 7.5–13.5 μm FPA: 324 × 256 Output: 14-bit digital NETD: <50 mK FOV: 18°× 14° (25 mm lens)	IR measurement
CCD camera	SONY NEX-VG20 	Spectral band: visible FPA: 16.1 M Exmore APS-HD CMOS Output: 1920 × 1080, 60 p FOV: 18–200 mm lens	Visible measurement
DSLR camera	5D Mark II 	Spectral band: visible FPA: 21.1 M CMOS FOV: 24–70 mm F2.8L	Recording environments
Thermal target	BMH-30 	Type: gas heater Temperature: 450°C Diameter: 160 mm	Simulating antitank missile
IR thermometer	DT-8865 	Type: IR sensor (8–14 μm) Range: −50~1000°C Resolution: 0.1°C Response time < 150 ms	Measuring target/environment temperature

Table 2: Measured meteorological data of the recording place (winter).

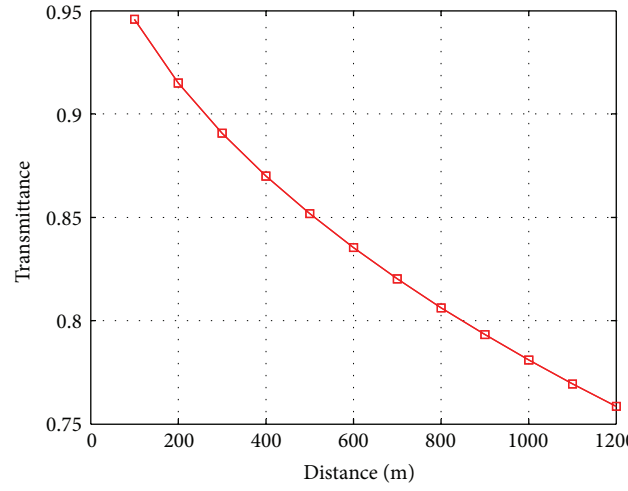
Time	Weather	Visibility (km)	Cloud (1/10)	Temperature (°C)	Humidity (%)
2013.01.18.10 H	Clear	11	0	−2.3	45
2013.01.18.11 H	Clear	11	0	−0.6	31
2013.01.18.12 H	Clear	15	0	1.5	25
2013.01.18.13 H	Clear	15	0	2.8	18
2013.01.18.14 H	Clear	18	0	3.9	15
2013.01.18.15 H	Clear	20	0	4.8	14
2013.01.18.16 H	Clear	20	0	4.6	14
2013.01.18.17 H	Clear	20	0	3.6	16
2013.01.18.18 H	Clear	18	0	2.2	19
2013.01.18.19 H	Clear	20	0	1.4	22
2013.01.18.20 H	Clear	20	0	0.6	26
2013.01.18.21 H	Clear	20	0	0.3	27
2013.01.18.22 H	Clear	20	0	−0.2	32
2013.01.18.23 H	Clear	20	0	−1.2	47
2013.01.19.00 H	Clear	20	0	−1.2	52
2013.01.19.01 H	Clear	20	0	−1.4	56
2013.01.19.02 H	Clear	20	0	−1.5	56
2013.01.19.03 H	Clear	13	0	−2.1	60
2013.01.19.04 H	Clear	13	0	−2.3	61
2013.01.19.05 H	Clear	11	0	−2.5	63
2013.01.19.06 H	Clear	11	0	−2.5	62
2013.01.19.07 H	Cloudy	11	4	−2.9	65
2013.01.19.08 H	Cloudy	15	7	−3.0	66
2013.01.19.09 H	Cloudy	18	8	−2.6	65

Table 3: Measured meteorological data of the recording place (spring).

Time	Weather	Visibility (km)	Cloud (1/10)	Temperature (°C)	Humidity (%)
2013.04.19.11 H	Cloudy	23	5	12.8	32
2013.04.19.12 H	Cloudy	23	3	14.9	29
2013.04.19.13 H	Cloudy	23	5	14.5	28
2013.04.19.14 H	Clear	25	2	15.2	28
2013.04.19.15 H	Clear	25	2	15.4	27
2013.04.19.16 H	Clear	25	2	14.8	28
2013.04.19.17 H	Clear	25	4	13.5	30
2013.04.19.18 H	Cloudy	25	6	12.0	34
2013.04.19.19 H	Cloudy	23	6	10.3	39
2013.04.19.20 H	Cloudy	23	6	9.5	41
2013.04.19.21 H	Cloudy	20	6	9.1	43
2013.04.19.22 H	Cloudy	20	7	8.8	44
2013.04.19.23 H	Cloudy	20	7	8.8	44
2013.04.20.00 H	Cloudy	20	7	8.8	44
2013.04.20.01 H	Cloudy	20	7	8.8	43
2013.04.20.02 H	Cloudy	20	7	8.7	43
2013.04.20.03 H	Cloudy	20	10	8.6	46
2013.04.20.04 H	Cloudy	20	10	8.5	47
2013.04.20.05 H	Cloudy	20	10	8.5	48
2013.04.20.06 H	Cloudy	20	10	8.6	48
2013.04.20.07 H	Cloudy	20	10	8.5	50
2013.04.20.08 H	Cloudy	20	10	7.5	60
2013.04.20.09 H	Rainy	9	10	5.6	75

MODTRAN parameters	Value
Observation height (m)	100
Band (μm)	7.5–13.49
Atmosphere	1976 US standard
Extinction	Rural-VIS-23 km
Type of atmospheric path	Horizontal path

(a)



(b)

Figure 6: MODTRAN simulation of the recording environment: (a) selected MODTRAN parameters, (b) transmittance versus distance.

and SCR values can be derived based on these three key parameters. Figure 10 shows the flow of the SCR computation of a test image. The test image was obtained from 14-bit raw data and the bright spot represents a gas heater (Figure 10(a)). SNR enhanced image can be obtained by applying a matched filter to the input image (Figure 10(b)). The background image can be estimated using a local mean filter with a 11 × 11 moving average kernel (Figure 10(c)). A contrast image can be obtained by subtracting the estimated background image from the pre-filtered input image (Figure 10(d)). A

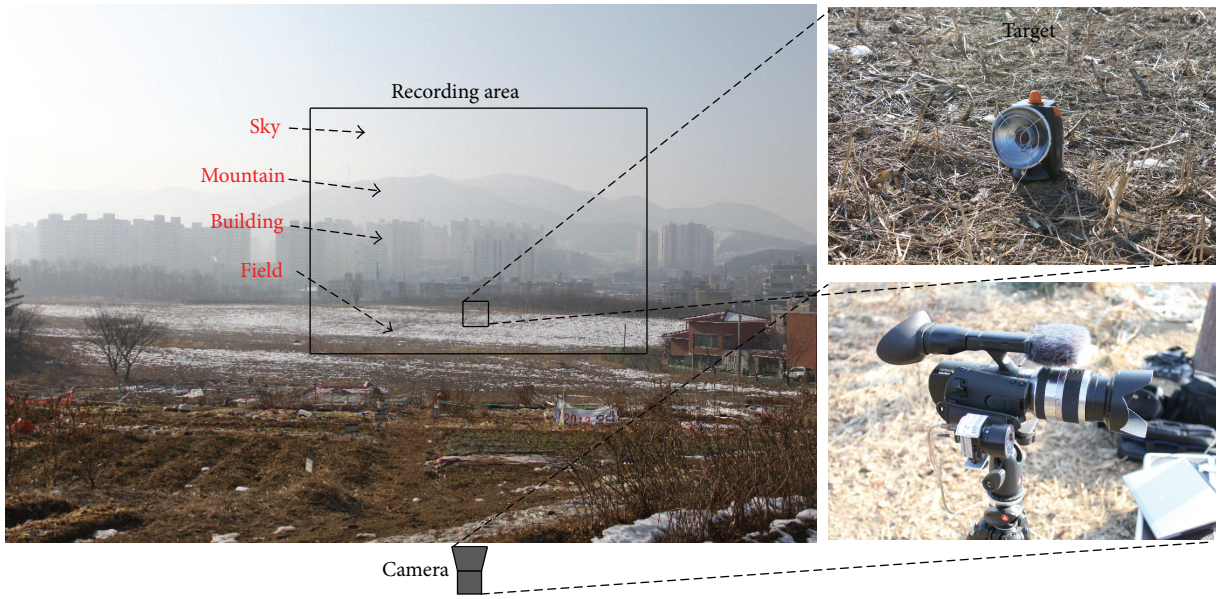


Figure 7: Recording area and locations of the target and cameras.

standard deviation map was estimated from the contrast image (Figure 10(e)). The final SCR map was generated using the contrast result and standard deviation map (Figure 10(e)).

The SCR-related parameters should be analyzed on different backgrounds because antitank missiles can exist anywhere. The backgrounds were classified as natural (sky, remote mountain, near field) and artificial background (man-made buildings). Figure 11 shows the corresponding regions indicated by the rectangles. Because this study was interested in the background variation effects on the SCR values, a fixed target signal, such as 7317 obtained by averaging the target intensities, was used. This assumption is reasonable because the target distance is within 1 km and the signal attenuation is negligible. For each pixel position $((i, j))$, the mean background intensity $(\mu_B(i, j))$, standard deviation $(\sigma_B(i, j))$, and SCR value $(\mathrm{SCR}(i, j))$ were calculated. A representative value for each region was obtained by averaging the corresponding values.

Variation Analysis Results. The effects of the recording time, temperature, and humidity on the SCR parameters were analyzed. From the analysis, the optimal small target detection time, temperature, and humidity conditions were obtained for different backgrounds. In addition, the evaluated data was used to control the detection thresholds to achieve a predefined detection rate.

In the first inspection, the SCR parameter variations were analyzed according to the recording time. In winter season, the recording time started at 10 a.m. and ended at 9 a.m. on the next day with a 1-hour recording interval. Figure 12(a) shows the average background intensity (μ_B) variations for the four types of backgrounds over a 24-hour period. The background intensities were relatively high during the day and low during the night. Sky background showed very low intensity and increased when a cloud appeared (09 H). Given a fixed target intensity, the contrast data can be calculated as shown in Figure 12(b). The contrast values showed the lowest value at noon and fluctuated during the night. The sky background showed the highest contrast all the time. During the day (10 H–15 H), the contrast magnitudes were as follows: sky > mountain > building > near field. During the night, the order of the changes was sky > near field > mountain > building. The contrast of the building and target decreased at night because humans use energy to warm rooms. The variations in the clutter level can also be checked using the standard deviation of the background, as shown in Figure 12(c). According to the graph, the clutter level increased during the day and decreased during the night. The standard deviation of the sky background almost showed the lowest values but increases when a cloud appeared (03 H, 08 H). The near field showed strong clutter during the day and evening. The building background showed almost constant clutter during the entire day. The clutter level of the mountain background showed a peak at noon and decreased. The final SCR versus time curve can be obtained from the above parameter variations, as shown in Figure 12(d). The mountain and near field background showed increasing SCR values according to the time and the building background showed an almost constant SCR curve. The sky background represents very high SCR values and decreased abruptly when the cloud appeared (03 H, 08 H). As higher SCR values would ensure a higher detection rate, the best operating time can be predicted for each background from the curve.

In spring season, the recording conditions are the same as the winter season except the starting time (we started at 11 a.m.). Figure 13(a) shows the average background intensity (μ_B) variations for the four types of backgrounds over a 23-hour period. The background intensities were relatively high during the day and low during the night except the sky background. It showed very low intensity and increased when

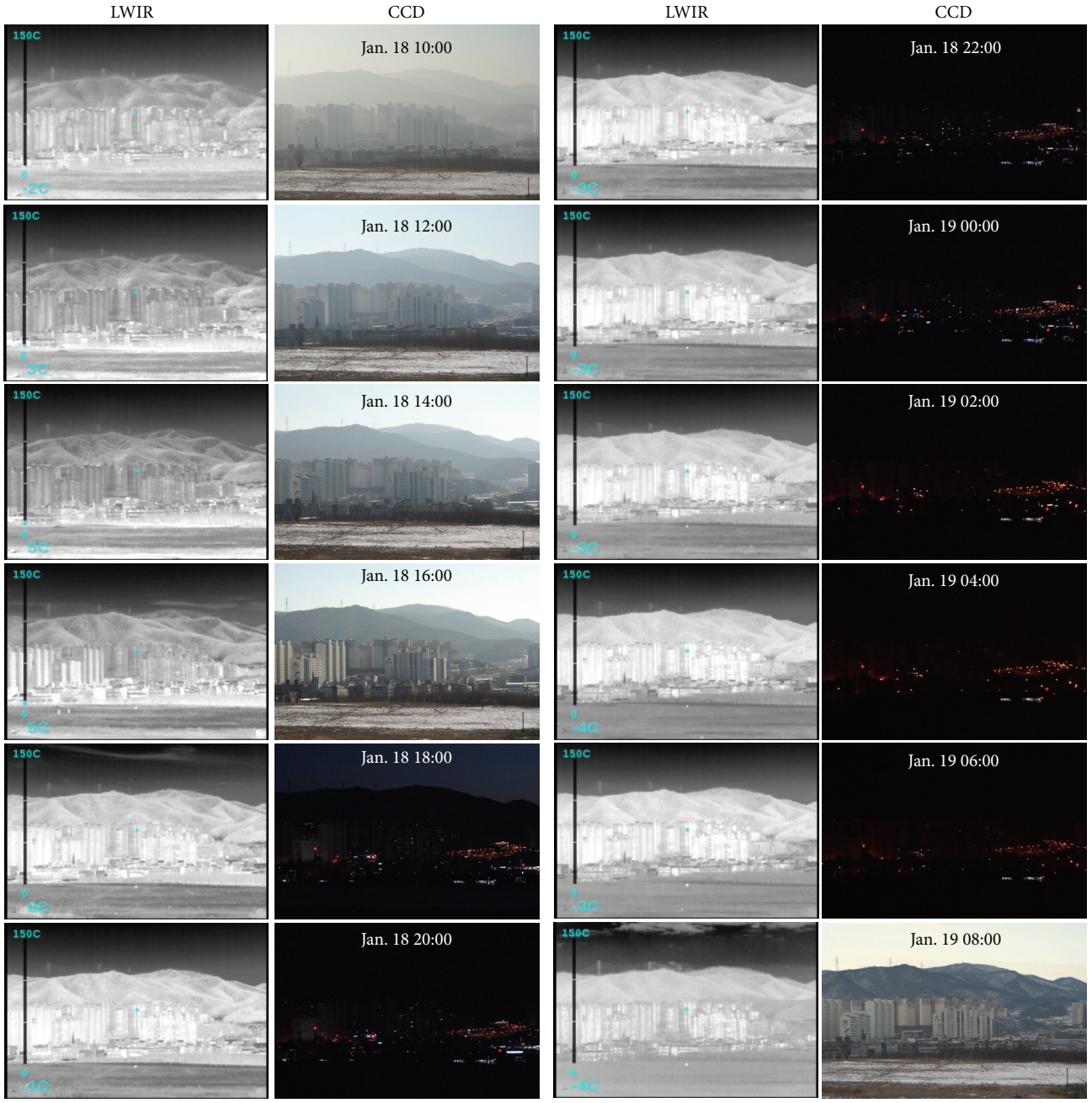


Figure 8: Partial examples of the acquired LWIR and CCD images over 24-hour period in Winter.

a cloud appeared (after 21 h). Given a fixed target intensity, the contrast data can be calculated as shown in Figure 13(b). The contrast values showed the lowest value at 15 h and remained constant during the night. The sky background showed the highest contrast all the time as the winter case. During the day (11 H–17 H), the contrast magnitudes were as follows: sky > mountain > building > near field. During the night (18 H–08 H), the order of the changes was sky > mountain ≥ near field > building. The contrast of the mountain and near field shows quite similar values and patterns. The variations in the clutter level can also be checked using the standard deviation of the background, as shown in Figure 13(c). According to the graph, the clutter level is high during the day and decreased during the night. The order of clutter level during night was building > mountain > near field > sky. The final SCR versus time curve can be obtained from the above parameter variations, as shown in Figure 13(d). The mountain, near field, and building background showed increasing SCR values according to the time and the sky background represents very high SCR values and fluctuated abruptly according to the

Figure 9: Partial examples of the acquired LWIR and CCD images over 24-hour period in spring.

level of cloud. Note that the SCR of building background was almost constant in winter and increased in spring during 23 hours.

If we compare both data (winter, spring), we can find interesting results as shown in Figure 14. Figures 14(a) and 14(b) represent the temperature and humidity variations according to time. The temperature increased during day and decreased during night. Conversely, the humidity decreased during day and increased during night. The temperature of spring is higher than that of winter and the humidity of spring fluctuates less than that of winter. Figures 14(c)–14(f) compare SCR values between winter and spring for the mountain region, building region, near field region, and sky region, respectively. The SCR curves of the mountain and near field show similar patterns in winter and spring season. However, the SCR curves of the building region show different characteristics: almost constant in winter and increasing pattern in spring. The SCR curves of sky region show quite random according to the level of cloud.

In addition, the SCR parameter variations were analyzed according to the temperature. Because the temperatures were recorded at each recording time, the SCR parameters were reordered according to the temperature. As shown in Figures 14(a) and 14(c)–14(f), SCR values of mountain and

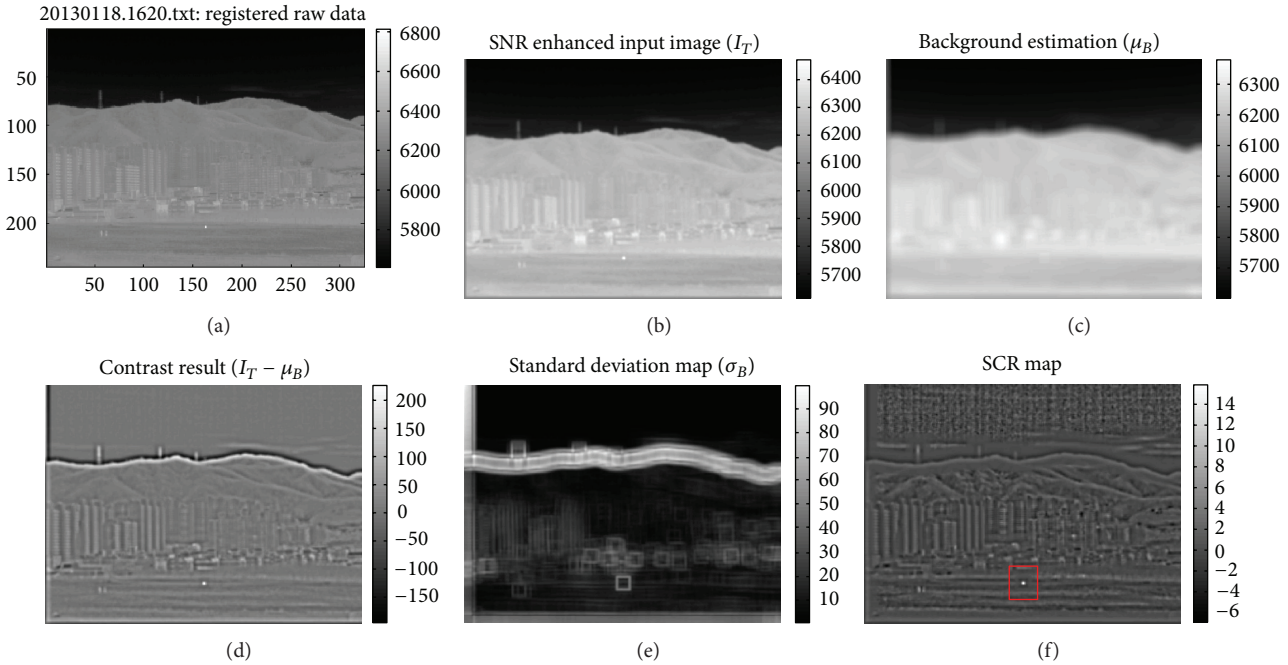


Figure 10: SCR computation flow: (a) registered raw input data (14 bits), (b) SNR enhanced input image, (c) background estimation result, (d) contrast result, (e) standard deviation estimation, and (f) computed SCR map.

Table 4: Summary of the SCR variations according to the time, temperature, and humidity for various backgrounds.

Background	Time (10 H-09 H)	Temperature	Humidity
Mountain	Increase	Decrease	Increase
Near field	Increase	Decrease	Increase
Sky	High, fluctuating	High, fluctuating	High, fluctuating
Building (winter)	Constant	Constant	Constant
Building (spring)	Increase	Decrease	Increase

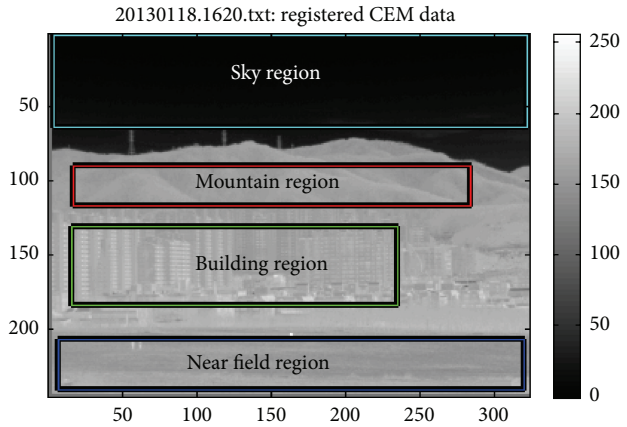


Figure 11: Analysis regions indicated on the image.

near field increased slightly according to the temperature decrease. Those of the sky background do not reveal such phenomena because it is more affected by cloud. Those of the sky background did not reveal such phenomena because it is affected more by cloud. The SCR values of the building background were relatively unaffected by temperature in winter except spring. From above inspection, small targets can be detected better when they are cold (night), particularly in natural backgrounds, such as mountains and near fields.

In the last inspection, the SCR parameter variations were analyzed according to the humidity. Because relative humidity was recorded at each recording time, the SCR parameters could be reordered according to the humidity level. As shown in Figures 14 (b) and 14(c)–14(f), SCR values of mountain and near field increased slightly according to the humidity increase. Those of the sky background did not reveal such phenomena because it is affected more by cloud. The SCR values of the building background were relatively unaffected by the humidity. Table 4 lists the overall evaluations.

4. Small Target Detection Robust to Target Signature Variations

Until now, we discussed the variability of infrared signature in terms of target detection parameter (SCR). A successful small target detection system should work regardless of the signature variations. We can consider two kinds of

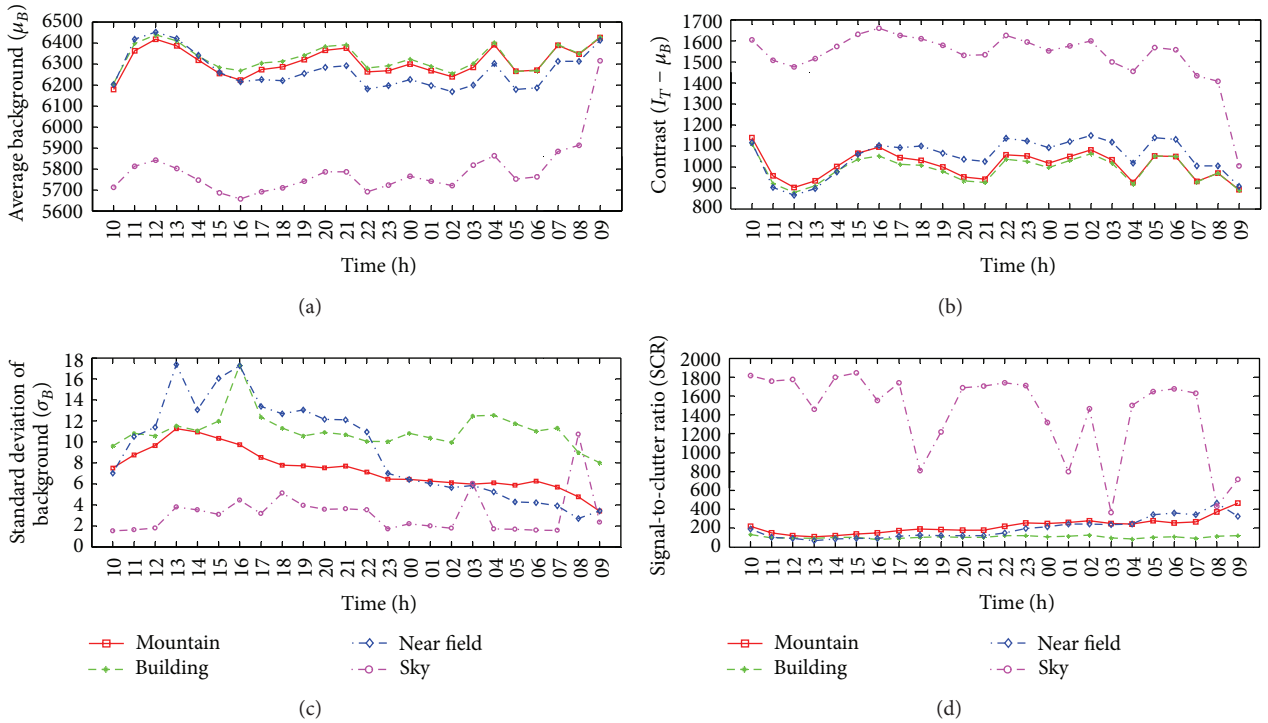


FIGURE 12: Variation analysis of winter according to the recording time: (a) average background versus time, (b) contrast versus time, (c) standard deviation versus time, and (d) SCR versus time.

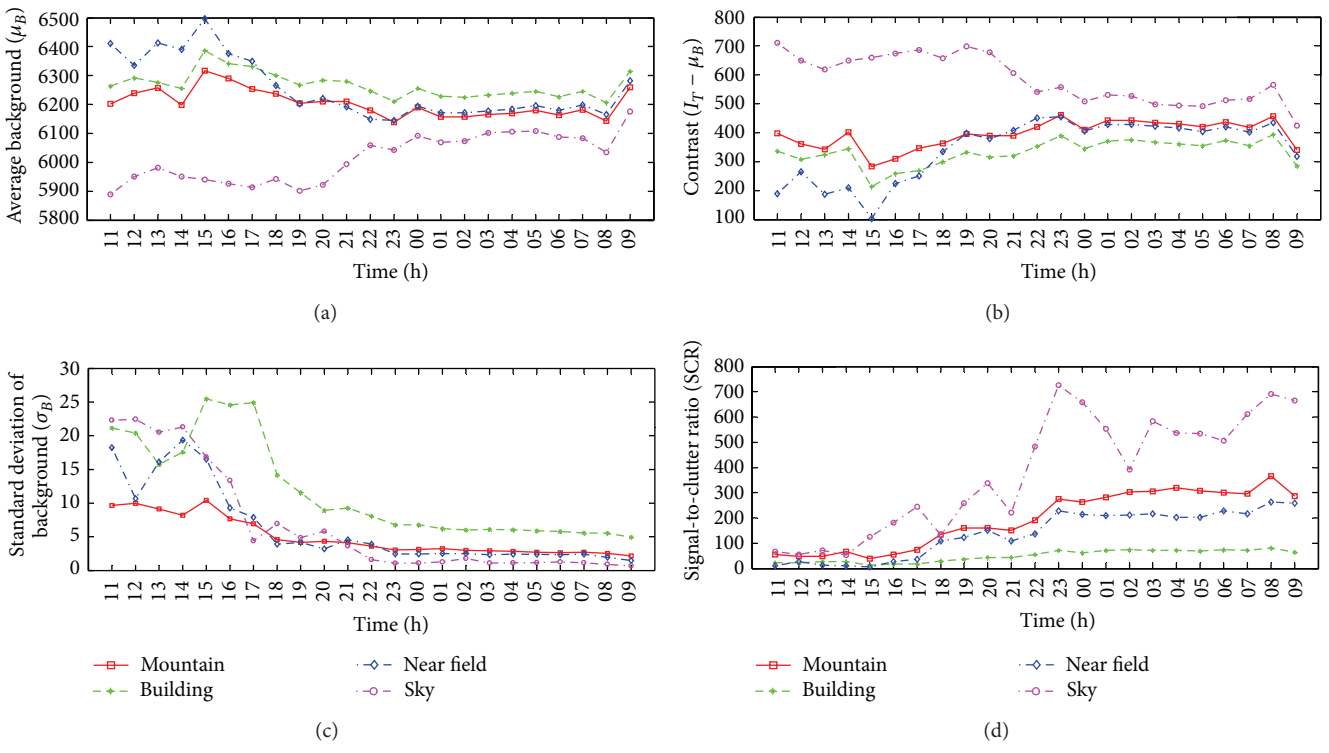


FIGURE 13: Variation analysis of spring according to the recording time: (a) average background versus time, (b) contrast versus time, (c) standard deviation versus time, and (d) SCR versus time.

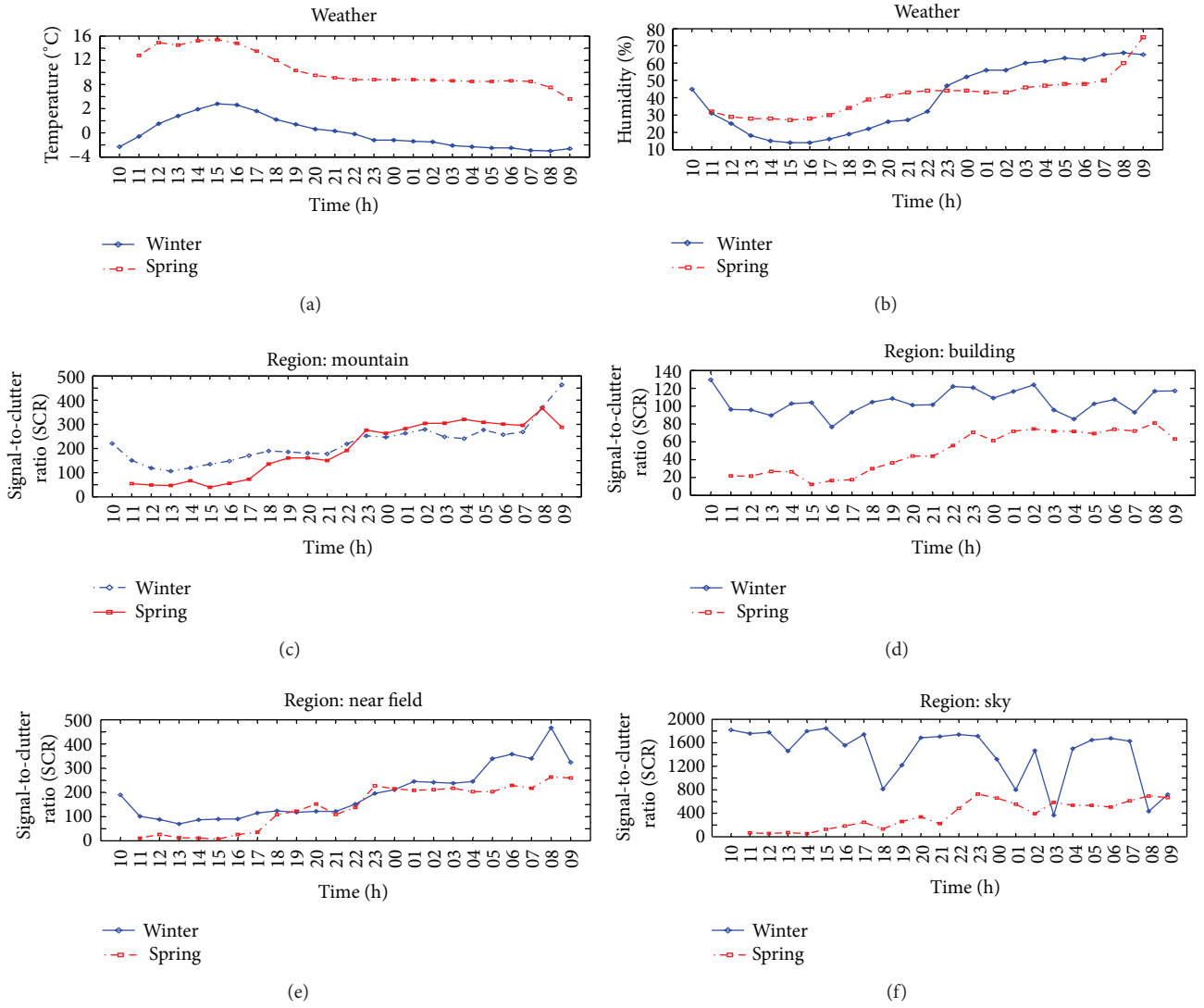


FIGURE 14: SCR comparison between winter and spring: (a) mountain, (b) building, (c) near field, and (d) sky.

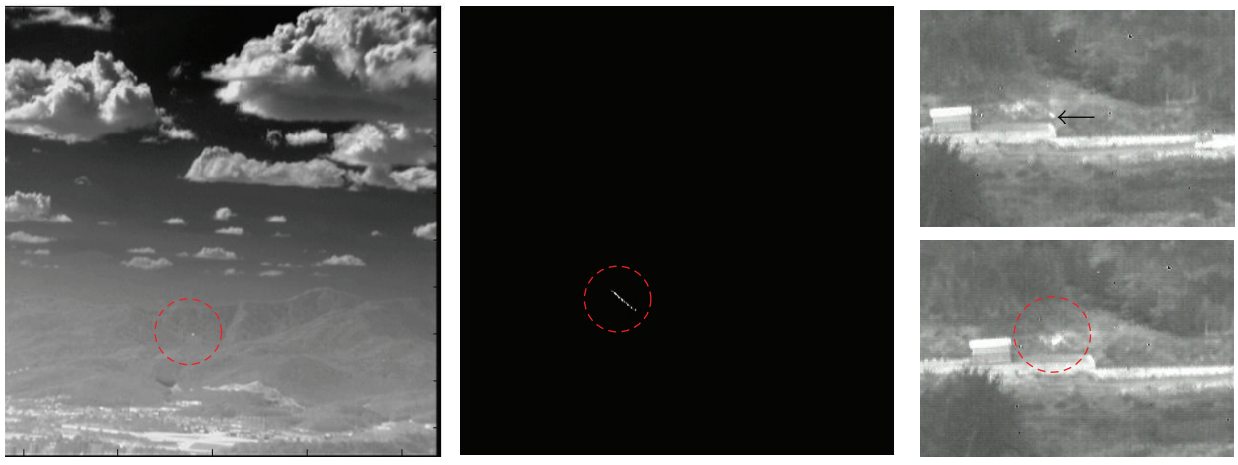


FIGURE 15: Limitations of TVF-based method: (a) ambiguity of target position and subpixel velocity, (b) target missing by the intersection of TVF and spatial filter.

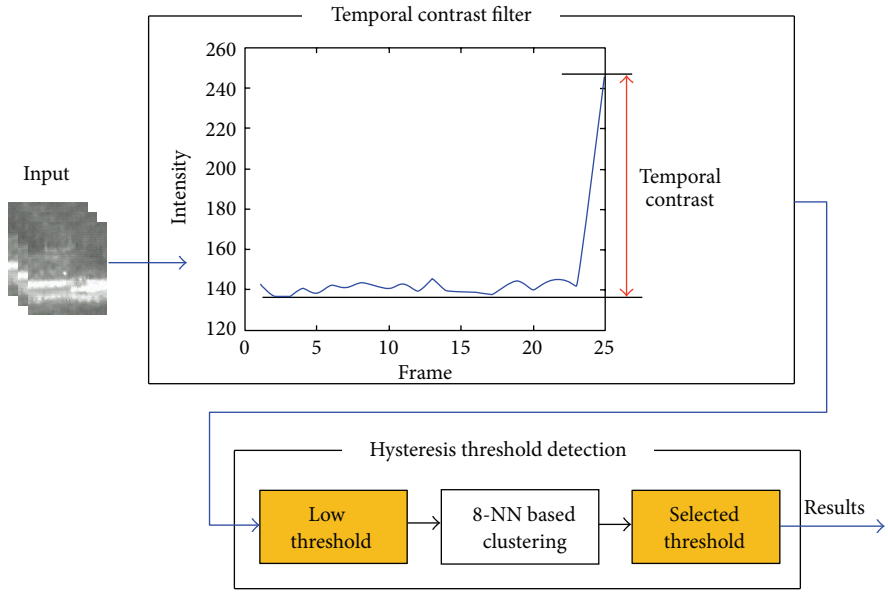


Figure 16: Proposed supersonic target detection system.

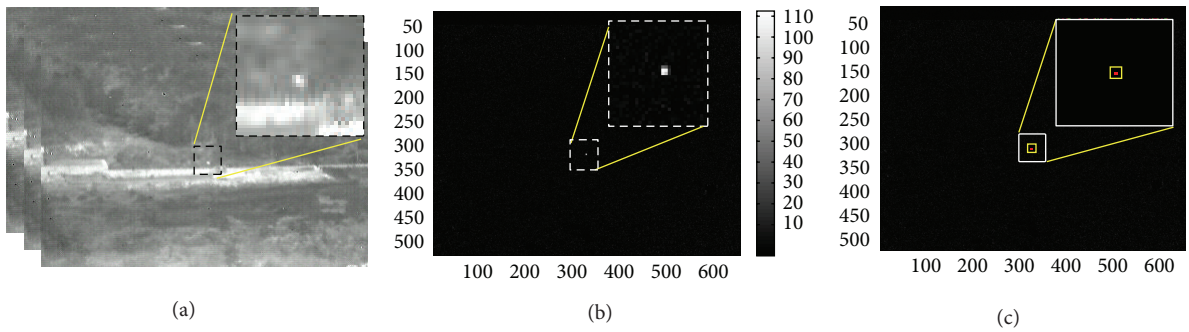


Figure 17: Supersonic target detection procedures with results: (a) input sequence, (b) TCF results, and (c) hysteresis threshold detection results.

overcoming approaches such as knowledge-based adaptive thresholding and robust detection filtering method.

4.1. Knowledge-Based Thresholding. SCR values change enormously according to recording time, temperature, humidity, season, background type, and so on. If we have a lot of databases, we can make a knowledge-based target detection system using (11) to handle the signature variation. An adaptive threshold (Th) is determined by a function with parameters of time (θ_t), temperature (θ_c), humidity (θ_h), season (θ_s), and background type (θ_b). In fact, such approach is time consuming and impractical to realize the knowledge database

$$\mathrm{SCR}(x, y) > \mathrm{Th}(\theta_t, \theta_c, \theta_h, \theta_s, \theta_b). \quad (11)$$

4.2. Robust Target Detection Filter. Background subtraction-based small target detection is sensitive to the IR signature variation and generates a lot of false detections by background clutter. However, if we use robust target detection filter, the problem can be mitigated. In the past decades, a variety of approaches have been developed. Among them, the temporal variance filter (TVF) of temporal profile has been used successfully to detect point targets moving at subpixel velocity [8, 9]. Slowly moving cloud clutter can be removed by subtracting the connecting line of the stagnation points (CLSP) [10]. Recently, the CLSP method is approximated for real time processing [11]. In supersonic missile detection with high frame rate camera, a detection algorithm should be simple but powerful performance of detection rate and localization accuracy to cover a wide range of target velocity (subpixel to pixel velocity).

Because the TVF-based method detects targets based on stripe patterns, it shows high detection performance. However, it has limitations such as the ambiguity of target position and subpixel velocity assumption as shown in Figure 15(a). The ambiguity of target position can be solved by the intersection of TVF and spatial filter, which leads to low detection performance in background clutter as shown in Figure 15(b). We solve these three problems by the hysteresis

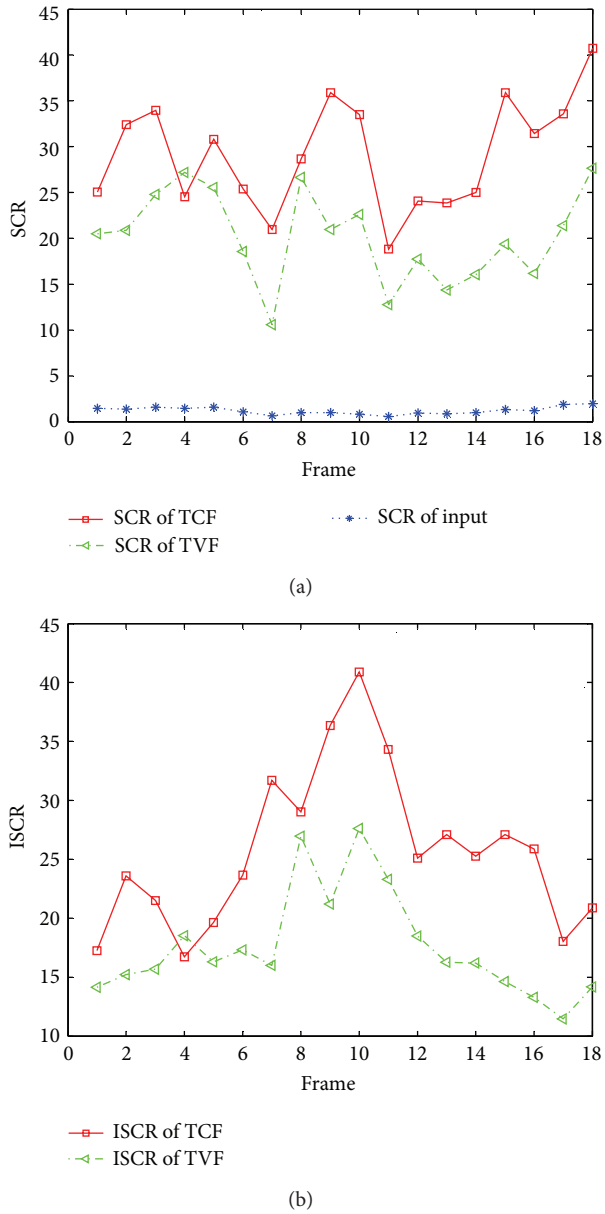


FIGURE 18: Comparative analysis of temporal filters for test sequence (Set 4): (a) SCR values for input, TCF, and TVF; (b) ISCR values for TCF and TVF.

threshold-based detection after the temporal contrast filter (TCF). The TCF can enhance the signatures of moving target pixels and the hysteresis threshold-based detection can localize targets accurately.

TCF-Based Supersonic Target Detection System. The proposed small target detection system consists of TCF part and detection part, as shown in Figure 16. The filtering part conducts the enhancement of target signature by applying the temporal contrast. If $I(i, j, k)$ denotes the intensity of (i, j) pixel at the current kth frame, the TCF at (i, j, k) is defined as (12). We assume that the buffer size is k and $k - 1$ frames are used to estimate background intensity. The key part of the TCF is the background signature estimation by the minimum filter to maximize the signal-to-noise ratio. Because the contrast is produced by the difference of the current intensity and previous intensity of a pixel, we can remove the ambiguity of target location. The rest of the detection system consists of a hysteresis thresholding method. The first threshold is selected to be as low as possible in order to find the candidate target region. Then the 8-nearest neighbor (8-NN) based clustering method is utilized to group the detected pixels. So, we can divide a considering region into a target region and a background region. The adaptive threshold detection is conducted by using (13) where $T_{\max}$ denotes the maximal TCF in a target region and μ_{BG}, σ_{BG} represent the average and standard deviation of the background region, respectively. If the signal-to-clutter ratio (SCR_{temp}) is larger than a predefined threshold t, the current considering region is declared as a detected target. Figure 17 summarizes the overall detection flow. Note that a moving target is enhanced by the proposed TCF and detected by the hysteresis threshold

$$\mathrm{TCF}(i, j, k) = I(i, j, k) - \min_{n=1,2,\ldots,k-1} I(i, j, n), \quad (12)$$

$$\mathrm{SCR}_{\mathrm{temp}} = \frac{T_{\max} - \mu_{BG}}{\sigma_{BG}}. \quad (13)$$

5. Experimental Results

We use the TVF as a baseline filtering method [8]. Targets are detected using the same hysteresis thresholding method. In addition, we compare the TCF with the modified TVF (modTVF) which uses both TVF and spatial filter (mean subtraction filter) to localize targets. We prepared two synthetic image sequences and two real target sequences (F-15, Metis-M) with frame rate of 120 Hz. The synthetic sequences are generated using the physics-based method [12]. A test target of Mach 3 is inserted in real ground clutter image with incoming path (Set 1) and passing-by path (Set 2). The first real sequence consists of four F-15s with dynamic motion in strong cloud clutter (Set 3). The second real sequence contains a real antitank missile (Metis-M) incoming near the IR camera (Set 4, Cedip, LWIR, 120 Hz). We evaluated the proposed method in terms of target detection performance as well as filtering performance. The filtering performance can be measured by the improvement of SCR (ISCR) that is defined as SCRout/SCRin. As shown in Figure 18, the proposed method outperforms the other in terms of ISCR for the test Set 4. Table 5 summarizes the statistical performance comparisons of the proposed TCF, TVF, and modTVF in terms of detection rate and false alarm rate. We use the same temporal threshold ($t = 7$) and buffer size ($k = 5$) for fare comparisons. According to the results, the proposed temporal filter produces a much higher number of correct detections and lower localization errors than those of other methods. Figure 19 shows small moving target detection results of cluttered images where the small rectangles represent detection. As indicated by the arrows, the TVF showed inaccurate target localizations due to the stripe patterns and the modTVF often missed true targets in clutter such as cloud edge and ground. Note the superior detection performance of the TCF-based

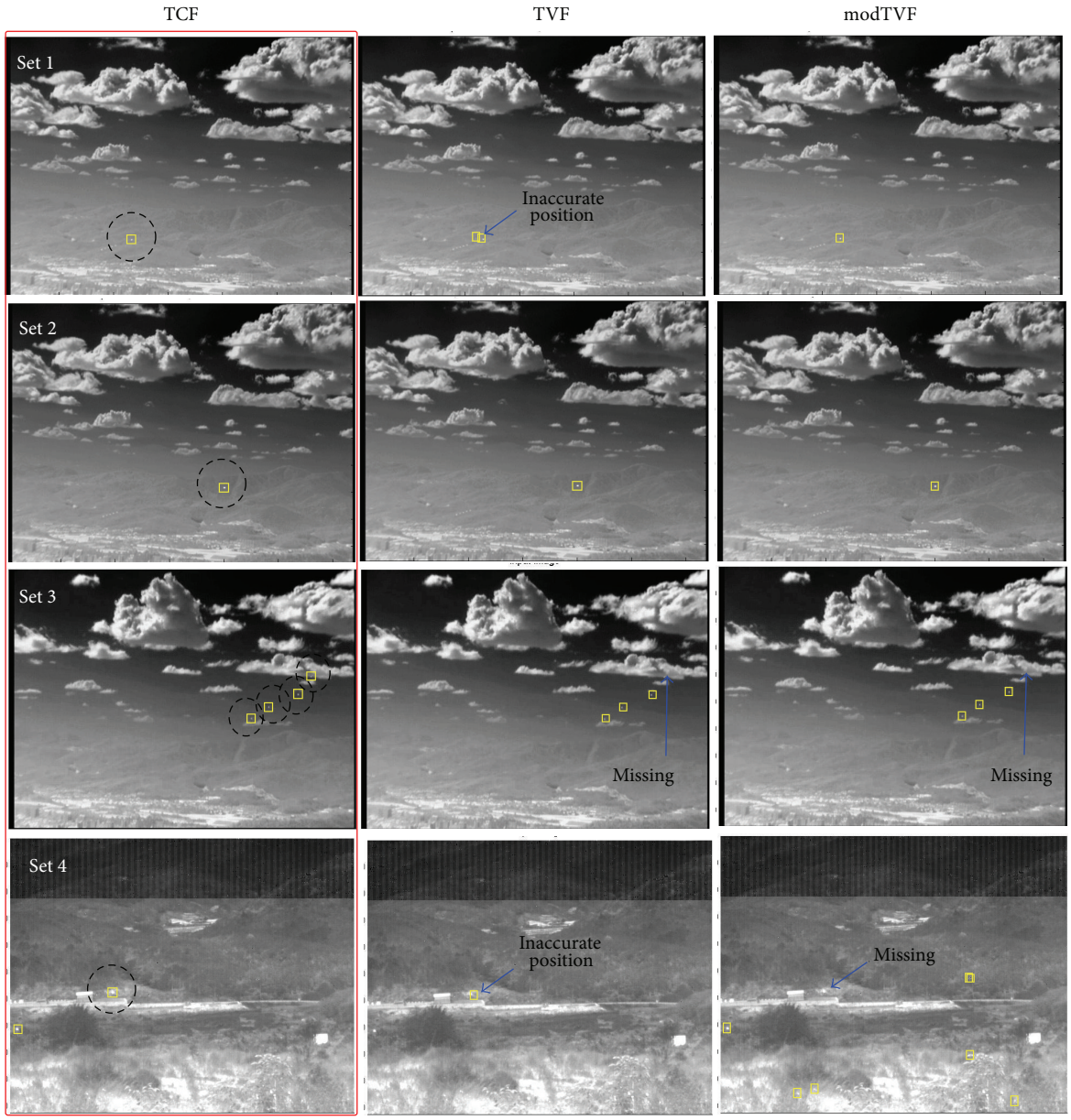


Figure 19: Performance comparison examples of small target detection methods for the four kinds of test sets. Circles represent ground truth and rectangles represent detected results.

Table 5: Statistical performance comparisons of small infrared target detection methods (DR: detection rate, FAR: number of false alarms per frame, PE: position error).

Method	Performance measure	Set 1: syn. incoming	Set 2: syn. passing by	Set 3: real F-15 Multi	Set 4: real metis-M
TCF	DR (%)	100 (100/100)	100 (11/11)	98.78 (3,161/3,200)	99.13 (114/115)
	FAR (number/frame)	0 (0/100)	0 (0/11)	0.0025 (2/811)	0.4 (46/115)
	PE (pixel)	0.13	0.12	0.15	0.116
TVF	DR (%)	97 (97/100)	100 (11/11)	92.19 (2,950/3,200)	89.57 (103/115)
	FAR (number/frame)	0.6 (60/100)	4 (44/11)	0 (0/811)	0.16 (18/115)
	PE (pixel)	5.61	0.28	4.85	4.33
modTVF	DR (%)	100 (100/100)	100 (11/11)	91.09 (2,915/3,200)	36.52 (42/115)
	FAR (number/frame)	0.69 (69/100)	0 (0/11)	0.011 (9/811)	1.91 (220/115)
	PE (pixel)	0.15	0.14	0.17	0.17

method. The supplementary material on the web shows the demo sequences of target detection in Set3.

6. Conclusions

An analysis of the effects of environment in IR-based small target detection is very important. In this study, the 24-hour IR data was recorded in winter and spring, and the IR variations were analyzed in terms of the small target detection parameters particularly the signal-to-clutter ratio (SCR), which is the first trial in this area. SCR variations were analyzed with regard to the recording time, temperature, and humidity. According to the analysis, the natural backgrounds, such as mountains and near field, behave similarly. The SCR values increased during the recording time (10 H–09 H) in these regions. In addition, the SCR values decreased with increasing temperature and humidity. The SCR values of the sky background were quite high and did not show a specific pattern but were affected strongly by cloud. The SCR values of the man-made background, such as buildings, were almost constant regardless of the recording time, temperature, and humidity except spring. Overall, the best conditions can be determined for optimal small target detection or for predicting the small target detection performance under different weather conditions and backgrounds.

In terms of optimal target detection, IR signature variations should be considered to obtain desirable target detection rate and false alarm rate. If the background-related SCR variations are used, the small target detection system can be upgraded by controlling the detection thresholds adaptively depending on the background and weather conditions. However, such approach is impractical because it requires huge number of IR databases according to environmental parameters. On the other hand, we can overcome the IR variation by proposing a robust method. This paper proposed a new simple but powerful supersonic small target detection method by the novel temporal contrast filter. As validated by a set of experiments, it can effectively find and localize true targets with the velocity from subpixel to pixel per frame for various clutter images including cloud and ground clutter. Due to the simplicity of the algorithm with powerful detection capability, the proposed method can be used for real-time military applications for staring infrared cameras.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by a Grant (UD1200630D) of the Agency for Defense Development (ADD) and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (no. 2012-0003252), Korea. This work was also supported by the STRL (Sensor Target Recognition Laboratory) program of Defense Acquisition Program Administration and Agency for Defense Development, Korea. The above foundations have no conflicting interests and recommend journal publications.

References

[1] P. A. Jacobs, *Thermal Infrared Characterization of Ground Targets and Backgrounds*, SPIE Press, 2nd edition, 2006.

[2] P. B. W. Schwering, D. F. Bezuidenhout, W. H. Gunter et al., "Optical characterisation of small surface targets," in *Electro-Optical Remote Sensing, Detection, and Photonic Technologies and Their Applications*, vol. 6739 of *Proceedings of SPIE*, Florence, Italy, September 2007.

[3] P. B. W. Schwering, D. F. Bezuidenhout, W. H. Gunter, F. P. J. Le Roux, and R. H. Sieberhagen, "IRST infrared background analysis of bay environments," in *Infrared Technology and Applications XXXIV*, vol. 6940 of *Proceedings of SPIE*, Orlando, Fla, USA, March 2008.

[4] A. N. de Jonga, P. B. W. Schweringa, P. J. Fritza, and W. H. Gunterb, "Optical characteristics of small surface targets, measured in the False Bay, South Africa; June 2007," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XX*, vol. 7300 of *Proceedings of SPIE*, Orlando, Fla, USA, April 2009.

[5] G. C. Holst, *Electro-Optical Imaging System Performance*, International Society for Optical Engineering, 2008.

[6] H. K. Hughes, "Beer's law and the optimum transmittance in absorption measurements," *Applied Optic*, vol. 2, no. 9, pp. 937–945, 1963.

[7] S. Kim, "Double layered-background removal filter for detecting small infrared targets in heterogenous backgrounds," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 32, no. 1, pp. 79–101, 2011.

[8] J. Silverman, C. E. Caefer, S. DiSalvo, and V. E. Vickers, "Temporal filtering for point target detection in staring IR imagery: II. Recursive variance filter," in *Signal and Data Processing of Small Targets 1998*, vol. 3373 of *Proceedings of SPIE*, pp. 44–53, Orlando, Fla, USA, April 1998.

[9] L. Varsano, I. Yatskaer, and S. R. Rotman, "Temporal target tracking in hyperspectral images," *Optical Engineering*, vol. 45, no. 12, Article ID 126201, 2006.

[10] D. Liu, J. Zhang, and W. Dong, "Temporal profile based small moving target detection algorithm in Infrared image sequences," *International Journal of Infrared and Millimeter Waves*, vol. 28, no. 5, pp. 373–381, 2007.

[11] Y. S. Jung and T. L. Song, "Aerial-target detection using the recursive temporal profile and spatiotemporal gradient pattern in infrared image sequences," *Optical Engineering*, vol. 51, no. 6, Article ID 066401, 2012.

[12] S. Kim, Y. Yang, and B. Choi, "Realistic infrared sequence generation by physics-based infrared target modeling for infrared search and track," *Optical Engineering*, vol. 49, no. 11, Article ID 116401, 2010.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 625280, 6 pages
http://dx.doi.org/10.1155/2014/625280

Research Article

A Stochastic Total Least Squares Solution of Adaptive Filtering Problem

Shazia Javed[1,2] and Noor Atinah Ahmad[1]

[1] *School of Mathematical Science, Universiti Sains Malaysia, 11800 Penang, Malaysia*
[2] *Lahore College for Women University, Lahore 54000, Pakistan*

Correspondence should be addressed to Shazia Javed; shaziafateh@hotmail.com

Received 30 August 2013; Accepted 5 December 2013; Published 3 February 2014

Academic Editors: S. Bourennane and J. Marot

An efficient and computationally linear algorithm is derived for total least squares solution of adaptive filtering problem, when both input and output signals are contaminated by noise. The proposed total least mean squares (TLMS) algorithm is designed by recursively computing an optimal solution of adaptive TLS problem by minimizing instantaneous value of weighted cost function. Convergence analysis of the algorithm is given to show the global convergence of the proposed algorithm, provided that the stepsize parameter is appropriately chosen. The TLMS algorithm is computationally simpler than the other TLS algorithms and demonstrates a better performance as compared with the least mean square (LMS) and normalized least mean square (NLMS) algorithms. It provides minimum mean square deviation by exhibiting better convergence in misalignment for unknown system identification under noisy inputs.

1. Introduction

Ordinary least squares methods are extensively used in many signal processing applications to extract the system parameters from input/output data [1, 2]. These methods yield an unbiased solution of adaptive least squares problem having no interference in both inputs and outputs or having interference only in the outputs of the unknown system and clean inputs. However, if interference exists in both input and output of the unknown system or adaptive filtering problem, the ordinary least squares solution gets biased [3].

Total least squares (TLS) method [4] is an efficient technique to achieve an unbiased estimate of the system parameters when both input and output are contaminated by noise. Golub and Van Loan [5] provided an analytical procedure to get an unbiased solution of the TLS problem using singular value decomposition (SVD) of data matrices. This technique is extensively used in data processing and control applications [4, 6, 7]. However, application of TLS methods in signal processing is still limited because computation of SVD requires a high complexity of $O(N^3)$ for an $N \times N$ matrix.

TLS solutions of adaptive filtering problem gained importance after the pioneer work done by Pisarenko [8]. He presented an efficient solution of adaptive TLS problem by adaptively computing the eigenvector corresponding to smallest eigenvalue of augmented input/output signal's autocorrelation matrix. Since then, several algorithms have been proposed based on the adaptive implementations of Pisarenko. The adaptive TLS algorithms proposed in [9–11] are able to achieve an unbiased TLS solution of adaptive filtering problem with a complexity of $O(N)$. However they are sensitive to the correlation properties of input signals and have a drawback of bad performance under correlated inputs.

In this paper, an iterative algorithm is presented to find an optimal TLS solution of adaptive FIR filtering problem. A stochastic technique similar to that of least mean squares (LMS) algorithm of adaptive least squares filtering is employed to develop a total least mean squares (TLMS) algorithm for adaptive total least squares problem. Instead of basing the approach on the minimum mean squares error as the LMS algorithm does, the proposed (TLMS) algorithm is based on the total mean squares, obtained by minimizing the weighted cost function for the TLS solution of adaptive filtering problem. The proposed algorithm has maintained the $O(N)$ complexity of adaptive TLS algorithms with an

additional quality of having steady state convergence under correlated inputs. Convergence analysis is presented to show the global convergence of the proposed algorithm under all kinds of inputs provided the stepsize parameter is suitably chosen.

This paper is outlined as follows: we start with a mathematical formulation of adaptive least squares problem in Section 2 and derivation of the TLMS algorithm is given in Section 3, including its convergence analysis in Section 3.1. After that efficiency of the proposed algorithm is tested in Section 4 by applying it for an unknown system identification problem and comparing the results with conventional LMS and normalized LMS (NLMS) algorithms. Concluding remarks are given in Section 5.

2. Mathematical Formulation of Adaptive Total Least Squares Problem

Consider an unknown system to be identified by adaptive FIR filter of length N and response vector $\mathbf{w}_n$, at time n, with an assumption that both input and output are corrupted by an additive white Gaussian noise (AWGN). The noise free input vector $\mathbf{a}_n \in \mathscr{R}^N$ is formed from the input signals $u(n)$, such that

$$\mathbf{a}_n = [u(n), u(n-1), \ldots, u(n-N+1)]^T \in \mathscr{R}^N. \quad (1)$$

The desired output of the unknown system is then given by

$$\widetilde{s}(n) = s(n) + \Delta s(n), \quad (2)$$

where $s(n) = \mathbf{w}_n^T \mathbf{a}_n$ is system's output and $\Delta s(n)$ an added white Gaussian noise of zero mean and variance $\sigma_{\Delta s}^2$.

The primary assumption of an adaptive least squares (ALS) problem is that perturbations occur in the output signals only and that the input signals are exactly known. This assumption is not practical enough, because perturbations due to sampling or modeling or measurement errors affect the input signals too. A sensible choice to overcome such situations is to introduce perturbations in input signals in addition to perturbations of output signals. A schematic diagram of an adaptive filter with perturbed input is depicted in Figure 1.

If $\Delta\mathbf{a}_n = [\Delta u(n), \Delta u(n-1), \ldots, \Delta u(n-N+1)]^T \in \mathscr{R}^N$, denote the perturbations in input vector $\mathbf{a}_n$, where $\Delta u(n)$ is an additive white Gaussian noise (uncorrelated from the output noise) of zero mean and variance $\sigma_{\Delta u}^2$, then noisy input vector is

$$\widetilde{\mathbf{a}}_n = \mathbf{a}_n + \Delta\mathbf{a}_n. \quad (3)$$

It is clear from Figure 1 that for every input signal $\widetilde{u}(n) = u(n) + \Delta u(n)$, the filter produces an estimated output $y(n) = \mathbf{w}_n^T \widetilde{\mathbf{a}}_n$, which is compared with $\widetilde{s}(n)$ to produce a least squares error signal $e(n) = y(n) - \widetilde{s}(n)$. Define the autocorrelation matrix $\mathbf{R}_{\widetilde{\mathbf{a}}_n}$ of noisy input vector $\widetilde{\mathbf{a}}_n$ as $\mathbf{R}_{\widetilde{\mathbf{a}}_n} = E\{\widetilde{\mathbf{a}}_n \widetilde{\mathbf{a}}_n^T\}$ and the cross-correlation vector of output signal with $\widetilde{\mathbf{a}}_n$ as $\mathbf{p}_{\widetilde{\mathbf{a}}_n} = E\{\widetilde{s}(n)\widetilde{\mathbf{a}}_n\}$.

At this stage the least squares solution, obtained by minimizing the cost function $J = E\{e^2(n)\}$, gives a poor estimation of the solution of adaptive filtering problem because of the presence of noise in filter input. Casting adaptive filtering problem as total least squares problem can, however, restructure the poor estimation of solution under noisy input [10, 11]. The following definition is made to adopt a more general signal model for ATLS-based filtering.

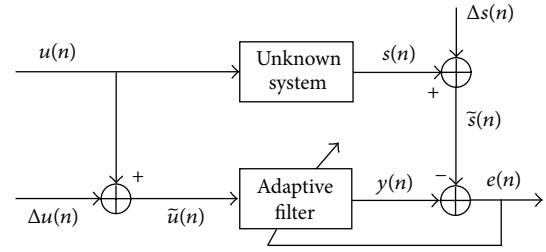


Figure 1: An unknown system identification model for adaptive filtering of noisy input signals.

Definition 1 (augmented data vector). Define an $(N+1) \times 1$ augmented data vector $\widetilde{\mathbf{z}}_n$ as

$$\widetilde{\mathbf{z}}_n = \left[\widetilde{\mathbf{a}}_n^T : \widetilde{s}(n)\right]^T = \begin{pmatrix} \widetilde{\mathbf{a}}_n \\ \widetilde{s}(n) \end{pmatrix}. \quad (4)$$

An alternate form of $e(n)$, in terms of augmented data vector of Definition 1, is obtained as follows:

$$\begin{aligned} e(n) &= y(n) - \widetilde{s}(n) = \mathbf{w}_n^T \widetilde{\mathbf{a}}_n - \widetilde{s}(n) \\ &= \left[\mathbf{w}_n^T - 1\right] \begin{pmatrix} \widetilde{\mathbf{a}}_n \\ \widetilde{s}(n) \end{pmatrix} = \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n, \end{aligned} \quad (5)$$

where $\widetilde{\mathbf{w}}_n = [\mathbf{w}_n^T - 1]^T$ denote the $(N+1) \times 1$ extended parameter vector.

The TLS solution of adaptive filtering problem is an eigenvector associated with the smallest eigenvalue of extended autocorrelation matrix $\widetilde{\mathbf{R}}_n$:

$$\widetilde{\mathbf{R}}_n = E\left\{\widetilde{\mathbf{z}}_n \widetilde{\mathbf{z}}_n^T\right\} = \begin{pmatrix} \mathbf{R}_{\widetilde{\mathbf{a}}_n} & \mathbf{p}_{\widetilde{\mathbf{a}}_n} \\ \mathbf{p}_{\widetilde{\mathbf{a}}_n}^T & \sigma_{\widetilde{s}(n)}^2 \end{pmatrix}, \quad (6)$$

where $\sigma_{\widetilde{s}(n)}^2 = E\{\widetilde{s}(n)\widetilde{s}(n)\}$.

Instead of minimizing the mean square error $E\{e^2(n)\}$, adaptive total least squares problem is concerned with minimizing the total mean square error $E\{\eta^2(n)\}$ and cost function $J(\widetilde{\mathbf{w}}_n) = E\{e^2(n)\}$, where the total error $\eta(n)$ is given by

$$\eta(n) = \frac{e(n)}{\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}} = \frac{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n}{\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}} = \frac{\widetilde{\mathbf{z}}_n^T \widetilde{\mathbf{w}}_n}{\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}}. \quad (7)$$

The TLS cost function $\widetilde{J}(\widetilde{\mathbf{w}}_n)$ is then defined in terms of total error as

$$\widetilde{J} = E\left\{\eta^2(n)\right\} = E\left\{\frac{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n \widetilde{\mathbf{z}}_n^T \widetilde{\mathbf{w}}_n}{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}\right\} = \frac{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{R}}_n \widetilde{\mathbf{w}}_n}{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}. \quad (8)$$

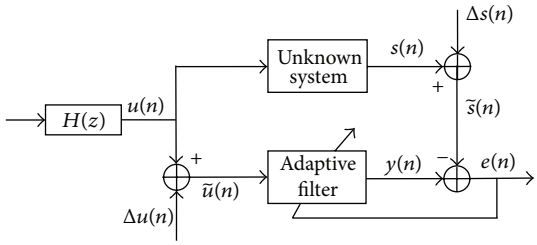


Figure 2: An unknown system identification model for adaptive filtering of correlated noisy signals.

The adaptive total least squares problem is a minimization problem of the form [10, 11]:

$$\min_{\widetilde{\mathbf{w}}_n \in \mathscr{R}^{N+1}} \widetilde{J} = \min_{\widetilde{\mathbf{w}}_n \in \mathscr{R}^{N+1}} \frac{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{R}}_n \widetilde{\mathbf{w}}_n}{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}. \tag{9}$$

Note that an optimal solution $\mathbf{w}_{\text{opt}}$ of the TLS problem (9) is an eigenvector corresponding to the smallest eigenvalue of $\widetilde{\mathbf{R}}_n$. In practice SVD technique is used to solve TLS problems since it offers lower sensitivity to the computational errors; however, it is computationally expensive [5]. An alternate choice to estimate eigenvector corresponding to smallest eigenvalue is to use an adaptive algorithm [1, 2].

3. Derivation of Total LMS Algorithm for Adaptive Filtering Problem

In adaptive least squares problem, conventional LMS algorithm is a steepest descent method which uses an instantaneous cost function $J = e^2(n)$ for computation of gradient vector [1]. Using a similar implementation in TLS problem, the total LMS (TLMS) algorithm is obtained by having an instantaneous estimate of the cost function (8) as $\widetilde{J} = \eta^2(n)$. The recursive update equation of TLMS algorithm is then given as

$$\widetilde{\mathbf{w}}_{n+1} = \widetilde{\mathbf{w}}_n - \mu \nabla_{\widetilde{\mathbf{w}}_n} \widetilde{J}, \tag{10}$$

where μ is the stepsize parameter or convergence parameter. Note that

$$\begin{aligned}
\nabla_{\widetilde{\mathbf{w}}_n} \widetilde{J} &= \frac{\partial}{\partial \widetilde{\mathbf{w}}_n} \left(\eta(n)^2\right) \\
&= 2\eta(n) \frac{\partial}{\partial \widetilde{\mathbf{w}}_n} \eta(n) \\
&= 2\eta(n) \frac{\partial}{\partial \widetilde{\mathbf{w}}_n} \left(\frac{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n}{\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}} \right) \\
&= \frac{2\eta(n)}{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n} \left\{ \widetilde{\mathbf{z}}_n \sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n} - \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n \frac{\widetilde{\mathbf{w}}_n}{\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}} \right\} \\
&= \frac{2\eta(n)}{\left(\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n\right)^{3/2}} \left\{ \widetilde{\mathbf{z}}_n \cdot \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n - \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n \cdot \widetilde{\mathbf{w}}_n \right\}.
\end{aligned} \tag{11}$$

Table 1: LMS-total (TLMS) algorithm for adaptive filtering.

Algorithm	×/÷	+/−
Initialization		
$\mathbf{w}_o = \mathbf{0}$	...	...
$\widetilde{\mathbf{w}}_o = [\mathbf{w}_o^T \quad -1]^T$	...	...
Update		
for $n = 0, 1, 2, \ldots$		
$\widetilde{\mathbf{z}}_n = \left[\widetilde{\mathbf{a}}_n^T : \widetilde{s}(n)\right]^T$	...	...
$e(n) = \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n$	$N+1$	N
$\text{norm}_{\text{sq}} = \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n$	$N+1$	N
$\widetilde{\mathbf{w}}_{n+1} = \widetilde{\mathbf{w}}_n + \frac{2\mu\, e(n)}{\text{norm}_{\text{sq}}^2} \left\{ e(n) \cdot \widetilde{\mathbf{w}}_n - \text{norm}_{\text{sq}} \cdot \widetilde{\mathbf{z}}_n \right\}$	$3N+7$	$2N+2$
$\mathbf{w}_{n+1} = -\frac{\widetilde{\mathbf{w}}_{n+1}(1:N)}{\widetilde{\mathbf{w}}_{n+1}(N+1)}$	N	...
Total	$6N+9$	$4N+2$

Using $e(n) = \widetilde{\mathbf{z}}_n^T \widetilde{\mathbf{w}}_n = \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n$ and $\|\widetilde{\mathbf{w}}_n\| = \sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}$, then above equation becomes

$$\nabla_{\widetilde{\mathbf{w}}_n} \widetilde{J} = \frac{2}{\left\|\widetilde{\mathbf{w}}_n\right\|^4} \left\{ \left\|\widetilde{\mathbf{w}}_n\right\|^2 e(n) \cdot \widetilde{\mathbf{z}}_n - e(n)^2 \cdot \widetilde{\mathbf{w}}_n \right\}. \tag{12}$$

Substituting (12) in (10), the updated equation of TLMS algorithm becomes

$$\widetilde{\mathbf{w}}_{n+1} = \widetilde{\mathbf{w}}_n + \frac{2e(n) \cdot \mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^4} \left\{ e(n) \cdot \widetilde{\mathbf{w}}_n - \left\|\widetilde{\mathbf{w}}_n\right\|^2 \widetilde{\mathbf{z}}_n \right\}. \tag{13}$$

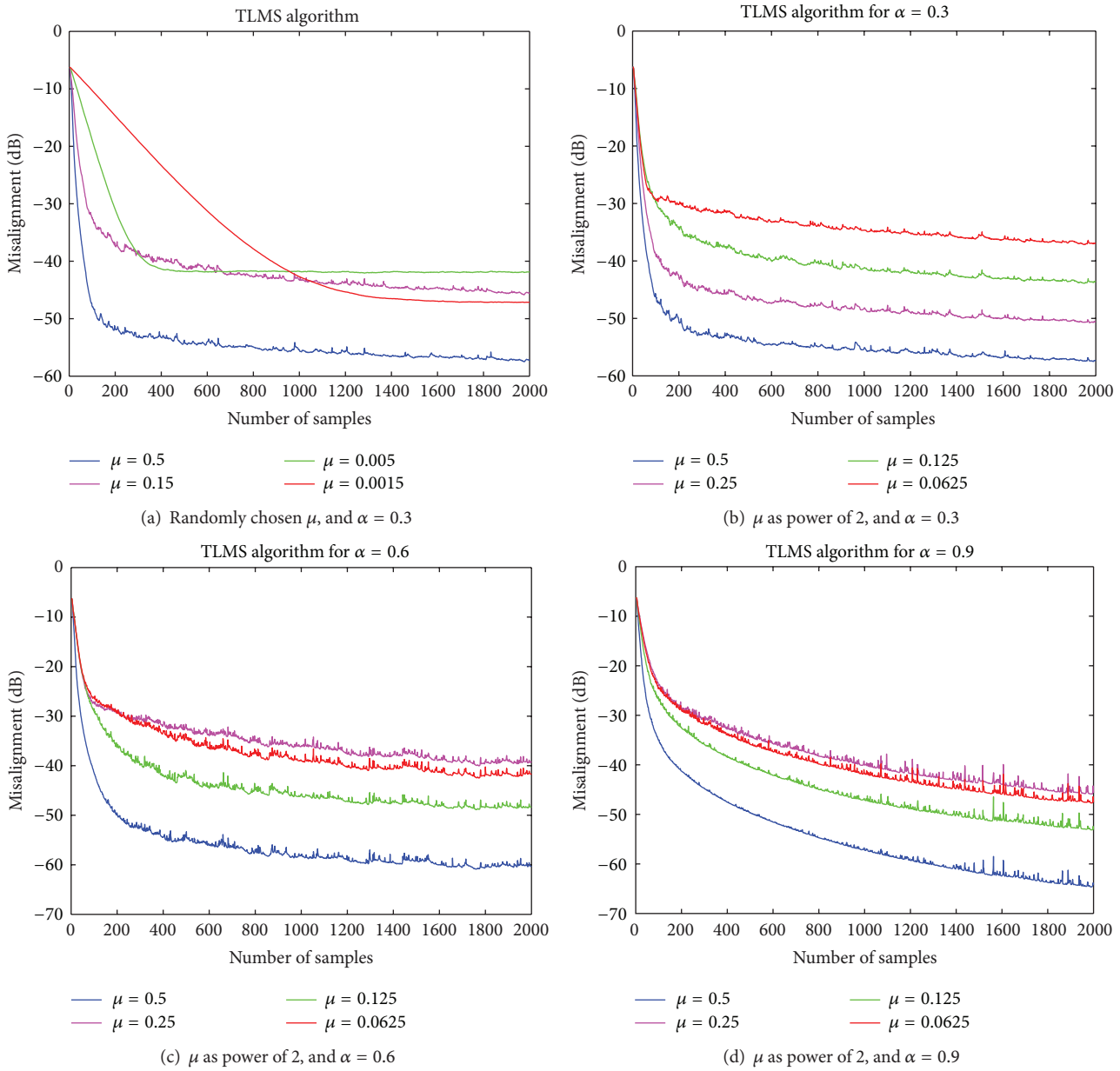
Once $\widetilde{\mathbf{w}}_{n+1}$ is computed using (13), the TLS solution update $\mathbf{w}_{n+1}$ is obtained by the following formula:

$$\mathbf{w}_{n+1} = -\frac{\widetilde{\mathbf{w}}_{n+1}(1:N)}{\widetilde{\mathbf{w}}_{n+1}(N+1)}. \tag{14}$$

The detailed TLMS algorithm is summarized in Table 1. A complexity measure of the algorithms shows that it is a computationally linear algorithm, requiring a total of $6N + 9$ multiplications/divisions per iteration. This computational simplicity of adaptive TLMS algorithm makes it a better choice than computationally expensive SVD based TLS algorithm, which requires $6N^3$ computations per iteration [5].

3.1. Convergence Analysis. In (13), inner product with $\widetilde{\mathbf{z}}_n$ yields,

$$\begin{aligned}
\widetilde{\mathbf{w}}_{n+1}^T \widetilde{\mathbf{z}}_n &= \left(\widetilde{\mathbf{w}}_n + \frac{2e(n) \cdot \mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^4} \left\{ e(n) \cdot \widetilde{\mathbf{w}}_n - \left\|\widetilde{\mathbf{w}}_n\right\|^2 \widetilde{\mathbf{z}}_n \right\} \right)^T \widetilde{\mathbf{z}}_n \\
&= \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n + \frac{2e(n) \cdot \mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^4} \left\{ e(n) \cdot \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n - \left\|\widetilde{\mathbf{w}}_n\right\|^2 \widetilde{\mathbf{z}}_n^T \widetilde{\mathbf{z}}_n \right\} \\
&= e(n) + \frac{2e(n) \cdot \mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^4} \left\{ |e(n)|^2 - \left\|\widetilde{\mathbf{w}}_n\right\|^2 \left\|\widetilde{\mathbf{z}}_n\right\|^2 \right\} \\
&= e(n) + \frac{2e(n) \cdot \mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^2} \left\{ |\eta(n)|^2 - \left\|\widetilde{\mathbf{z}}_n\right\|^2 \right\}.
\end{aligned} \tag{15}$$



(a) Randomly chosen μ, and $\alpha = 0.3$

(b) μ as power of 2, and $\alpha = 0.3$

(c) μ as power of 2, and $\alpha = 0.6$

(d) μ as power of 2, and $\alpha = 0.9$

FIGURE 3: Learning curves of total misalignment of TLMS algorithm.

Since $e(n) = \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n$, Cauchy-Schwarz inequality [12] gives

$$|e(n)|^2 \le \|\widetilde{\mathbf{w}}_n\|^2 \|\widetilde{\mathbf{z}}_n\|^2; \tag{16}$$

that is,

$$\begin{aligned} \left|\frac{e(n)}{\|\widetilde{\mathbf{w}}_n\|}\right|^2 &\le \|\widetilde{\mathbf{z}}_n\|^2, \\ |\eta(n)|^2 &\le \|\widetilde{\mathbf{z}}_n\|^2, \\ |\eta(n)|^2 - \|\widetilde{\mathbf{z}}_n\|^2 &\le 0. \end{aligned} \tag{17}$$

Let $\delta_n = \|\widetilde{\mathbf{z}}_n\|^2 - |\eta(n)|^2 \ge 0$ then (15) becomes:

$$\begin{aligned} \widetilde{\mathbf{w}}_{n+1}^T \widetilde{\mathbf{z}}_n &= e(n) + \frac{2e(n)\cdot\mu}{\|\widetilde{\mathbf{w}}_n\|^2}\{-\delta_n\} \\ &= \left\{1 - \frac{2\mu\delta_n}{\|\widetilde{\mathbf{w}}_n\|^2}\right\} e(n) \end{aligned} \tag{18}$$

or,

$$\widetilde{\mathbf{w}}_{n+1}^T \widetilde{\mathbf{z}}_n = \left\{1 - \frac{2\mu\delta_n}{\|\widetilde{\mathbf{w}}_n\|^2}\right\} \widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{z}}_n. \tag{19}$$

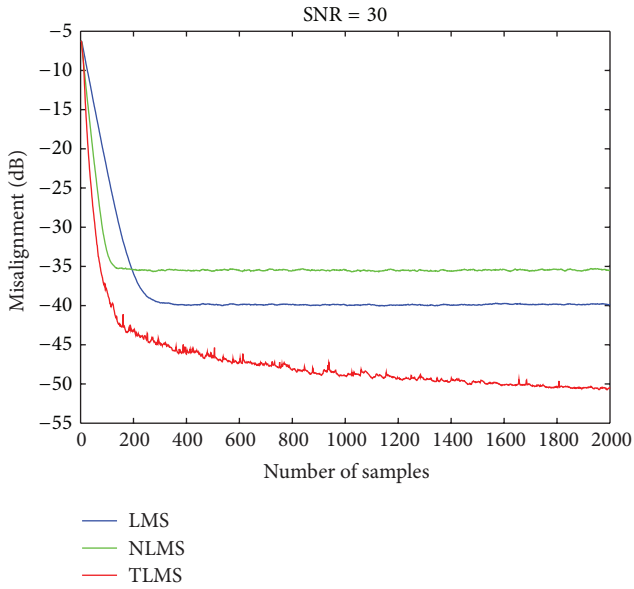


Figure 4: Comparison of TLMS, LMS, and NLMS algorithms.

Since $\widetilde{\mathbf{z}}_n \neq \mathbf{0}$,

$$\widetilde{\mathbf{w}}_{n+1} = \left\{1 - \frac{2\mu\delta_n}{\left\|\widetilde{\mathbf{w}}_n\right\|^2}\right\} \widetilde{\mathbf{w}}_n \tag{20}$$

which shows that $\{\widetilde{\mathbf{w}}_n\}$ is a geometric progression. It would converge to an optimal solution if

$$\left|1 - \frac{2\mu\delta_n}{\left\|\widetilde{\mathbf{w}}_n\right\|^2}\right| < 1 \tag{21}$$

or

$$0 < \frac{\mu}{\left\|\widetilde{\mathbf{w}}_n\right\|^2} < \frac{1}{\delta_n}. \tag{22}$$

This shows that the proposed algorithm is a variable stepsize algorithm, with $\widetilde{\mu} = \mu/\|\widetilde{\mathbf{w}}_n\|^2$. An appropriate way to choose μ is to initialize the algorithm such that $\|\mathbf{w}_o - \mathbf{w}_{\text{opt}}\|$ is less than $2\|\mathbf{w}_{\text{opt}}\|$ [13]. According to this result for $\mathbf{w}_o = \mathbf{0}$, $\|\widetilde{\mathbf{w}}_o\| = 1$ and $\widetilde{\mu} = \mu$, while $\delta_o = \|\widetilde{\mathbf{z}}_n\|^2 - \widetilde{s}(n) > 0$.

4. Application of TLMS Algorithm in System Identification

To examine the performance of proposed TLMS algorithm, an unknown system identification model, shown in Figure 2, is used.

A white Gaussian input signal of variance $\sigma^2 = 1$ is passed through a coloring filter with frequency response [1]:

$$H(z) = \frac{\sqrt{1 - \alpha^2}}{1 - \alpha z^{-1}}, \tag{23}$$

where $|\alpha| < 1$, α is a correlation parameter and controls the eigenvalue spread of input signals. $\alpha = 0$ corresponds to the case when eigenvalue spread of input signals is close to 1, and eigenvalue spread increases with an increase in the value of α.

A white Gaussian noise of SNR = 30 dB is added in the input signal $u(n)$ to get noisy signal $\widetilde{u}(n)$, and an output signal $\widetilde{s}(n)$ is obtained by corrupting the output signal $s(n)$ with an additive white Gaussian noise of SNR 30 dB. Proposed TLMS algorithm is compared with LMS and NLMS algorithms of [1] to get an FIR vector for a filter of length $N = 10$. Least squares misalignment $\|\mathbf{w}_n - \mathbf{w}_{\text{opt}}\|$ is compared with the total least squares misalignment $\|\mathbf{w}_n - \mathbf{w}_{\text{opt}}\|/\sqrt{\widetilde{\mathbf{w}}_n^T \widetilde{\mathbf{w}}_n}$, and simulations results are recorded for 2000 iterations with an ensemble average of 1000 independent runs.

4.1. Convergence Behavior Corresponding to Stepsize Parameter μ. Although TLMS algorithm converges for all values of μ, satisfying (22), but steady state convergence TLMS algorithm is observed when stepsize parameter μ is a power of 2. In Figures 3(b)–3(d), four learning curves of total misalignment of TLMS algorithm are shown, corresponding to $\mu = 0.5, 0.25, 0.125$, and 0.0625, and it is observed that robustness increases uniformly with an increase in the value of μ. On the other hand if μ is chosen randomly, then a change in the convergence behavior is random, though Figure 3(a) shows that algorithm still converges.

4.2. Convergence Behavior Corresponding to Correlation Parameter α. To check effect of changes in correlation parameter α on the steady state convergence behavior of TLMS algorithm, different simulations are presented in Figure 3, each showing four learning curves of total misalignment of TLMS algorithm corresponding to $\mu = 0.5, 0.25, 0.125$, and 0.0625. In the first two simulations $\alpha = 0.3$ in Figures 3(a) and 3(b), it is 0.6 in Figure 3(c), and 0.9 in Figure 3(d). It is clear from the results of all these simulation curves that increase in correlation of data signals has not affected the steady state performance of the algorithm. Although the convergence speed seems to slow down, but all the curves converge to optimal solution.

4.3. Comparison. Figure 4 shows the comparison of misalignment of three algorithms, that is, LMS, NLMS, and TLMS algorithms. The first two compute a least squares solution of adaptive total least squares problem, while the third one computes TLS solution of adaptive total least squares problem. Taking $\alpha = 0.3$, stepsize parameter for LMS algorithm is chosen as 0.015, for NLMS algorithm as 0.3, and for TLMS algorithm, it is 0.25. The results in this simulation show that the convergence of TLMS algorithm increases with an increase in the iteration, and it presents a better solution of adaptive TLS problem.

5. Conclusion

In this paper, an efficient TLMS algorithm is presented for the total least squares solution of adaptive filtering problem. The proposed algorithm is derived by using cost function of

weighted instantaneous error signals and an efficient computation of misalignment in terms of mean squares deviation. TLMS algorithm has better ability to tackle with perturbations of both input and output signals, because it is chiefly derived for the purpose. Since in real life problems, both input and output signals are contaminated by noise, therefore TLMS algorithm has great applicability. Convergence analysis shows that the proposed algorithm has global convergence, provided that the stepsize parameter is chosen appropriately. Furthermore, it is computationally simple and requires only $O(N)$ complexity, while other algorithms for TLS problems either require higher complexity or are sensitive to correlation properties of data signals.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to the anonymous referees for their valuable comments, appreciation, and recommendation to publish this paper. This work is supported by School of Mathematical Sciences, Universiti Sains Malaysia.

References

[1] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, John Wiley & Sons, New York, NY, USA, 1998.

[2] N. A. Ahmad, "Comparative study of iterative search method for adaptive filtering problems," in *Proceedings of the International Conference On Applied Mathematics (ICAM '05)*, Bedong, Indonesia, August 2005.

[3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 2nd edition, 1991.

[4] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, Pa, USA, 1991.

[5] G. H. Golub and C. F. Van Loan, *An Analysis of Total Least Squares Problem*, SIAM, Philadelphia, Pa, USA, 1980.

[6] M. A. Rahman and K.-B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1440–1454, 1987.

[7] Y. Hua and T. K. Sarkar, "On the total least squares linear prediction method for frequency estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2186–2189, 1990.

[8] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical Journal of the Royal Astronomical Society*, vol. 33, no. 3, pp. 347–366, 1973.

[9] P. A. Thompson, "Adaptive spectral analysis technique for unbiased frequency estimation in the presence white noise," *Proceedings of the 13th Asilomar Conference on Circuits, Systems & Computers*, pp. 529–533, 1980.

[10] D.-Z. Feng, "Efficient recursive total least squares algorithm for FIR adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, pp. 268–280, 1994.

[11] D.-Z. Feng, X.-D. Zhang, D.-X. Chang, and W. X. Zheng, "A fast recursive total least squares algorithm for adaptive FIR filtering," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2729–2737, 2004.

[12] E. Kreyszig, *Introductory Functional Analysis with Applications*, vol. 1, John Wiley & Sons, New York, NY, USA, 1989.

[13] B. E. Dunne and G. A. Williamson, "Stable simplified gradient algorithms for total least squares filtering," in *Proceedings of the 34th Asilomar Conference on Signals, Systems, Computers*, pp. 1762–1766, November 2000.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 983871, 8 pages
http://dx.doi.org/10.1155/2014/983871

Research Article

Local Analysis of Human Cortex in MRI Brain Volume

Sami Bourouis[1,2]

[1] *Université de Tunis El Manar, École Nationale d'Ingénieurs de Tunis, BP 37 Belvedere Tunis, 1002 Tunis, Tunisia*
[2] *College of Computers and Information Technology, Taif University, P.O. Box 888, Taif 21974, Saudi Arabia*

Correspondence should be addressed to Sami Bourouis; sami_bourouis@yahoo.fr

Received 3 August 2013; Accepted 29 October 2013; Published 29 January 2014

Academic Editors: S. Bourennane and J. Marot

This paper describes a method for subcortical identification and labeling of 3D medical MRI images. Indeed, the ability to identify similarities between the most characteristic subcortical structures such as sulci and gyri is helpful for human brain mapping studies in general and medical diagnosis in particular. However, these structures vary greatly from one individual to another because they have different geometric properties. For this purpose, we have developed an efficient tool that allows a user to start with brain imaging, to segment the border gray/white matter, to simplify the obtained cortex surface, and to describe this shape locally in order to identify homogeneous features. In this paper, a segmentation procedure using geometric curvature properties that provide an efficient discrimination for local shape is implemented on the brain cortical surface. Experimental results demonstrate the effectiveness and the validity of our approach.

1. Introduction

1.1. Problem Statement. In a medical context, cortical surfaces analysis has motivated many researchers in anatomy, visualization, image registration, and pattern recognition [1–5]. The interest in this analysis comes in part from the variability of the convoluted shape of the cerebral cortex which is considered as the key of human intelligence and one of the most important factors regarding the comparison of brain anatomy and function [3]. However, the cortical surface has a complex structure comprised of folds "gyri" and fissures "sulci" (Figure 1). The accurate identification and labeling of these structures is helpful in human brain mapping studies, yet very challenging. Sulcal and gyral features are often considered as anatomical landmarks for automatically parcellating the cortex into features of interest that are functionally distinct. They also vary greatly from one individual to another because they have different geometric properties. So, we can consider them as a useful benchmark for comparison in the case of morphometric study [5]. For all these reasons, an accurate analysis of the cortical surface is very helpful for specialists in order to clarify the evolution differences between humans and animals, to study nervous disorders, and to preserve the main functions of a patient during brain surgery.

The labeling of cortical surface would be a significant aid in the study of the cortex anatomy. Indeed, the neurosurgeon would need to locate the regions of interest versus cortical folds (sulci and gyri) of the patient. However, this operation is not easy due to the complexity of anatomical shapes and the interpatient variability. Therefore, the extraction and labeling of cortical surface automatically and objectively become of great importance. It should be noted that this segmentation addresses various problems such as 3D visualization, topological analysis of the surface, brain mapping, and the analysis and the interpretation of brain activity. In addition, we think that automatic extraction can improve the computational time, the quality, and the reproducibility of this process.

1.2. Related Work. In this context, several studies have been developed to quantify cortical morphometric dissimilarities and then to facilitate the labeling of the human cortex. Unfortunately, some of them require manual intervention which penalizes the reproducibility of this task. Proposed approaches in the literature are divided into two main categories. The first one is based on spatial normalization. In this case, a coordinate system (i.e., 3D Talairach grid) is used to make a comparison between the patient brain and the reference (template) one. Template brain serves to give

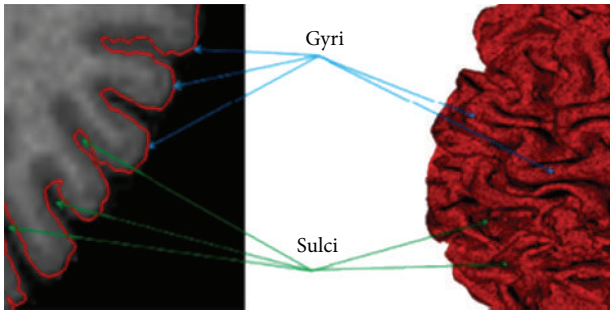


Figure 1: Illustration of sulcal and gyral patterns.

precise information on the point location. For example, Jaume et al. [6] proposed an algorithm to match an atlas labeled mesh to the patient brain mesh in a multiresolution way. Then, they transfer the labels from the matched mesh to label the patient mesh without manual intervention. Labeling the patient brain surface provides a map of the brain folds where the neurosurgeon can easily track the features of interest. The second category is based on the extraction of relevant anatomical regions (i.e., sulci, subcortical nuclei, gyri) in order to make comparison with shape descriptors or any other kind of descriptor. The manual delineation of these features requires a tedious process because of the large inter- and intraindividual variability [7]. So, the development of automatic segmentation tools based on specific and reproducible features is highly recommended [8–10]. The main features used for segmenting the cortical surface are based on geometric properties such as curvature [11–14], geodesic depth [15–17], medial axis [18], and so forth. For example, curvature-based approaches make the first-order approximation that sulci are concave and gyri are convex, and geodesic depth-based approaches use ad hoc techniques to distinguish these two complex structures. Kao et al. [17] apply depth thresholding to extract sulcal regions. They compute a geometric depth measure for each point on the cortical surface, and they extract sulcal regions by checking the connectivity above a depth threshold. Finally, they delineate the fundus by thinning each connected region keeping the endpoints fixed.

Other segmentation methods use other techniques such as watersheds of curvature function [19, 20]. Nevertheless, the watershed method shows often sensitivity to the choice of the depth threshold parameter and to noise, so, the division of the surface into regions cannot give accurate information about the differences between local regions. Moreover, deformable models [21] are applied to discriminate sulci and gyri. However, parameters describing the elasticity of the model affect the definition of some areas such as sulcal ones. The proposed approach in [14] is intended to describe a surface by dividing it into homogeneous regions according to the discrete mean and Gaussian curvature estimates. The surfaces are obtained from three-dimensional imaging datasets by isosurface extraction after data presmoothing. A hierarchical multiresolution representation of the isosurface is then generated. Finally, segmentation is performed at various levels of detail to detect the main features of the surface. This low-resolution description is used to determine constraints for the segmentation at the higher resolutions. Methods of extracting the cortical surfaces from MRI brain volumes have facilitated studies on inter- and intraindividual variability of sulcal and gyral patterns. In a clinical routine, these patterns are often delineated by a manual labeling process. This process is extremely tedious, time consuming, and probably leads to measurement errors.

1.3. Motivations and Contributions. According to the literature reviewed, we noted that the local curvature provides an effective shape measure and could be applied to discrete surface analysis. Moreover, measures related to the curvatures are used in several works to segment subcortical features. For example, the average principal curvatures (mean curvature) can describe the local folding of the surface. The gyri correspond to large values of the mean curvature and are detected via thresholding. Moreover, the Gaussian curvature is an intrinsic measure typically used to classify the surfaces in different primitive forms. Now, if we proceed by segmenting the cortex at high resolution, it would be difficult to interpret and compare the results because fine structures like gyri and sulci are very complex. To resolve this problem, we have to keep only main folds while eliminating the details. So, it is important to create a low-resolution representation of the initial mesh. This need is due also to rendering speed reasons and to allow fast transmission of 3D models in network-based applications.

This work takes place in this growing area and it proposes an efficient method and tool to segment and describe locally the subcortical structures. Unlike our previous work [22], the proposed work herein is more relevant and complete. Indeed, we treat in this paper the problem of local segmentation of human cerebral cortex. Our main purpose is to analyze finely obtained discrete surfaces after a segmentation operation of MRI volume. So, we propose here a fully automatic method for parcellating the cerebral cortex into gyri and sulci features of interest. It is based mainly on discrete differential geometry operators, which are the key to distinguish the two patterns, and on multiresolution representation to simplify and accelerate processing time.

The rest of this paper is organized as follows. The suggested method is presented in the next section. The experimental results are given in Section 3. Finally, Section 4 concludes the paper.

2. Methods and Materials

2.1. Method Outline. The proposed method is based mainly on two essential steps: the first one is intended to simplify the initial mesh and the second one to classify the surface as containing homogeneous attributes by using invariant local descriptors based on local geometric curvature. Because the cortex is composed of a large set of folds, it is important to detect only the main folds and not all small ones on the cortical surface. Progressive mesh simplification solves this problem. Indeed, the segmentation in high resolution could

(1) 3D MRI volume segmentation.
(2) Gray matter/white matter interface extraction.
(3) Topology preserving mesh decimation.
(4) Local cerebral cortex analysis based on local curvature estimation.

Algorithm 1

fail to properly distinguish between the major folds and it could lead to misinterpretation. Mesh simplification, whose role is to keep only the main folds while removing small details must be made with caution. Indeed, if we are interested in both sulcal and gyral patterns, it is better to simplify the entire cortical surface in a uniform manner to preserve the topology of these structures. If we are interested in a particular structure such as sulci, it is interesting in this case to simplify the surface more while preserving the topology of the sulci. So, simplification may be done without topology preservation in some subarea and with topology preservation in other ones. We introduce here the main steps performed by our method to have a rapid and efficient characterization of cortical surfaces (see Figure 3 and Algorithm 1).

2.2. Discrete Cortical Surface Detection. To study the human cortex folding patterns, we may either use the interface between gray matter (GM) and white matter (WM) as the cortical surface representation is reliable [13]. In recent years, there has been a considerable effort in developing methods for this purpose. A range of methods including classification-based, region-based, contour-based, and knowledge-based approaches have been proposed for MR brain image segmentation. Manual editing is highly accurate and has been one of the most employed techniques, but tedious and laborious, because boundaries are usually traced by hand. Automating this process is a challenging problem.

2.2.1. Gray Matter/White Matter Interface Extraction. Recently, we have proposed an efficient fully automatic method for brain MRI volume segmentation [23]. The approach performs the segmentation using a succession of operations involving a registration step from known data, a classification step, and a segmentation step based on the level-set technique. The role of the registration and the classification is to accurately initialize the active model and to control its evolution. Recently, we have proposed a new formulation [24] for the evolution of the variational model which is expressed as

$$\frac{\partial \psi}{\partial t} = \left[\alpha_r F_{\text{region}}(I) + \alpha_b F_{\text{boundary}}(I)\right] |\nabla \psi|. \quad (1)$$

F_{boundary} causes the evolving of the front to be more strongly attracted to image edges and is given by

$$F_{\text{boundary}}(I) = \text{sign}\left(F_{\text{boundary}}\right) \cdot \frac{c+k}{1+|\nabla I|}, \quad (2)$$

where

$$\text{sign}\left(F_{\text{boundary}}\right) = \begin{cases} +1, & \text{if } F_{\text{region}} < 0, \\ -1, & \text{otherwise.} \end{cases} \quad (3)$$

F_{region} controls the evolution of the model and segments the cancer region based on the following equation:

$$F_{\text{region}}(I) = \begin{cases} I - (m_T - \epsilon_T), & \text{if } I < m_T, \\ (m_T + \epsilon_T) - I, & \text{otherwise,} \end{cases} \quad (4)$$

where ϵ_T is a constant parameter and m_T is the mean value of the bone cancer region. This value is calculated on the estimated region after the classification step. ϵ_T controls the brightness of the region to be segmented and defines a range of greyscale values that could be considered inside the expected region of interest. More technical details could be found in [23, 24].

2.2.2. Decimation and Topology Preserving. Representing the surface as explicit geometry is efficient when used with the conventional computer graphics approaches for shading and viewing. Further, it greatly reduces the necessary data storage and provides a data structure that can be measured. In our case, a triangulated surface mesh was generated from the segmented white/gray matter using a standard isosurface "Marching Cubes" algorithm [25]. It has the advantage of providing an accurate three-dimensional polygonal representation that can be used for other image processing tasks. Nevertheless, the marching cubes output usually contains multiple small "useless" meshes which are physically disconnected from each other. Moreover, it produces more than the necessary number of polygons needed to represent an object accurately. The result contains an enormous number of extremely small triangles that prevent an interactive rendering of models. A high resolution mesh generated from volume data is generally very hard to work with. In order to reduce aliasing artifacts on images, we are led to apply the algorithm proposed in [26]. On the other hand, we suggest simplifying the created mesh in order to accelerate the overall mesh analyzing process. In fact, a low resolution mesh is an efficient way to convey fine surface details while maintaining a simple underlying geometry. To achieve mesh simplification, we use VTK's Quadric Decimation [27, 28], an algorithm to reduce the large number of triangles in the mesh. In Hoppe's scheme, the topology preserving operations (EdgeCollapse and EdgeSplit) are sufficient to transform the full resolution mesh into a simpler base mesh. This optimization procedure is able to provide a good approximation to the original mesh, preserves its geometry, and conserves its overall appearance (material

identifiers, color values, normals, and texture coordinates). The algorithm is based on repeated edge collapses until the requested mesh reduction is achieved. Edges are placed in a priority queue based on a quadric error metric to delete the edge. So, an optimal collapse point can be computed. The process is repeated until the desired reduction level is reached or until topological constraints prevent further reduction.

2.3. Local Cerebral Analysis. Different approaches can be used to study fine details of the cortical surface folding patterns. For example, depth maxima have been used to detect a concept similar to sulcal roots in [15]. On the other hand, a surface's behavior can be described by dividing the surface into distinct regions of elliptic and hyperbolic behaviors. The regions of elliptic behavior can be classified into convex and concave regions by considering the direction of the surface normal. Curvature is one of the most useful criteria for intrinsic structure description of a given surface. For example, Gaussian and mean curvatures may be used to classify the surface into meaningful structures such as valleys or ridges. Through these structures, it is possible to discriminate the surface into connected elliptical or hyperbolic regions. In this study, we suggest segmenting the discrete surface into homogeneous small regions according to different criteria which are based mainly on the local principal curvatures. We compute Gaussian and mean curvatures, and later we use them to classify the vertices into different categories. To estimate the curvature information of each vertex we have applied the approximations proposed by Meyer et al. [29]. We recall here the expression of the mean K_H and the Gaussian K_G curvature operators:

$$K_H(x_i) = \frac{1}{2}\left\| \frac{1}{2A_M} \sum_{i \in N(i)} \left(\cot\alpha_{ij} + \cot\beta_{ij}\right)\left(x_i - x_j\right) \right\|$$
$$K_G(x_i) = \frac{2\Pi - \sum_{j=1}^{\sharp f} \Theta_j}{A_M}, \tag{5}$$

where α_{ij} and β_{ij} are the two angles opposite to the edge in the two triangles sharing the edge (x_i, x_j) as depicted in Figure 2. Θ_j is the angle of the jth face at the vertex x_i and $\sharp f$ denotes the number of faces around this vertex. The maximum principal curvature $k_{\max}$ and the minimum principal curvatures $k_{\min}$, which are related to the Gaussian and the mean curvatures, are also calculated at the vertex x_i as:

$$k_{\max}(x_i) = K_H(x_i) + \sqrt{\Delta(x_i)},$$
$$k_{\min}(x_i) = K_H(x_i) - \sqrt{\Delta(x_i)}, \tag{6}$$
$$\Delta(x_i) = \max\left(K_H^2(x_i) - K_G^2(x_i), 0\right).$$

By this mean, each vertex should belong to a gyral or sulcal compartment based on:

(i) Its mean and Gaussian curvature values [14],

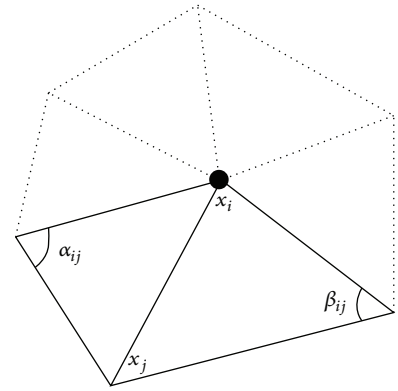


Figure 2: One-ring neighbors and angles opposite to an edge.

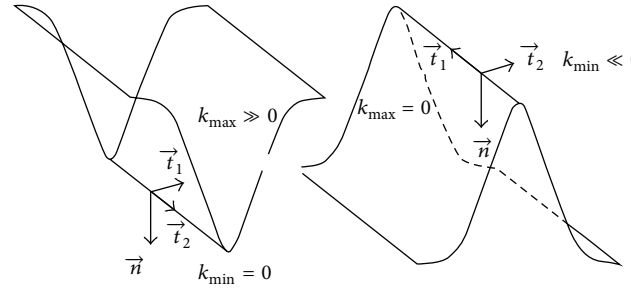


Figure 3: Sign of the principal curvatures relative to the surface behavior.

(ii) A function of the principal curvatures, $k_{\max}$ and $k_{\min}$, so-called Shape Index (SI) [30] which is independent of translation, rotation, and scaling. This measure is given by

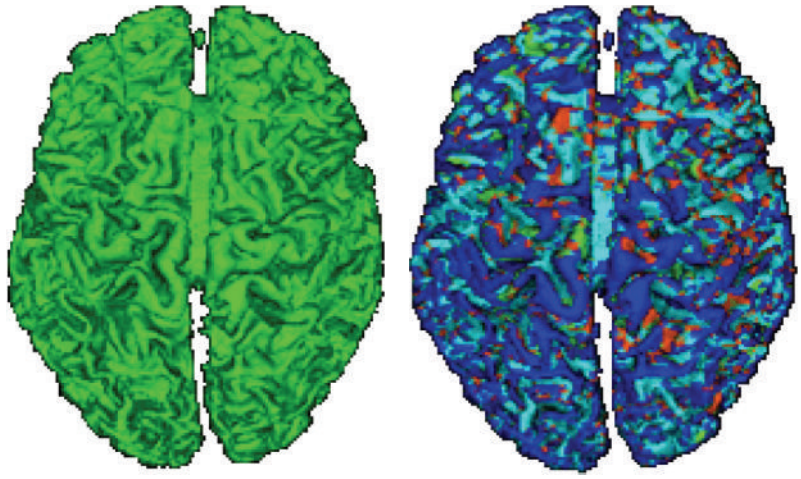
$$\mathrm{SI} = -\frac{2}{\pi}\arctan\left(\frac{k_{\max} + k_{\min}}{k_{\max} - k_{\min}}\right). \tag{7}$$

The value of the Shape Index varies in $[-1, 1]$ where negative values represent concave surface, whereas positive values correspond to convex one. The segmentation process consists of classifying vertices according to Table 1 by encoding each surface type with a particular color. Indeed, elliptic regions on the surface must be separated from hyperbolic regions. Taking into account these criteria guarantees an efficient way to classify all vertices of the discrete surface into gyral or sulcal features.

3. Experimental Results

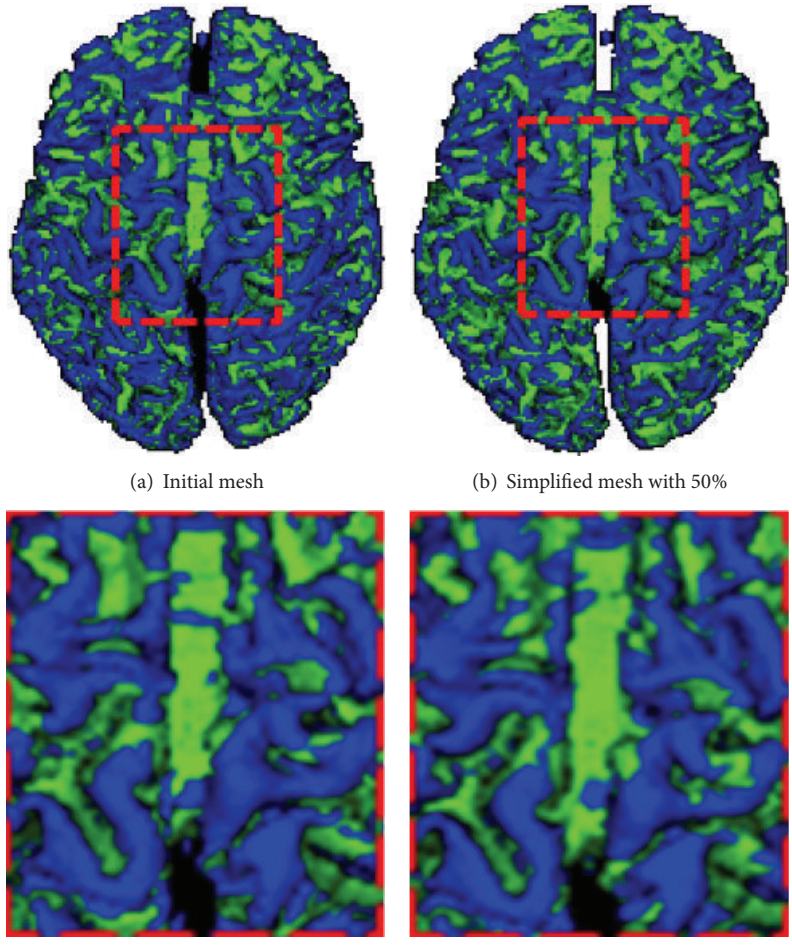
We have performed a series of experiments on brain MR images. Resulting labels are depicted in Figure 4. This figure shows the partitioning of the cortical mesh into elliptical convex regions (cyan), elliptical concave regions (red), hyperbolic convex regions (green), and hyperbolic concave regions (blue). These segmentations are presented also in a multiresolution setting (Figure 6).

To validate the proposed method, we limited ourselves to a qualitative assessment of results. Unfortunately, we were



(a) (b)

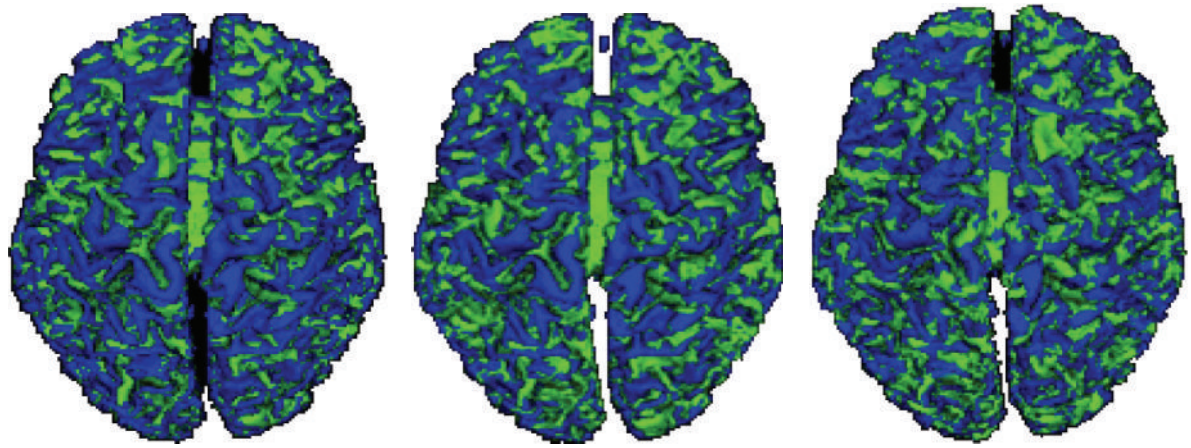
Figure 4: An example of local cortical surface segmentation. The blue color is associated with sulcal regions, red with minimum values of sulcal regions, green with gyral regions, and cyan with maximum values of gyral regions.



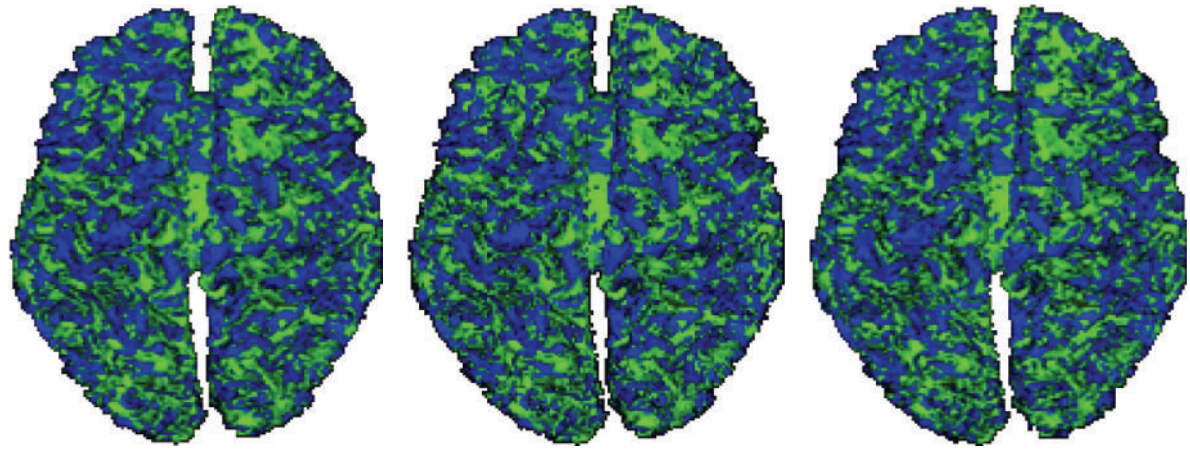
(a) Initial mesh (b) Simplified mesh with 50%

(c) Zoom on initial mesh (d) Zoom on simplified mesh

Figure 5: Cortex surface segmentation for two different resolutions: initial mesh and simplified mesh with 50%. This figure shows the possibility of identifying the sulcal and gyral structures even after a strong simplification.



(a) 100% initial mesh (b) Simplification of 30% (c) Simplification of 60%



(d) Simplification of 70% (e) Simplification of 80% (f) Simplification of 90%

Figure 6: Local segmentation of cortical surface for different resolutions.

Table 1: Possible combinations of surface types according to different descriptors.

Sign (K_H)	−	−	+	+
Sign (K_G)	+	−	−	+
Shape Index (SI)	[3/8, 1]	[1/8, 3/8]	[−3/8, −1/8]	[−5/8, −1]
Region type	Elliptic	Hyperbolic	Hyperbolic	Elliptic
Region color	Green	Cyan	Blue	Red

not able to compare our results with other works because of the lack of published papers with a complete quantitative result on the same images. We recall that we are mainly interested in investigating the ability of the method to identify the required structures in different resolutions. We present in this section some obtained results. Figure 5 shows two results for the segmentation (classification) of the same cortex: the first result (Figure 5(a)) is associated with the initial mesh and the second one (Figure 5(b)) with the simplified mesh (with 50%). We find that sulcal and gyral regions are well-preserved and still identifiable even after a significant reduction in the initial mesh. This result justifies our choice for an intermediate step of mesh simplification.

Figure 6 shows a hierarchy of the segmented cortex into gyral and sulcal structures. From this result, we see that the topology of large structures is conserved even after several simplifications of the initial mesh. It is noteworthy that this observation has been made also in [27]. It was also noted that gyral regions were better preserved than sulcal regions. Indeed, sulcal regions are thinner than gyral ones; thus, a simplifying operation will alter the topology of these small structures. From this result, we note also that the analysis (classification of the cortical surface) of low resolution has the advantage of clearly identifying the structures of interest with elliptic and hyperbolic forms. It should be noted at this point that the analysis of a higher resolution (more complex resolution) could be easily deduced from the other coarser resolutions. This reconstruction can be performed using basic inverse operators on the current mesh. Thereby, we obtain significant gain in computing time when we performed our algorithm on simplified meshes instead of initial complex ones.

The obvious difficulty in analyzing the obtained results is that there is no clear definition of what is correct? In some papers, authors rely on neurology experts to locate such sulci and gyri structures in the brain. Otherwise, the results can be interpreted based on the following idea: gyrus

is located in the upper zone of the cortex (that is to say, the upper folds) and sulcus is placed in the basins of the cortex. Consequently, it will be easy to distinguish between elliptical and hyperbolic shapes. The obtained results show that it is possible to detect main features on low resolution (only 30% of the original data), whereas, if we want to do the same thing on higher resolution (100%), we cannot interpret clearly these results nor recognize the local shape of the region of interest. In fact, the classification may include small irrelevant details, which have no effect on the overall shape. On the other hand, this process is considered time-consuming since discrete surfaces are so dense. Current tests were performed on data extracted from the ICBM (International Consortium for Brain Mapping Data Base). In the future, quantitative validations will be carried out to guarantee the performance of the method. We suggest also to study other datasets showing a notable morphological variability.

4. Conclusion

The 3D medical image analysis, especially the cortical surface analysis problem, is both an important and difficult task. Its applications are numerous such as in neurology. The goal of this work was to develop a method for local shape segmentation based on surface curvature and multiresolution scheme. Local curvature measures properties provide an effective shape measure and could be effectively applied to discrete surface analysis. Moreover, the multiresolution way provides the neurologist with a map of the patient brain in a short time. Overall, our method shows qualitatively interesting results. Although the experimental results are satisfying, there are some future works to do. As a perspective for further work, we plan to address other issues including (1) the integration of other geometrical criteria to improve the surface characterization, (2) the consideration of the connection between all vertices having same criteria values, (3) the conduction of quantitative evaluations to demonstrate the robustness of the algorithm and (4) the comparison of our results with other methods that could benefit for further research. So, further investigations are required to extend the algorithm to a large range of meshes showing notable morphological variability.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] G. le Goualher, C. Barillot, L. le Briquer, J. C. Gee, and Y. Bizais, "3d detection and representation of cortical sulci," *Computer Assisted Radiology*, pp. 234–240, 1995.

[2] P. Thompson and A. W. Toga, "A surface-based technique for warping three-dimensional images of the brain," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 402–417, 1996.

[3] D. C. van Essen, H. A. Drury, S. Joshi, and M. I. Miller, "Functional and structural mapping of human cerebral cortex: solutions are in the surfaces," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 3, pp. 788–795, 1998.

[4] Y. Wang, B. S. Peterson, and L. H. Staib, "Shape-based 3D surface correspondence using geodesics and local geometry," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 644–651, June 2000.

[5] P. G. Batchelor, A. D. Castellano Smith, D. L. G. Hill, D. J. Hawkes, T. C. S. Cox, and A. F. Dean, "Measures of folding applied to the development of the human fetal brain," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 953–965, 2002.

[6] S. Jaume, B. M. Macq, and S. K. Warfield, "Labeling the brain surface using a deformable multiresolution mesh," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI '02)*, vol. 2488 of *Lecture Notes in Computer Science*, pp. 451–458, Springer, 2002.

[7] J.-J. Kim, B. Crespo-Facorro, N. C. Andreasen et al., "An MRI-based parcellation method for the temporal lobe," *NeuroImage*, vol. 11, no. 4, pp. 271–288, 2000.

[8] D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin, and J. Régis, "Automatic recognition of cortical sulci of the human brain using a congregation of neural networks," *Medical Image Analysis*, vol. 6, no. 2, pp. 77–92, 2002.

[9] J. F. Mangin, F. Poupon, D. Riviere et al., "3d moment invariant based morphometry," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, vol. 2879 of *Lecture Notes in Computer Science*, pp. 505–512, Springer, Montreal, Canada, 2003.

[10] J. F. Mangin, D. Rivire, A. Cachia et al., "Object-based strategy for morphometryof the cerebral cortex," in *InformationProcessing in Medical Imaging*, C. Taylor and J. Noble, Eds., vol. 2732, pp. 160–171, 2003.

[11] G. L. Goualher, E. Procyk, D. Louis Collins, R. Venugopal, C. Barillot, and A. C. Evans, "Automated extraction and variability analysis of sulcal neuroanatomy," *IEEE Transactions on Medical Imaging*, vol. 18, no. 3, pp. 206–217, 1999.

[12] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system," *NeuroImage*, vol. 9, no. 2, pp. 195–207, 1999.

[13] A. Cachia, J.-F. Mangin, D. Rivière et al., "A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns," *IEEE Transactions on Medical Imaging*, vol. 22, no. 6, pp. 754–765, 2003.

[14] F. Vivodtzev, G. P. Bonneau, L. Linsen, B. Hamann, K. I. Joy, and B. A. Olshausen, "Hierarchical isosurface segmentation based on discrete curvature," in *Proceedings of the symposium on Data visualisation*, pp. 249–258, 2003.

[15] G. Lohmann and D. Y. von Cramon, "Automatic labelling of the human cortical surface using sulcal basins," *Medical Image Analysis*, vol. 4, no. 3, pp. 179–188, 2000.

[16] M. E. Rettmann, C. Xu, D. L. Pham, and J. L. Prince, "Automated segmentation of sulcal regions," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI '99)*, vol. 1679 of *Lecture Notes in Computer Science*, pp. 158–167, Springer, New York, NY, USA, 1999.

[17] C.-Y. Kao, M. Hofer, G. Sapiro, J. Stern, K. Rehm, and D. A. Rottenberg, "A geometric method for automatic extraction of sulcal fundi," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 530–540, 2007.

[18] G. le Goualher, C. Barillot, and Y. Bizais, "Modeling cortical sulci with active ribbons," *International Journal of Pattern*

Recognition and Artificial Intelligence, vol. 11, no. 8, pp. 1295–1315, 1997.

[19] A. P. Mangan and R. T. Whitaker, "Partitioning 3D surface meshes using watershed segmentation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 4, pp. 308–321, 1999.

[20] S. Pulla, A. Razdan, and G. Farin, "Improved curvature estimation for watershed segmentation of 3-dimensional meshes. manuscript," *IEEE Transactions on Visualization and Computer Graphics*, 2001.

[21] M. E. Rettmann, X. Han, C. Xu, and J. L. Prince, "Automated sulcal segmentation using watersheds on the cortical surface," *NeuroImage*, vol. 15, no. 2, pp. 329–344, 2002.

[22] S. Bourouis, K. Hamrouni, and M. Dhibi, "Multiresolution mesh segmentation of MRI brain using classification and discrete curvature," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications (VISAPP '08)*, pp. 421–426, January 2008.

[23] S. Bourouis and K. Hamrouni, "3d segmentation of mri brain using level set and unsupervised classification," *International Journal of Image and Graphics*, vol. 10, no. 1, pp. 135–154, 2010.

[24] S. Bourouis and K. Hamrouni, "Fully automatic brain tumor segmentation based on multi-modality mri and level-set," *Journal of Ubiquitous Systems and Pervasive Networks*, vol. 3, no. 2, pp. 47–54, 2011.

[25] W. E. Lorensen and H. E. Cline, "Marching cubes: a high-resolution 3-d surface construction algorithm," *ACM Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[26] R. T. Whitaker, "Reducing aliasing artifacts in iso-surfaces of binary volumes," in *Proceedings of the IEEE Symposium on Volume Visualization*, pp. 23–32, 2000.

[27] H. Hoppe, "New quadric metric for simplifying meshes with appearance attributes," in *Proceedings of the 10th IEEE Visualization Conference (VIS '99)*, pp. 59–66, October 1999.

[28] http://www.vtk.org/.

[29] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," in *Visualization and Mathematics III, Mathematics and Visualization*, pp. 113–134, 2002.

[30] S. Petitjean, "A survey of methods for recovering quadrics in triangle meshes," *ACM Computing Surveys*, vol. 34, no. 2, pp. 211–262, 2002.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 780835, 13 pages
http://dx.doi.org/10.1155/2014/780835

Research Article

Optimization of a Biometric System Based on Acoustic Images

Alberto Izquierdo Fuente,[1] Lara Del Val Puente,[2] Juan J. Villacorta Calvo,[1] and Mariano Raboso Mateos[3]

[1] *Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática, Universidad de Valladolid, Paseo Belén 15, 47011 Valladolid, Spain*
[2] *Departamento de Ciencia de los Materiales e Ingeniería Metalúrgica, Expresión Gráfica de la Ingeniería, Ingeniería Cartográfica, Geodesia y Fotogrametría Ingeniería Mecánica e Ingeniería de los Procesos de Fabricación, Universidad de Valladolid, Paseo del Cauce 59, 47011 Valladolid, Spain*
[3] *Facultad de Informática, Universidad Pontificia de Salamanca, Compañia 5, 37002 Salamanca, Spain*

Correspondence should be addressed to Alberto Izquierdo Fuente; alberto.izquierdo@tel.uva.es

Received 26 August 2013; Accepted 20 November 2013; Published 28 January 2014

Academic Editors: S. Bourennane and J. Marot

On the basis of an acoustic biometric system that captures 16 acoustic images of a person for 4 frequencies and 4 positions, a study was carried out to improve the performance of the system. On a first stage, an analysis to determine which images provide more information to the system was carried out showing that a set of 12 images allows the system to obtain results that are equivalent to using all of the 16 images. Finally, optimization techniques were used to obtain the set of weights associated with each acoustic image that maximizes the performance of the biometric system. These results improve significantly the performance of the preliminary system, while reducing the time of acquisition and computational burden, since the number of acoustic images was reduced.

1. Introduction

Biometric identification [1–3] is a subject of active research, where new algorithms and sensors are being developed. The most widely used identification systems are based on fingerprints, hand geometry, retina, face, voice, vein, signature, and so forth. The fusion of information from multiple biometric systems is also improving the performance of identification and verification systems [4].

Radar-based systems require expensive hardware and can be unreliable due to the very low reflection intensity from humans. Acoustic imaging provides a simple and cheap sensor alternative that allows obtaining very precise range and angular information. Particularly, in the acoustic field, there are two accurate and reliable classification systems for targets:

(i) animal echolocation, performed by mammals such as bats, whales, and dolphins, where nature has developed specific waveforms for each type of task [5, 6] such as the classification of different types of flowers [7];

(ii) acoustic signatures used in passive sonar systems [8, 9], which analyse the signal received by a target in the time-frequency domain.

There are few papers working on acoustic imaging in air for the detection of human beings. Moebus et al. [10, 11] worked with the ultrasonic band (50 kHz) using a 2D array and beamforming in reception. They analysed solid objects (poles and a cuboid on a pedestal) in their first work and human images more recently. They showed that humans have a distinct acoustic signature and proposed to model the echoes from the reflection parts of objects in the scene by a Gaussian mixture-model. Based on the parameters of this model, a detector could be designed to discriminate between person and nonperson objects.

In previous works, the authors of this paper developed multisensor surveillance and tracking systems based on

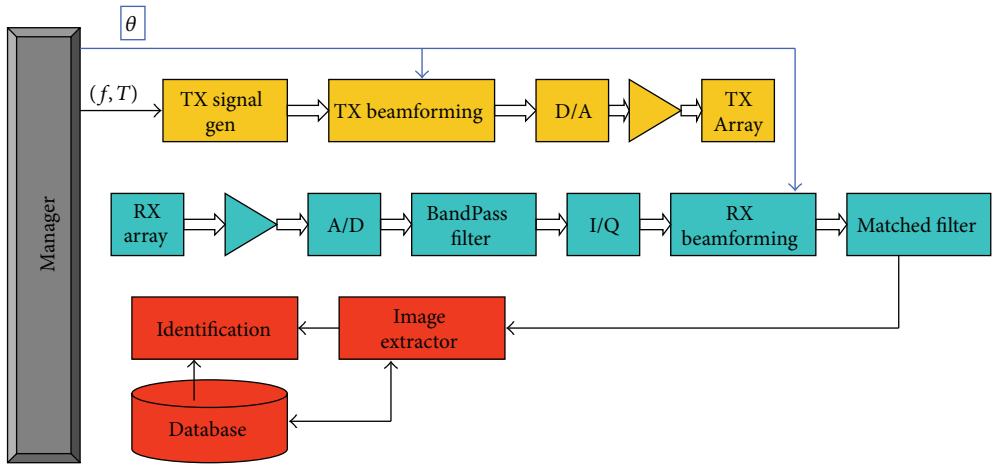


Figure 1: Functional description block diagram.

acoustic arrays and image sensors [12, 13]. In November, 2011, authors were working on the development of a novel biometric system, based on acoustic images acquired with electronic scanning arrays [14, 15]. Humans were acoustically scanned by an active system working from 6 to 12 kHz (audioband) that registered their acoustic images. Thus, the system could identify people by comparing the acquired acoustic images with a previously acquired database of said images. This system used beamforming with a linear microphone array and a linear tweeter array in transmission and reception, respectively [16]. This paper was the first one related to acoustic imaging in air for biometric identification of humans in the literature.

That previous work was based on 4 positions and 4 frequencies, and it evaluated the mean square error (MSE) between the acoustic images, assuming that all these images had the same weight in the error calculation and that all the images provided relevant information.

This new work has examined the contribution that each of the images associated with a position and a frequency has in the performance of the biometric system and has optimized the weights associated with the selected images.

On a first phase, the contribution of the acoustic images was analysed, assuming that their weights on the MSE were unitary or null, and working with a variable set of images from only 1 image to up to 16 images. After that, on a second phase, a weight optimization was done on the set of selected acoustic images, so that each image contributed to the calculation of the MSE proportionally to the information provided to the biometric identification between individuals.

In this paper, Section 2 describes the system including its functional description, its hardware architecture, the acoustic array, and the acoustic profiles. Section 3 describes the results previously published, which, implicitly, use unitary weights for all the images. This section also analyses the contribution of each image either individually or grouped with other images and the system performance by optimizing the weights for the selected images. Finally, Section 4 presents our conclusions.

2. Material and Methods

2.1. Functional Description. Based on basic radar/sonar principles [17, 18], an acoustic sound detection and ranging system for biometric identification was proposed [16], according to the block diagram in Figure 1.

This system performed three main tasks: (i) person scanning and detection, (ii) acoustic images acquisition, and (iii) person identification based on a database of acoustic images.

For each steering angle, the system performed: (i) transmission beamforming, (ii) reception beamforming, and (iii) match filtering. After processing all the steering angles, a two-dimensional matrix was formed and stored that this represented the acoustic image.

The application software developed had four operation modes:

(i) *Channel calibration*. A calibration procedure to ensure that all channels had the same phase and gain [19].

(ii) *Surveillance*. The system detected and estimated the position of the targets in the chamber, visualizing an acoustic image.

(iii) *Image acquisition*. The system captured the acoustic image of a person for a predefined set of frequencies and positions.

(iv) *Biometric identification*. For the person under analysis, the system got the acoustic images and compared them with a set of acoustic images of X individuals, previously stored in a database.

2.2. Hardware Architecture. The biometric system had four elements:

(i) a computer with a real-time acquisition system for 16 channels, based on 1.5 M gate FPGA Xilinx Spartan-3 DSP and two Omnibus I/O Daughter Card sites;

(ii) a preamplifier and amplifier system;

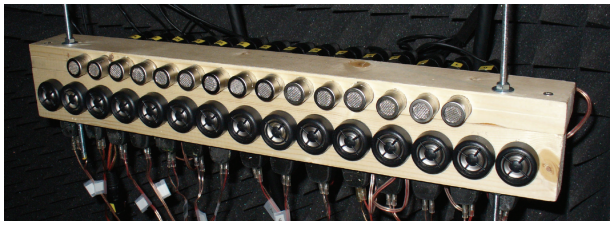


Figure 2: Transmitter and receiver arrays.

(iii) a transmitter (Tx) uniform linear array (ULA) with 15 tweeters and a receiver (Rx) ULA with 15 microphones, as it is shown in Figure 2;

(iv) an acoustic anechoic chamber with a 5 × 3 × 2.5 m working area which was designed for a 500 Hz cutoff frequency.

Figure 3 shows a block diagram of the system and the interconnection between its elements.

2.3. Acoustic Array

2.3.1. Spatial Aperture Selection. Two ULAs with 15 $\lambda/2$-equispaced sensors were employed. These arrays had different spatial apertures in order to reduce sidelobe levels on the final beampattern (Tx + Rx).

A transmission array with a 50 cm spatial aperture and a reception array with a 40 cm spatial aperture were used. On the transmission array, the tweeters were placed so as to occupy the minimum space.

2.3.2. Frequency Band Selection. On the basis of the angular resolution (3-dB beamwidth of the mainlobe), the absence of grating lobes, the frequency response of the microphone-tweeter pair, and the frequency response of a person, four different frequencies that guarantee the independence of the obtained images were selected [16]: 6 kHz (f_1), 8 kHz (f_2), 10 kHz (f_3), and 12 kHz (f_4), where the frequency gap was the maximum in order to obtain independent images.

The maximum steering angle was determined by the size of the person, his/her distance from the array, and the nonappearance of grating lobes. Based on these considerations, the following parameters were selected:

(i) the positioning area was located 3 m from the array;

(ii) the maximum width of a person with outstretched arms was 2 m.

Therefore, for the scanning and positioning area, the selected angle excursion was ±15°, as shown in Figure 4.

2.3.3. Angle Resolution Cells and Number of Beams. Given a ULA, Δu is defined as the 3-dB beamwidth of the mainlobe in the $\sin(\theta)$ *space*, where $\Delta u = \sin \Delta\theta$, having $\Delta\theta$ the 3-dB beamwidth of the mainlobe in degrees. Beamwidth in $\sin(\theta)$ *space* does not depend on the steering angle and, therefore, assuming that beams are 3-dB overlapped, the number of beams necessary to cover the exploration zone will be [20]:

$$M = \text{round}\left(\frac{2 \cdot \sin\theta_{\max}}{\Delta u}\right), \quad (1)$$

where $\theta_{\max} = 15^\circ$ is the angular excursion.

The number of beams for each frequency, M_k, is shown in Table 1.

Table 1: Number of beams versus frequency.

F (Hz)	$\Delta\theta$ (degrees)	Δu	M_k
6000	4.20°	0.0732	7
8000	3.20°	0.0558	9
10000	2.56°	0.0447	11
12000	2.12°	0.0370	13

2.4. Acoustic Profiles. Following the previous design considerations, the system retrieved the acoustic image associated with a rectangle of 2 m × 2.5 m (width × depth) dimensions, where the person under analysis had to be located 3m away from the line array, as described in Figure 4.

A 2 ms pulse width and a sampling frequency f_s = 32 kHz were used. This value was a trade-off between range resolution and received energy. The acoustic images were collected from 2.0 m to 4.5 m, in the range coordinate, and from −15° to 15°, in the azimuth coordinate, using M_k steering angles.

The selected positions for the person under analysis were front view with arms folded on both sides (p_1), front view with arms outstretched (p_2), back view (p_3), and side view (p_4). Figure 5 shows the four positions using a test subject.

The acoustic profile, P_i, associated with person i, included the 16 acoustic images obtained for the positions (p_1, p_2, p_3, and p_4), evaluated at the frequencies (f_1, f_2, f_3, and f_4).

Figure 6 shows the acoustic images for (i) the front view position (p_1) where the head and trunk of the subject can be clearly identified, (ii) the front view position with arms outstretched (p_2) where the head and arms of the subject can be clearly identified, (iii) the back view position (p_3) where the back of the head can be identified, and (iv) the side view position (p_4) where the closest shoulder and side of the head can be identified.

3. Results and Discussion

3.1. Previous Study: Biometric Identification via Mean Square Error

3.1.1. Metric Based on Mean Square Error (MSE). The identification implemented by the acoustic biometric system was based on the mean square error (MSE) between acoustic images from two different profiles [21].

First, a function $E_p^f[i, j]$ was defined as the mean square error between an acoustic image $I_i(r, s)$ from profile P_i and an

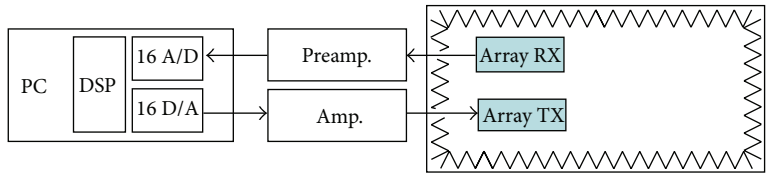


Figure 3: Hardware architecture block diagram.

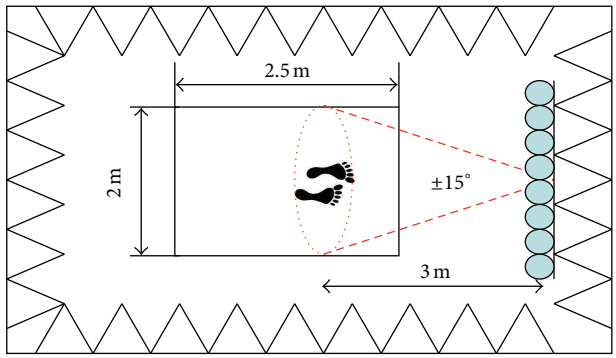


Figure 4: Scanning and positioning area.

acoustic image $I_j(r, s)$ from profile P_j, for a specific frequency f and position p:

$$E_p^f[i, j] = \sum_{r=1}^{R} \sum_{s=1}^{S} \left(I_i(r, s) - I_j(r, s)\right)^2, \quad i, j = 1 \cdots NP, \tag{2}$$

where $I(r, s)$ is a $R \times S$ matrix and NP is the number of acoustic profiles stored in the database.

Then, the multifrequency error function $E_p[i, j]$ was defined as the sum of the errors at each frequency for a specific position p:

$$E_p[i, j] = E_p^{6\,\mathrm{kHz}}[i, j] + E_p^{8\,\mathrm{kHz}}[i, j] + E_p^{10\,\mathrm{kHz}}[i, j] + E_p^{12\,\mathrm{kHz}}[i, j]. \tag{3}$$

Finally, the global error function $E[i, j]$ was defined as the sum of the multifrequency errors at each position p:

$$E[i, j] = E_{p_1}[i, j] + E_{p_2}[i, j] + E_{p_3}[i, j] + E_{p_4}[i, j]. \tag{4}$$

If P_k was an unknown profile to be identified, the algorithm associated the profile, P_k, to the person "i" in the database whose profile P_i had the minimum $E[k, i]$ value. The normalized global error was defined as the distance or metric used by the acoustic biometric system.

3.1.2. False Match Rate (FMR), False Nonmatch Rate (FNMR) and Receiver Operating Characteristic (ROC) Curve. Based on the methodology to characterize a biometric system [22] and assuming that there were no errors in the acquisition, FNMR and FMR parameters were calculated.

False match rate (FMR) is the probability of the system matching incorrectly the input acoustic profile to a non-matching template in the database. It measures the percent of invalid inputs which are incorrectly accepted. Thus, FMR was obtained by matching acoustic profiles of different people.

The global error $E[i, j]$ was calculated for all these cases. And then the FMR parameter was calculated as the percentage of matching whose error value was equal or less than distance d:

$$E[i, j] \leq d, \tag{5}$$

where distance d is the set of possible values of the global error.

False nonmatch rate (FNMR) is the probability of the system not matching the input acoustic profile to a matching template in the database. It measures the percent of valid inputs which are incorrectly rejected. Hence, FNMR was obtained by matching acoustic profiles of the same people.

Again, the normalized global error was calculated for all these cases. Then the FNMR parameter was calculated as the percentage of matching whose error value was greater or equal than distance d:

$$E[i, j] \geq d. \tag{6}$$

A receiver operating characteristic (ROC) curve, is a graphical plot which illustrates the performance of a classifier system as its discrimination threshold, distance d in this case, is varied. This ROC curve is a visual characterization of the trade-off between the FNMR and the FMR obtained. It was created by plotting the FMR values versus the FNMR values, at various threshold/distance settings.

3.1.3. Test Scenario. This acoustic biometric system, based on an electronic scanning array using sound detection and ranging techniques, was analysed in order to find the feasibility of employing acoustic images of a person as a biometric feature.

In this previous study [16], 10 people (5 men and 5 women with different morphological features, as shown in Table 2) were scanned in the four selected positions with a narrow acoustic beam, employing four pulsed tone signals, with the selected frequencies.

To evaluate this system, acoustic profiles were captured 10 times for each of the 10 people under test during 10 days. In the analysis, all people wore an overall, as common reference clothing, in order to eliminate clothing as a distinctive factor.

Figure 7 shows the FMR and the FNMR functions versus the normalized distance d obtained in the analysis.

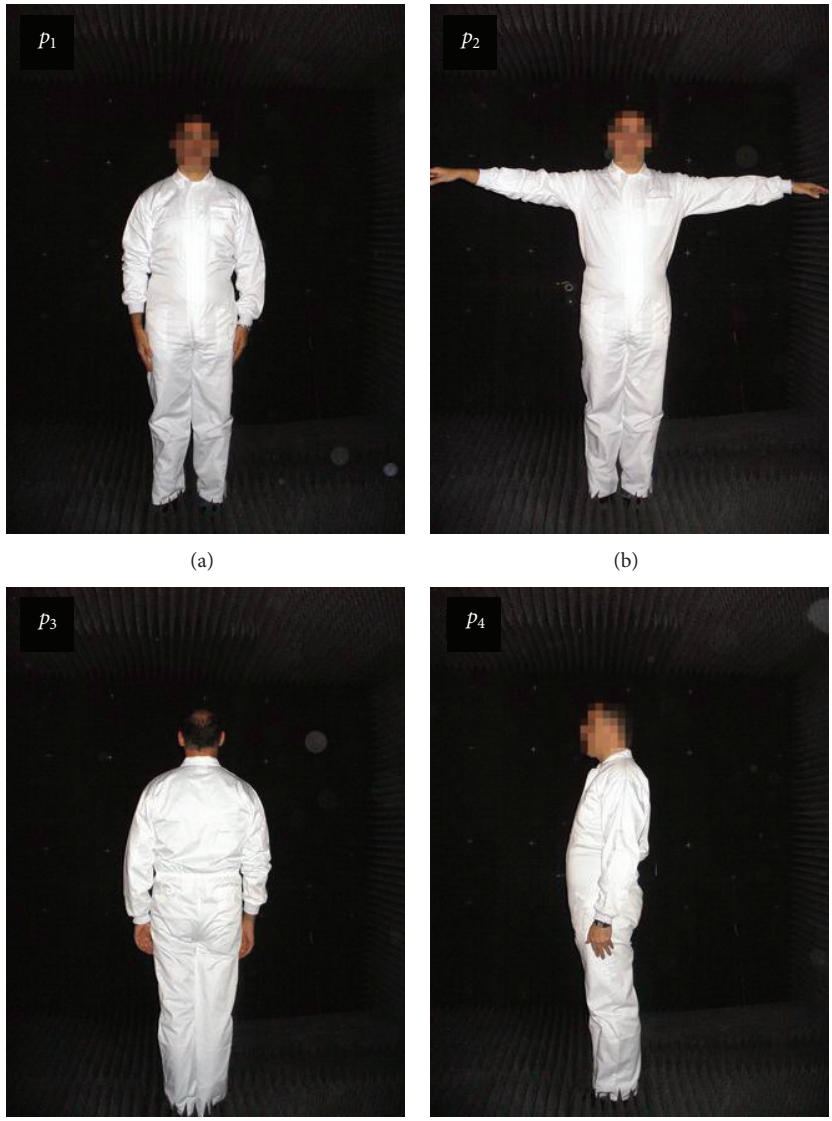


Figure 5: Person positions.

It can be observed that the value of the equal error rate (EER)—the crossing point between FMR and FNMR functions—was 6.22%, for a distance $d = 0.35$ m.

The corresponding ROC curve is shown in Figure 8.

The FNMR, FMR, and ROC curves obtained were comparable to those of commercial biometric systems, confirming the feasibility of using acoustic images in biometric systems.

3.2. Contribution of the Images to the Acoustic Profile. The global error $E[i, j]$ used in Section 3.1.1 can be reformulated as the sum of the errors due to each acoustic image associated with a frequency and a position:

$$E[i, j] = \sum_{f=1}^{4} \sum_{p=1}^{4} E_p^f [i, j]. \quad (7)$$

Generalizing this expression, the weighted global error $E_w[i, j]$, where the contribution associated with each image is weighted by a value w_p^f, can be defined according to the following expression:

$$E_w [i, j] = \sum_{f=1}^{4} \sum_{p=1}^{4} w_p^f E_p^f [i, j], \quad (8)$$

where the weights are defined between 0 and 1:

$$0 \leq w_p^f \leq 1. \quad (9)$$

For the case where all the images contribute to a unitary weight, the global error coincides with the weighted global error:

$$E[i, j] = E_w [i, j], \quad w_p^f = 1 \ \forall p \ \forall f. \quad (10)$$

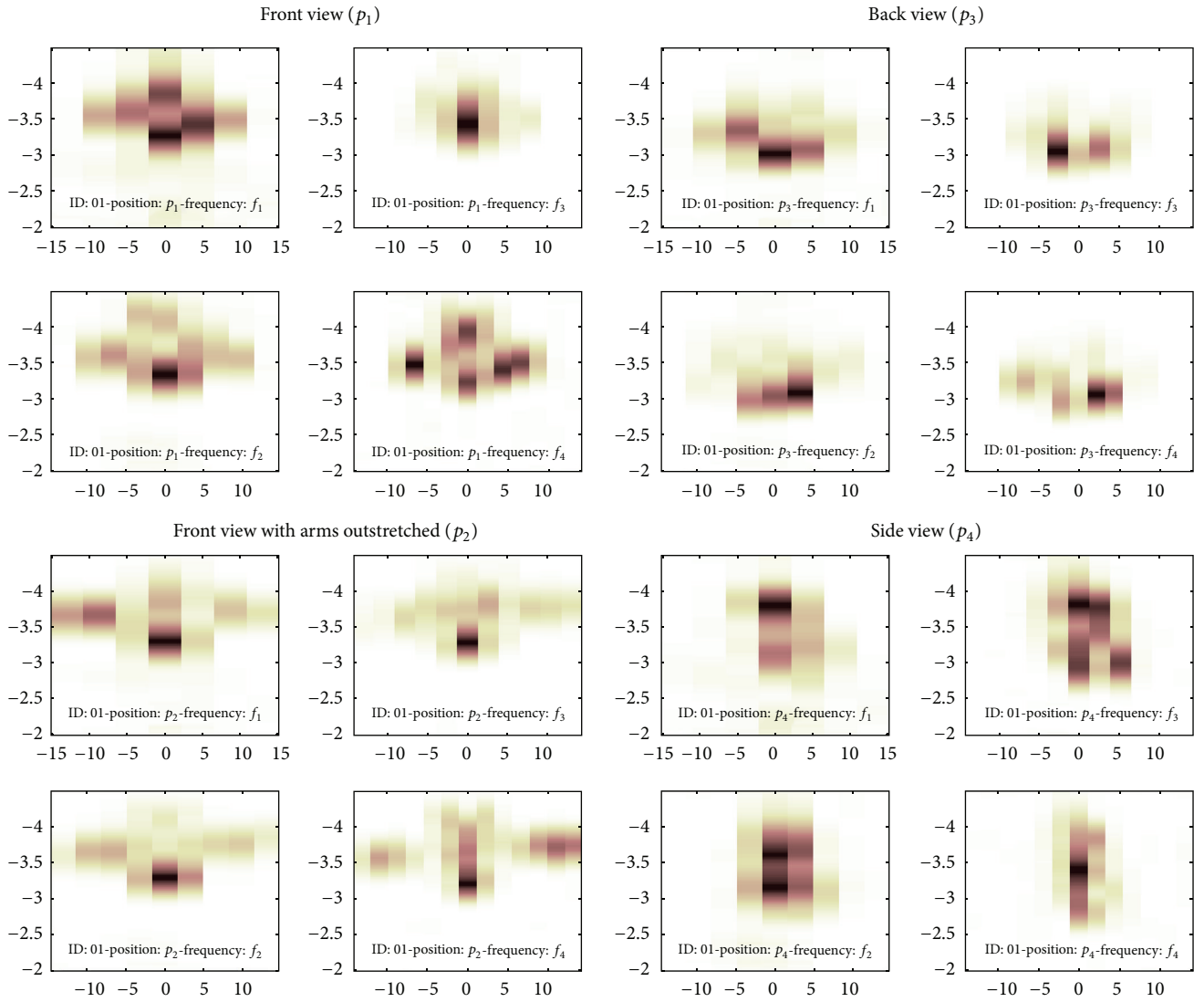


Figure 6: Acoustic images. x-axis: angle (degrees); y-axis: range (m).

An analysis to determine if all the images contribute equally in determining the ROC curve of the biometric system was performed. The hypothesis was that there will be images (associated with a position and a frequency) that provide more information than others. The goal was twofold: on the one hand, to detect the most relevant frequencies/positions and, on the other hand, to reduce the complexity of the system by eliminating those frequencies/positions that provided less information.

At this point, the information that an image provided had to be evaluated not only individually but also collectively to establish which images provided supplementary information. The ultimate goal was to obtain the set of images that allowed us to minimize the EER value associated with the system, taking the corresponding value using the global error as a reference and where all images contributed to unitary weights.

To evaluate the different hypotheses, a weight $w_p^f = 1$—to select an image—and a weight $w_p^f = 0$—not to select it—were defined.

The following studies were carried out:

(i) system analysis using a single image;

(ii) system analysis using all the images associated with a position;

(iii) system analysis using all the images associated with a frequency;

(iv) system analysis discarding all the images associated with a position;

(v) system analysis discarding all the images associated with a frequency;

(vi) System analysis discarding any N images.

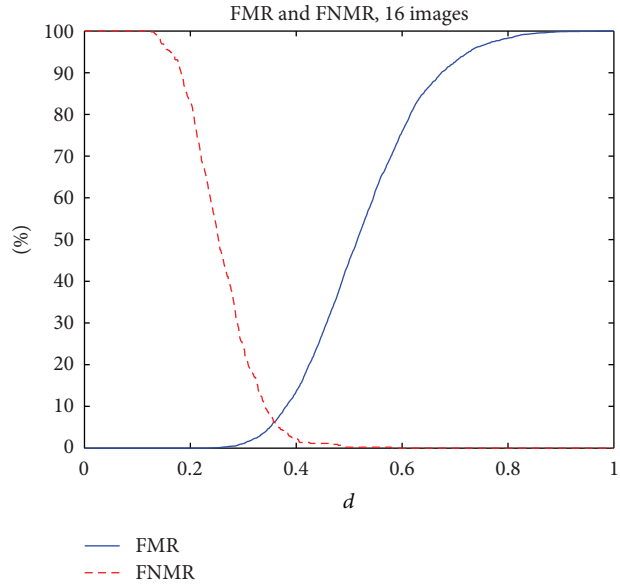


Figure 7: Functions FMR and FNMR versus distance d.

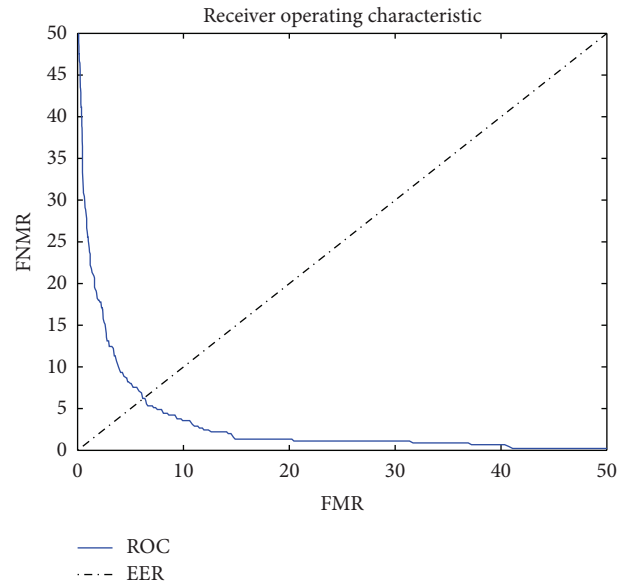


Figure 8: ROC.

3.2.1. Individual Images. In this case, the individual information—corresponding to a frequency f_i and a position p_i—that each acoustic image provided to the biometric system was analysed, assuming that the rest of the images were not present. In order to achieve this objective, a unitary weight was assigned to the image that corresponds to the selected position and frequency, while the rest of the images had null weights.

The result was equivalent to a biometric system consisting only of an acoustic image. Calculating the EER value of the system for each of the images, the following results were obtained, as shown in Table 3.

Table 2: Morphological features.

	Properties		
ID	Gender	Constitution	Height
00	Male	Very strong	Tall
01	Male	Strong	Average
02	Male	Strong	Average
03	Male	Thin	Tall
04	Male	Normal	Tall
05	Female	Thin	Tall
06	Female	Strong	Small
07	Female	Thin	Average
08	Female	Strong	Average
09	Female	Normal	Small

Table 3: Equal error rate using 1 image.

	EER-1 image			
Frequency	p_1 (front)	p_2 (front + arms)	p_3 (back)	p_4 (side)
f_1 (6 kHz)	30.38	16.97	31.81	23.93
f_2 (8 kHz)	33.44	**16.61**	28.67	27.09
f_3 (10 kHz)	**35.66**	20.96	33.33	26.28
f_4 (12 kHz)	34.60	16.00	33.89	26.44

It can be checked that the obtained values for each case were very different, resulting in a minimum value of EER = 16.61, for p_2 position (front with arms outstretched) evaluated at frequency f_2 (8 kHz) and a maximum value of the EER = 35.66, for p_1 position (front) evaluated at frequency f_3 (10 kHz). The ratio between EER maximum and minimum values was 2.14.

These results also highlighted that the images associated with position p_2 (front with arms outstretched) were the ones that provide the most information and, on the other hand, images associated with position p_1 (front) and p_3 (back) were those which provide the least information, since they were the columns that had higher EER values.

It became clear that each type of images provided different information and, therefore, it was not reasonable to assign all images the same contribution/weight to the error function.

It was also verified that the EER value for a single acoustic image was far superior to the value obtained when the 16 images were combined with unit weights (EER = 6.22). This indicated that a single image was not enough to constitute a biometric system based on acoustic signatures and that the combination of various frequencies/positions was essential to improve the system performance.

However, when minimizing the complexity of the system, the number of positions and frequencies was a relevant parameter. So, it was of great interest to determine whether the information associated with a position or a frequency provided more or less information than the remaining positions/frequencies. Therefore, the following two sections show the analysis of the performance of the system when using all the images associated with a frequency or a position.

Table 4: Equal error rate using 4 images associated with a position.

	EER-4 images			
	p_1 (front)	p_2 (front + arms)	p_3 (back)	p_4 (side)
$f_1 + f_2 + f_3 + f_4$	**25.80**	**9.52**	24.25	16.52

Table 5: Equal error rate using 4 images associated with a frequency.

	EER-4 images			
	f_1 (6 kHz)	f_2 (8 kHz)	f_3 (10 kHz)	f_4 (12 kHz)
$p_1 + p_2 + p_3 + p_4$	**13.78**	15.44	**17.56**	14.16

3.2.2. Images Associated with a Position. In this case, the joint information of the 4 images associated with a specific position, assuming that the rest of the images were not present, was analysed. This was achieved by assigning a unitary weight to those images corresponding to position p_i and a null weight to the rest of the images.

The result was equivalent to a biometric system consisting only of 4 acoustic images. Calculating the EER value of the system for each of the positions, the following results were obtained, as shown in Table 4.

This gave a minimum value of the EER = 9.52 for position p_2 (front with arms outstretched) and a maximum value of EER = 25.80 for position p_1 (front). The ratio between EER maximum and minimum values was 2.71.

It was evident that the use of 4 images associated with different frequencies improved substantially the EER values of the individual case. However, it was surprising that the EER value using a single image at position p_2 (the one associated with frequency f_2 = 8 kHz) was lower than some values obtained using 4 images (associated with positions p_1 or p_3).

Clearly, there were significant differences in the information associated with the different spatial positions.

3.2.3. Images Associated with a Frequency. In this case, the joint information of the 4 images associated with a specific frequency, assuming that the rest of the images were not present, was discussed. For this case, a unitary weight was assigned to the images corresponding to the frequency f_i and a null weight to the rest of the images.

The result was equivalent to a biometric system formed only by 4 acoustic images. Calculating the EER of the system for each of the frequencies, the following results were obtained, as shown in Table 5.

A minimum value of the EER = 13.78 for frequency f_1 (6 kHz) and a maximum value of the EER = 17.56 for frequency f_3 (10 kHz) were obtained. The ratio between EER maximum and minimum values was 1.27.

These results showed that using 4 images associated with different positions substantially improved the EER values of the individual case. In this case, the value of EER using a single image for the position p_2 (16.61) presented a value that was equivalent to the EER value using 4 images (13.78 –17.56).

Clearly, there were no significant differences in the information associated with the different frequencies. An EER = 9.52 using 4 frequencies for the position p_2 (front with arms outstretched) was obtained, clearly better than the EER = 13.78 using 4 positions for the frequency f_1 (6 kHz).

The EER values obtained with 4 images were superior to the EER values obtained using all the 16 images, so it was necessary to extend the information by increasing the number of images.

In the next two sections, 12 images were used, discarding the images that correspond to a particular position or frequency.

3.2.4. Images Discarding a Position. This case analysed the information from 12 images associated with three of the four positions, assuming that the rest of the images were not present. A null weight was assigned to the images corresponding to the position discarded, p_i, and a unitary weight to the rest of the images.

The result was equivalent to a biometric system consisting only of 12 acoustic images. Calculating the EER of the system for each of the cases, the following results were obtained, as shown in Table 6.

A minimum value of the EER = 5.79, excluding position p_1 (front), and a maximum value of the EER = 12.55, excluding position p_2 (front with arms outstretched), were obtained. Clearly, there were significant differences associated with the discarded positions (EER maximum and minimum ratio = 2.16).

The first conclusion was that better results can be obtained with 12 images (EER = 5.79) than with 16 images (EER = 6.22). Therefore, there were images associated with positions that clearly provided information that degraded the biometric system, rather than providing information to improve it.

In view of the previous results, position p_1 (front) was not significant in the presence of the information obtained from positions p_2, p_3, and p_4. It seemed evident that the system could remove the images associated with position p_1 in order to reduce its complexity.

As a second conclusion, in relation to the results obtained using 4 images associated with a position, except for the combination that excludes p_2, working with 12 images improved the performance of the biometric system. Note that the above combination did not use position p_2, which was shown to be the one that contributed to the most information to the system.

3.2.5. Images Discarding a Frequency. This case analysed the information from 12 images associated with three of the four frequencies, assuming that the rest of the images were not present. A null weight was assigned to the images corresponding to the frequency discarded, f_i, and a unitary weight to the rest of the images.

The result was equivalent to a biometric system formed only by 12 acoustic images. Calculating the EER of the system for each of the frequencies, the following results were obtained, as shown in Table 7.

A minimum value of the EER = 7.34, excluding frequency f_1 (6 kHz), and a maximum value of the EER = 8.68,

Table 6: Equal error rate using 12 images, discarding a position.

	EER-12 images			
	$p_2 + p_3 + p_4$	$p_1 + p_3 + p_4$	$p_1 + p_2 + p_4$	$p_1 + p_2 + p_3$
$f_1 + f_2 + f_3 + f_4$	**5.79**	**12.55**	7.94	8.19

Table 7: Equal error rate using 12 images, discarding a frequency.

	EER-12 images			
	$f_2 + f_3 + f_4$	$f_1 + f_3 + f_4$	$f_1 + f_2 + f_4$	$f_1 + f_2 + f_3$
$p_1 + p_2 + p_3 + p_4$	8.07	**7.34**	**8.68**	7.70

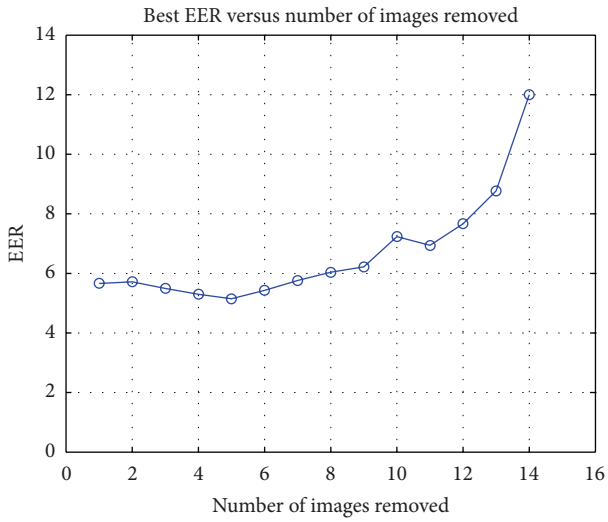


Figure 9: Equal error rate versus number of removed images.

excluding frequency f_3 (10 kHz) were obtained. Clearly the differences associated with the discarded frequencies were of little significance (EER maximum and minimum ratio = 1.18).

In this case, with 12 images (EER = 7.34), the system did not work better than with 16 images (EER = 6.22). Therefore, the use of multiple frequencies upgraded the biometric features of the system.

3.2.6. Discarding N Images. In view of the results, it was interesting to analyse the behaviour of the system when N images were discarded, where N was any number between 1 and 14. In preliminary studies, 15 images, 12 images, and 4 images were discarded but always grouped by frequency or by position.

In principle, discarding images means a reduction of information, which should be reflected as an increase in EER. However, in the previous section, it was shown that discarding 4 images associated with a position provided the best results. If this process of elimination of any frequency and position was generalized, lower EER values could be obtained.

This study was carried out to obtain the results shown in Figure 9.

The EER value had a minimum for the case $N = 5$, where the two combinations with lower EER were selected. For these two cases the images included/discarded are presented in Tables 8 and 9.

Given these results, and since the difference in the value of EER was small, the case with a value of EER = 5.19 was the selected candidate. This case allowed the complete elimination of all the images of position p_1 and, therefore, simplified the capturing of images of the person from 4 positions to 3. This represented a 25% reduction in acquisition time and in storage space.

By analysing the case $N = 4$ the EER function had a minimum value of 5.29. The two combinations with the smallest EER values were selected. Their results are shown in Tables 10 and 11.

Note that if combination number 4, which eliminated position p_1, was selected, a value of EER = 5.79, higher than the selected for $N = 5$, could be obtained. On the other hand, removing p_3-f_4 image improved the quality of the system, since both for $N = 4$ and for $N = 5$ the candidates with lower EER values did not include this image.

In conclusion, combination number 2 was selected with a value of EER = 5.19.

3.3. Weight Optimization. If, instead of quantifying the weights with unitary or null values, the value of the weights was optimized to minimize the weighted global error $E_w[i, j]$, a value of EER lower than the results of the previous section could be obtained.

The goal was to obtain the weights that minimized the weighted global error, defined by

$$\min(E_w[i,j])\big|_{w_p^f} = \sum_{f=1}^{4}\sum_{p=1}^{4} w_p^f E_p^f[i,j]. \quad (11)$$

Solving this optimization problem was complex because it was a multivariate optimization problem whose computational burden grew exponentially with the number of variables or weights.

The analysis of the total number of possible combinations required a very high computational cost in the order of C^{16}, where C was the number of different discretized weight values, making the problem directly unfeasible. Considering the results of Section 4, the number of weights to be optimized could be reduced from 16 to 11, decreasing the computational burden, although the process time was still too high.

A preliminary analysis of the error function indicated that it was a nonconcave space with multiple local minima, so

Table 8: Equal error rate with 5 images removed (combination number 1).

EER = 5.14															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
1	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1

Table 9: Equal error rate with 5 images removed (combination number 2).

EER = 5.19															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1

Table 10: Equal error rate with 4 images removed (combination number 3).

EER = 5.29															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1

Table 11: Equal error rate with 4 images removed (combination number 4).

EER = 5.79															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1

those algorithms based on the technique of the gradient could not be used. In practice, an optimization algorithm based on Powell's method [23] was used. This algorithm was based on directional searches and recursion and it significantly reduced the computational burden.

Firstly, the 11 weights associated with the images that were selected in the previous section were optimized. After that, the optimization with 16 images was performed in order to verify whether the exclusion of images had reduced the system performance.

3.3.1. Optimization with 11 Images. The obtained results yielded a value of EER = 4.17. The optimal vector of weights is shown in Table 12.

Then, in order to validate whether the deleted information contained in the 5 discarded images could improve the biometric performance of the system, an optimization was carried out for the 16 images.

3.3.2. Optimization with 16 Images. The obtained results yielded a value of EER = 4.00. The optimal vector of weights is shown in Table 13.

Note that the weights associated with position p_1 were much lower than the weights for the other positions. This validates the hypothesis that the data associated with this position provided very little information to the biometric system.

Since the EER value obtained with 16 images was lower than the EER value achieved with 11 images, the next step was to analyse whether increasing the number of images could improve the performance of the system.

3.3.3. Optimization with 12 Images. In this case, multiple combinations were tested, obtaining a value of EER = 4.0 for the case that discarded all the images of position p_1, as shown in Table 14.

It is not necessary to analyse the results of a larger number of images, since in this case with $N = 12$ images, the obtained EER value was equivalent to the case of $N = 16$. Therefore, discarded images did not provide meaningful information to the biometric system.

Optimal vectors for $N = 12$ and $N = 16$ were quite different. However, in both cases, the net information was the same, due to the fact that the value of the obtained EER was equivalent. Using 16 images, the information was redundant and therefore the information could be distributed among multiple images. But, using 12 images the information could only be obtained from the 12 selected images.

In any case, it should be noted that there were multiple combinations of weights which lead to the same value of EER for a fixed number of images. This fact showed that the function has multiple minima, as it was a very complex error surface.

Figure 10 shows the ROC functions for the case of $N = 12$ with optimized weights compared to the case $N = 12$ with unitary weights.

It is observed that the optimization process had significantly improved the performance of the biometric system.

Table 12: Equal error rate with 11 images with optimized weights.

EER = 4.17															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0	**0**	**0**	**0**	0.35	0.80	0.64	0.35	0.51	0.64	0.13	**0**	0.99	0.71	0.30	0.51

Table 13: Equal error rate with 16 images with optimized weights.

EER = 4.00															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0.05	0.02	0.06	0.07	0.63	0.70	0.26	0.49	0.46	0.52	0.17	0.38	0.65	0.52	0.13	0.39

Table 14: Equal error rate with 12 images with optimized weights.

EER = 4.00															
p_1				p_2				p_3				p_4			
f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
0	**0**	**0**	**0**	0.51	0.59	0.36	0.26	0.30	0.44	0.11	0.15	0.92	0.52	0.20	0.31

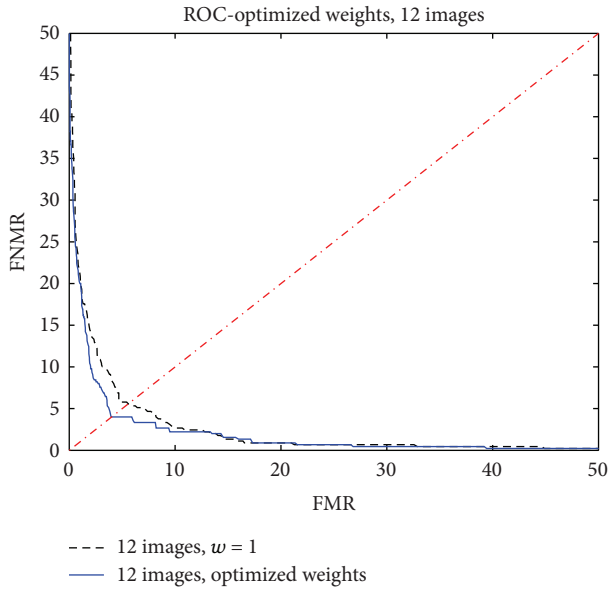


Figure 10: ROC function with 12 images: unitary weights versus optimized weights.

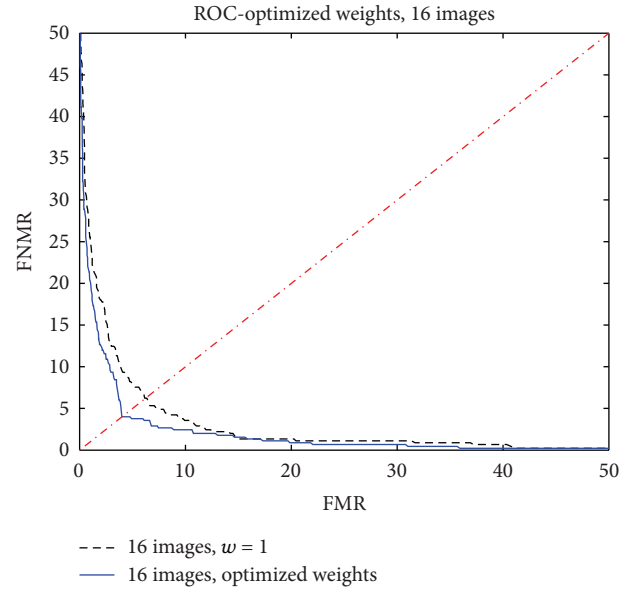


Figure 11: ROC function with 16 images: unitary weights versus optimized weights.

In a similar way, Figure 11 shows the ROC functions for the case of $N = 16$ with optimized weights compared to the case $N = 16$ with unitary weights, previously published and summarized in Section 3.1.

Again, it can be observed that the optimization process had improved the performance of the biometric system.

Finally, Figure 12 shows ROC functions for the case of $N = 12$ with optimized weights, comparing it to the case $N = 16$, also with optimized weights.

It is observed that the performance obtained with 12 images was equivalent to the one obtained with 16 images.

This study highlighted that the selection of 12 images along with optimization techniques allowed a substantial improvement in the performance of the biometric system while reducing the number of images required.

The original biometric system using 16 images and unitary weights yielded a value of EER = 6.22, and the new system using 12 images and optimized weights yielded a value of EER = 4.00. There was an improvement of over 30%.

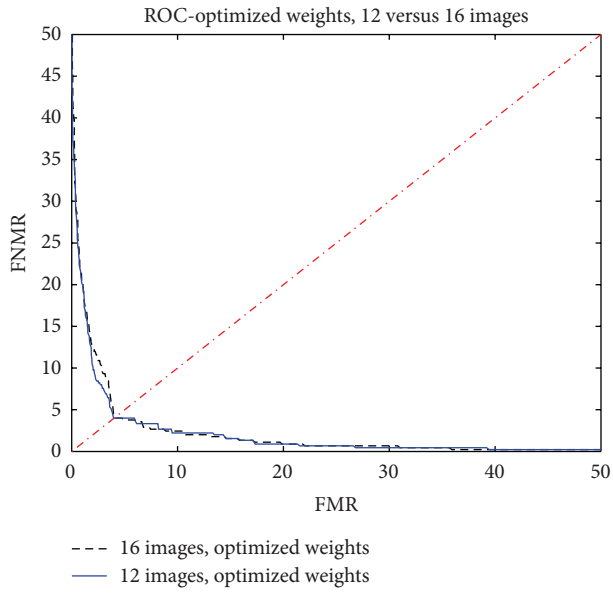


Figure 12: ROC function with 12 images versus 16 images with optimized weights.

4. Conclusions

Based on the results obtained in a preliminary publication, where 16 acoustic images of a person—working with 4 frequencies and 4 positions—were used, a methodology for the selection of the most significant images in the face of the biometric system performance was developed.

Each acoustic image that is associated with a position and a frequency provides and shares information that allows to discriminate people from each other.

On a first stage, the contribution of each acoustic image to the biometric system was analysed, assuming that all the images had a unitary or a null weight. We reached the conclusion that with 11 images we can obtain the same performance that with the 16 images. In addition, the images associated with the front position (p_1) are those that provide less information, since much of it can be obtained from the images of the remaining positions. This analysis was carried out measuring the value of EER and selecting an increasing number of images, until the value of EER was minimized.

Afterwards, on a second stage, weights for 11 images were optimized, where the EER value obtained was close to the one obtained optimizing 16 images. We arrived at the conclusion that using 12 acoustic images, which correspond to the positions front with arms outstretched, side and back, the minimum value of the EER can be obtained. This EER value coincides with the value obtained for 16 images.

On the basis of the developed methodology, the selection of acoustic images made on the first stage reduced the number of images and, therefore, significantly reduced the computational burden of the optimization. It was confirmed that the selected acoustic images are essentially the images that must be included in the optimization stage.

Currently, the research group is analyzing the system performance using new frequencies and new metrics not based on MSE.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[2] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proceedings of the 46th International Symposium Electronics in Marine (Elmar '04)*, pp. 184–193, June 2004.

[3] International Biometric Group, "Comparative Biometric Testing. Round 7 Public report V1.2," 2009, http://l.b5z.net/i/u/6084428/i/CBT7_IBGReport.pdf.

[4] T. Joshi, S. Dey, and D. Samanta, "Multimodal biometrics: state of the art in fusion techniques," *International Journal of Biometrics*, vol. 1, no. 4, pp. 393–417, 2009.

[5] C. Baker, M. Vespe, and G. Jones, "50 million years of waveform design," in *Proceedings of the IET Forum on Waveform Diversity and Design in Communications, Radar and Sonar*, pp. 7–21, London, UK, November 2005.

[6] A. Balieri, K. Woodbridge, C. Baker, and M. Holderied, "Flower classification by bats: radar comparisons," *IEEE Aerospace and Electronic Systems Magazine*, vol. 24, no. 5, pp. 4–7, 2009.

[7] D. V. Helversen, M. W. Holderied, and O. V. Helversen, "Echoes of bat-pollinated bell-shaped flowers: conspicuous for nectar-feeding bats?" *Journal of Experimental Biology*, vol. 206, no. 6, pp. 1025–1034, 2003.

[8] L. F. Chevalier, *Principles of Radar and Sonar Signal Processing*, Artech House, Boston, Mass, USA, 1st edition, 2002.

[9] D. W. Ricker, *Echo Signal Processing*, Kluwer, Dordrecht, The Netherlands, 1st edition, 2003.

[10] M. Moebus and A. M. Zoubir, "Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. 961–964, Honolulu, Hawaii, USA, April 2007.

[11] M. Moebus, A. M. Zoubir, and M. Viberg, "Parametrization of acoustic images for the detection of human presence by mobile platforms," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 3538–3541, Dallas, Tex, USA, March 2010.

[12] J. Duran de Jesus, A. Izquierdo Fuente, and J. J. Villacorta Calvo, "Multisensorial modular system of monitoring and tracking with information fusion techniques and neural networks," in *Proceedings of the IEEE 33rd Annual International Carnahan Conference on Security Technology*, pp. 59–66, October 1999.

[13] A. Izquierdo-Fuente, J. Villacorta-Calvo, M. Raboso-Mateos, A. Martines-Arribas, D. Rodríguez-Merino, and L. Del Val-Puente, "A human classification system for a video-acoustic detection platform," in *Proceedings of the IEEE 38th Annual International Carnahan Conference on Security Technology*, pp. 145–152, Albuquerque, NM, USA, October 2004.

[14] P. S. Naidu, *Sensor Array Signal Processing*, CRC Press, Boca Raton, Fla, USA, 1st edition, 2001.

[15] B. D. van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[16] A. Izquierdo-Fuente, L. Del Val, M. I. Jiménez, and J. J. Villacorta, "Performance evaluation of a biometric system based on acoustic images," *Sensors*, vol. 11, no. 10, pp. 9499–9519, 2011.

[17] D. K. Barton, *Radar System Analysis and Modeling*, Artech House, Norwood, Mass, USA, 2005.

[18] M. I. Skolnik, *Introduction to Radar Systems*, McGraw Hill, New York, NY, USA, 3rd edition, 2001.

[19] L. del Val, M. I. Jiménez, J. J. Villacorta-Calvo, A. Izquierdo, and M. Raboso, "A calibration methodology for an acoustical array," in *Proceedings of the Sensor Signal Processing for Defense*, London, UK, 2011.

[20] W. D. Wirth, *Radar Techniques Using Array Antennas*, IEE Radar, Sonar, Navigation and Avionics Series 10, The Institution of Electrical Engineers, London, UK, 2001.

[21] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley and Sons, San Diego, Calif, USA, 2nd edition, 2001.

[22] A. K. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*, Springer, New York, NY, USA, 1st edition, 2008.

[23] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, vol. 7, no. 2, pp. 155–162, 1964.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 393958, 10 pages
http://dx.doi.org/10.1155/2014/393958

Research Article

An Improved Real Time Image Detection System for Elephant Intrusion along the Forest Border Areas

S. J. Sugumar[1] and R. Jayaparvathy[2]

[1] *Department of Electrical and Electronics Engineering, Coimbatore Institute of Technology, Avinashi Road, Civil Aerodrome PO, Coimbatore, Tamil Nadu 641 014, India*
[2] *Department of Electronics and Communication Engineering, SSN College of Engineering, Kalavakkam, Old Mahabalipuram Road, Chennai, Tamil Nadu 603 110, India*

Correspondence should be addressed to R. Jayaparvathy; jayaparvathy14@gmail.com

Received 12 August 2013; Accepted 21 October 2013; Published 20 January 2014

Academic Editors: S. Bourennane and J. Marot

Human-elephant conflict is a major problem leading to crop damage, human death and injuries caused by elephants, and elephants being killed by humans. In this paper, we propose an automated unsupervised elephant image detection system (EIDS) as a solution to human-elephant conflict in the context of elephant conservation. The elephant's image is captured in the forest border areas and is sent to a base station via an RF network. The received image is decomposed using Haar wavelet to obtain multilevel wavelet coefficients, with which we perform image feature extraction and similarity match between the elephant query image and the database image using image vision algorithms. A GSM message is sent to the forest officials indicating that an elephant has been detected in the forest border and is approaching human habitat. We propose an optimized distance metric to improve the image retrieval time from the database. We compare the optimized distance metric with the popular Euclidean and Manhattan distance methods. The proposed optimized distance metric retrieves more images with lesser retrieval time than the other distance metrics which makes the optimized distance method more efficient and reliable.

1. Introduction

The Asian elephant (*Elephas maximus*) is highly threatened by habitat fragmentation, habitat loss, and human-elephant conflict. India hosts 60% of Asian elephant population, nearly two-thirds of the elephant population lives either close to or within human-dominated landscapes. Southern India harbors half of India's elephant population containing about 6300 elephants [1]. The increase in human population in India propelled by agricultural and industrial growth has led to the conversion of the forest lands into human settlements. Due to this, the wild elephant and other animal populations face acute shortage of resources such as water and food, making them move often into the human habitat. Hence, there has been severe man-elephant conflict. The conflict has been on the rise in the forest border areas with herds of wild pachyderms straying into human habitation [2]. The surveillance and tracking of these herds are difficult due to their size and nature of movement. The time to recover from the danger is negligible; hence, the loss due to destruction in the farms is more. The elephants are also subject to attack by humans resulting in danger to the life of elephants. According to the authors in [3], poaching for ivory had indeed become a threat with 100–150 tuskers being lost annually to illegal killings.

Human-elephant conflict (HEC) is a key example of the growing competition between people and wildlife for space and resources throughout Africa and Asia. This study explores the correlation of reported HEC incidents within 58 villages between 80 km from the boundary of Kallar to Walayar, Coimbatore, Tamil Nadu, India. Habitat loss and fragmentation is the biggest threat to the continuing survival of Asian elephants in this region. In addition to food crops, forests are being logged for their timber or cleared to make space for cash crop plantations such as rubber, tea, and palm. As the human population has inexorably risen, the forest

wild lands in which elephants live have been disappearing. Human-elephant conflict is on the rise and it is a battle that the elephant is losing. As elephant habitat diminishes, the elephants are pushed into increasingly smaller areas. This increases the population density to beyond sustainable levels and food availability grows short. The shortage of fodder has a negative impact on rates of reproduction; hence, normal birth rates begin to decrease. The serious consequence of the shortage of wild food leads to a corresponding increase of crop raiding and incidents of human-elephant conflict [3].

Human-elephant conflict is a rapidly expanding area of research, with conservationists working hard to understand the circumstances under which tensions are the highest between humans and elephants. A number of factors contribute to such conflicts, including population density of humans, elephant habitat structure, weather, time of year, and animal life [4]. A study made in the region of interest shows that elephants move into human habitation due to many reasons.

(i) Fences and trenches compromised by people who need access to forests.

(ii) Farm lands may funnel them to unprotected adjacent villages.

(iii) Badly planned barriers that do not take elephant behavior into consideration.

(iv) Denying elephant access to a critical water source or foraging area.

(v) Human activities create abundant secondary vegetation that brings elephants closer to human settlements.

(vi) Artificially maintained water sources attract elephants during drought.

(vii) Traditional migration routes severed by human intervention (e.g., canals, power installations, and cattle fences).

The obvious conclusion to be drawn is that there is no single cause or explanation to account for human-elephant conflict; situations are circumstantial and complex. Rather, elephants and agriculture mix in numerous ways with varying consequences. Human population growth and land occupation for settlement may heighten conflict with elephants. However, it is generally the borders of forests that are the focal points of conflicts. Minimizing human-elephant conflict to reduce the risk of life of both human beings and elephants is of utmost importance. Elephant conservation issues can be divided into two distinct categories:

(1) activities that affect elephants directly such as hunting/poaching and capture;

(2) developmental activities and human activities leading either to the loss of elephant habitat or its qualitative degradation.

Many methods are followed to avoid HEC. Construction of elephant proof trenches is being done all over the world. In [5], Fernando et al. discussed solar fencing to avoid elephant-human conflict. In [6], King et al. presented the concept of using beehives to mitigate elephant crop depredation. In [7], Loarie et al. discussed about the role of the artificial water sources which allow elephants to reside in forests during dry seasons. In [8], the authors discussed the potential use of satellite technology for conflict mitigation. The elephants tagged with radio collars react violently and damage it and even the elephants die. In [9], Venter and Hanekom proposed the possibility of using the elephant-elephant communication (elephant rumbles) to detect the presence of a herd of elephants in close proximity, In this work, the authors have recorded the low frequency infrasound pattern, but they do not compare with that of other animals to confirm an elephant occurrence. In [10], Vermeulen et al. proposed unmanned aircraft system to survey elephants, in which the elephant images are acquired at a height of 100 m but the small flight time and being expensive do not make it viable. In [11], Dabarera and Rodrigo proposed appearance based recognition algorithms for identification of elephants. Given the frontal face image of an elephant, the system searches the individual elephant using vision algorithms and gives the result as, already identified elephant, or as a new identification. In [12], Ardovini et al. present an elephant photo identification system based on the shape comparison of the nicks characterizing the elephant's ears. In [13], Goswami et al. addressed identifying elephants from photographs, and comparing resultant capture recapture-based population parameter estimates using supervised visual identification of individual variations in tusk, ear fold and lobe shape. The authors show that this is a reliable technique for individual identification and subsequent estimation of population parameters. But in real time, the capture of elephant's front image is not possible.

It is easier to chase elephants before they enter fields and therefore most damage can be averted [3]. Guarding from watch towers, patrolling, and trip wire alarms provide farmers with advance warning of approaching elephants. Once the animals are detected, active crop guarding devices using light and noise are deployed to chase them away. An early warning system to minimize the human-elephant conflict in the forest border areas is proposed in this paper. The system helps mitigate such conflicts in two ways:

(i) providing warning to people about the anticipated entry of elephants into human habitation;

(ii) providing advance information to the authorities to take action to chase the pachyderms back to the forest.

An early warning system to minimize the human-elephant conflict in the forest border areas using image processing is proposed in this paper. The system helps to detect the elephants even in the presence of other wild animals like Bison, Tiger, and Deer, and so forth. The system also identifies the elephants coming in groups. The reliability of elephant detection is tested and the time to detect the elephant images is optimized with the proposed optimized distance metric.

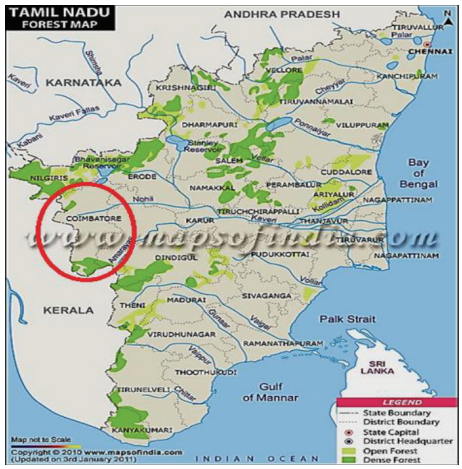


Figure 1: Map showing study area of the Coimbatore district in Tamil Nadu.

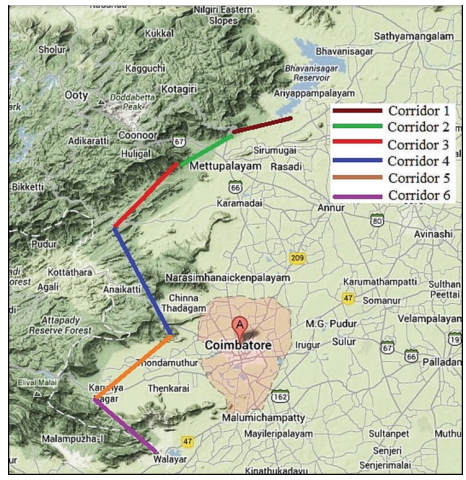


Figure 2: Corridors used by Elephants in the forest border area of Coimbatore.

2. Study Area

Studies were made in the Coimbatore Forest Division, Tamil Nadu, India, as shown in Figure 1. The data were collected from the Coimbatore Forest Department website and interviews with village people affected by elephants and relevant literature. Coimbatore district is richly endowed with hills, forests, rivers, and wildlife. Geographical area of the district is 74,433.72 sq. km with a forest area of 693.48 sq. km (9.33%). The forest of Coimbatore district is divided into two divisions.

South of Palghat Gap lies in the Anamalai Wildlife Sanctuary, which has been designated as a Tiger Reserve in 2008. North of Palghat lies in the Coimbatore forest division. This division is bounded in the north and northwest by Sathyamangalam, Erode, Nilgiris North, and Nilgiris South forest divisions and in the west and southwest by Palghat forest division of Kerala. Coimbatore Forest Division is spread over 693.48 sq. km in six ranges, of which 400 sq. km is conflict prone. The division has 58 villages and 315 km of forest boundary.

Six elephant corridors within the Coimbatore forest division are shown in Figure 2. Namely, Jaccanari-Vedar Colony (Corridor 1) in which the length of the corridor is about 12 km and width ranges from 0.2 to 1.5 km. In Kallar-Jaccanari (Corridor 2), the length of the corridor is 7 km and width ranges from 0.2 to 1.5 km. In Kallar-Nellithurai (Corridor 3), the length of the corridor is 10.8 km and width ranges from 0.25 to 3 km. In Anaikatti-Veerapandi (Corridor 4), the length of the corridor is 21 km and effective width ranges from 0.1 to 1.5 km. In Maruthamalai-Thanikandy (Corridor 5), the length of the corridor is 13 km and effective width ranges from 0.4 to 1.5 km. And in Kalkothi-Walayar (Corridor 6), the length of the corridor is 21 kms the effective width ranges from 0.2 to 0.9 km.

In total, there are 85 kms of elephant corridor in the district needs to be protected from human-elephant conflict. The region is also a part of the crucial elephant corridor in this region [4] amounting to a total of 691–914 elephants found in this region. The elephant is one of the most conflict-prone wildlife species in India, causing large-scale damage to crops and human lives. Each year, nearly 400 people and 100 elephants are killed in conflict related instances in India, and nearly 500,000 families are affected by crop damage. Several reasons including habitat fragmentation, degradation of habitat quality, loss of forest cover, laxity in management of physical barriers, and other causes have been cited for the human-elephant conflict (HEC) in the country.

The human habitats bordering the forests around Coimbatore city in south western India are seeing severe human-elephant conflict as the expanding human population propelled by industrial and agricultural growth is increasingly fragmenting elephant habitat in this landscape. The number of incidents of elephants straying into farm lands was 680 in 2011, 844 in 2010, and 560 in 2009. The number of people killed in elephant attacks in Coimbatore was 13 in 2012, 8 in 2011, 15 in 2010, and 11 in 2009. The number of elephants killed by such conflicts was 4 in 2012, 1 in 2011, 1 in 2010 and 2 in 2009 as shown in Figure 3 (Courtesy: the Hindu, Coimbatore, February 5, 2013).

In this region, train hit accidents occur frequently when the elephants try to pass the rail track Walayar, Coimbatore, railway section which is on the forest border area. Thus it has resulted in the death of 20 elephants in the last five years. However, it is generally the edges of forest that are to be protected as those are the focal points of conflict.

Humans go into the forest to graze cattle in day time and guard crops at night and therefore run a higher risk of being killed by elephants. They also try to ride elephants by throwing stones and sticks for which the elephants react hard and even kill the humans. Elephants that wait near villages for nightfall to eat crops have also been known to kill people [3]. During the period 1999–2011, there had been 1,822 incidents of crop damage and 53 cases of property damage caused by elephants and the compensation disbursed was Rs. 2.19 crore.

3. Material and Methods

In our previous work [2], seismic geophones used as sensors are buried under the ground to detect the movements of

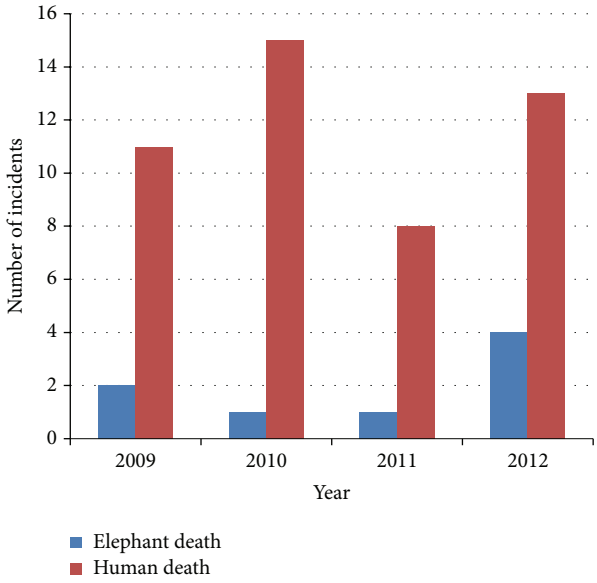


Figure 3: Human and elephant loss in Coimbatore district.

elephants in forest border areas. Elephants walking in the sensing range of the geophones produce vibration which is converted to an electrical signal, processed in an embedded controller and an SMS is sent to the forest officials for necessary action. In such intrusion detection systems, there is a possibility of insufficient vibrations sensed by the geophones as a consequence of weather conditions like rain and soil moisture. Due to this, it may miss the event, that is, elephant movement. An image processing based approach is developed as a solution to the above said problem to identify an intruding elephant in human living areas. The elephant comes out of the forest through certain pockets to enter into human living areas for getting food and water. The cameras mounted on towers or trees capture the image of the intruding elephant which is sent to the base station using RF network. The received image is processed in a PC at the base station and is compared with the stored database image of elephants. The snapshot from the video is taken every 5 seconds and compared with the database image. This image is also updated in database and added. On an image match, an SMS is sent to the forest officials through the GSM transceiver connected with the PC. The hardware setup consists of wireless camera, PIR sensor with signal control module, Atmega microcontroller, GSM module, and the power supply. PIR motion sensor detects the movement and it switches the camera to the capture mode; the camera captures the image over a 20-meter distance. The whole hardware setup is shielded with metal cast to protect from the rain. The elephant pockets in the corridors are identified and these setups are installed to monitor the movement of these herds.

3.1. Elephant Image Detection System. Elephant Image Detection System (EIDS) algorithm is developed in this work. A database of 114 images is created by capturing 2 elephants in different postures. Images of elephants in the Sadivayal elephant camp in Coimbatore, South India, have been taken and used in this work. The elephant database images are feature-extracted using Haar wavelet technique and clustered into groups by using K-means clustering. A similarity comparison is made by determining the number of significant coefficients in common between the query signature and the signatures of the database using F Norm theory. The searched elephant images are then arranged according to the similarity value obtained in a decreasing order. If the matched images are more than 5, an elephant detected message is sent through the GSM to the mobile phone numbers stored in the system. Once the procedure is complete, the system captures the next image and performs the same steps to detect elephants. The Haar wavelet decomposition of elephant image in RGB color space is represented at multiple scales [14]. The Haar wavelet decomposition is computed by iterating difference d_i and average a_i between odd and even samples s_i of the elephant image. Averaging and differencing the elephant image elements are done as follows:

$$a_i = \frac{s_i + s_{i+1}}{2}, \qquad d_i = \frac{s_i - s_{i+1}}{2}. \tag{1}$$

If an elephant image data set $S_1, S_2, \ldots, S_{N-1}$ contains N elements, there will be $N/2$ averages and $N/2$ wavelet coefficient values [15]. The averages are stored in the upper half of the N element array and the difference coefficients are stored in the lower half of the array. The averages become the input for the next step in the wavelet computation, for iteration $i + 1$, $N_i = N_{i/2}$. The recursive iterations continue until a single average and a single difference are calculated [16]. The scaling and wavelet values are represented by h_i and g_i, respectively, and are given in (2) and (3). The values of scaling coefficients are given as

$$h_0 = 0.5, \qquad h_1 = 0.5, \tag{2}$$

and the values of wavelet coefficients are given as

$$g_0 = 0.5, \qquad g_1 = -0.5. \tag{3}$$

The Haar transform is shown in matrix form as follows:

$$\begin{matrix} h_0 & h_1 & 0 & 0 & \cdots \\ g_0 & g_1 & 0 & 0 & \cdots \\ 0 & 0 & h_0 & h_1 & \cdots \\ 0 & 0 & g_0 & g_1 & \cdots \end{matrix} \tag{4}$$

The Haar transform for an eight element signal is shown in the following equation. Here, the signal is multiplied by the forward transform matrix A:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ d_0 \\ d_1 \\ d_2 \\ d_3 \end{bmatrix} = A \cdot \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix}, \tag{5}$$

where

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.5 \end{bmatrix}. \quad (6)$$

Since the columns of the A_i's are orthogonal to each other, each of these matrices is invertible with respect to A_i. The elephant database images are decomposed into multilevel coefficients from -1 to $-J$ levels. After decomposition, feature vectors for all the elephant images in the database are obtained using F-norm theory [17] as given in (7) and (9). Every image is considered as a square matrix. A is a square matrix and A_i is its ith-order submatrix. The F-norm of A_i is given as

$$\|A_i\|_F = \left[\sum_{K=1}^{i} \sum_{I=1}^{i} |a_{KI}|^2 \right]^{1/2}. \quad (7)$$

Let

$$\Delta A_i = \|A_i\|_F - \|A_{i-1}\|_F, \qquad \|A_0\|_F = 0. \quad (8)$$

The feature vector of A is defined as

$$V_{AF} = \{\Delta A_1, \Delta A_2, \ldots, \Delta A_n\}. \quad (9)$$

Vector elements in the feature vector are represented by ΔA_i and ΔB_i. The similarity between the two images is given by the following similarity criteria [16]. Let α_i be the similarity of ΔA_i and ΔB_i as follows:

$$\alpha_i = \begin{cases} \dfrac{\min(\Delta A_i, \Delta B_i)}{\max(\Delta A_i, \Delta B_i)}, & \Delta A_i \neq 0 \text{ or } \Delta B_i \neq 0, \\ 1, & \Delta A_i = \Delta B_i = 0. \end{cases} \quad (10)$$

Similarity between the two images lies in between $0 \leq \alpha \leq 1$. The images in the database are arranged according to the similarity match with the query image.

3.2. K-Means Clustering Algorithm. Clustering is a process of grouping the similar objects from a given data set. The most popular and reliable clustering algorithm is the K means clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. Let $S = \{S_i, i = 1, 2, \ldots, N\}$ be the n-dimensional data points to be clustered into a set of K-clusters, $C = \{C_1, C_2, C_3, \ldots, C_K\}$ [18] from the given elephant data set $X = \{x_1, \ldots, x_N\}$, $x_n \in E^d$. The M-clustering problem aims at partitioning the elephant data set into M disjoint subsets (clusters) $C_1, \ldots, C_M$. The most widely used clustering criterion is the Euclidean distance [19]. Based on this criterion, the clustering of elephant images is grouped depending on the cluster centers $m_1, \ldots, m_M$ as given below

$$E(m_1, \ldots, m_M) = \sum_{i=1}^{M} \sum_{k=1}^{M} I(x_i \in C_k) \|x_i - m_k\|^2, \quad (11)$$

where $I(X) = 1$ if X is true and 0 otherwise.

3.3. Proposed Optimized Distance Metric. In this paper, a novel distance metric called optimized distance measure integrated with K-means clustering algorithm to improve retrieval time is proposed. We have used the distance metrics in the work for (i) finding similarity between two images and (ii) ordering a set of images based on their distances from a given image. In many image retrieval systems, Euclidean [18] and Manhattan [20] are the popular distance measure algorithms used. We carried out a study on the above two similarity measures and proposed a new distance method called optimized distance measure. The proposed method retrieved more images with faster retrieval rate than the other two methods. The Euclidean distance measures [18] are suitable for the correlation between quantitative and continuous variables and are not suitable for ordinal data and it is given as

$$D = \sqrt{(R_c - R_g)^2 + (G_c - G_g)^2 + (B_c - B_g)^2}, \quad (12)$$

where (R_c, G_c, B_c) are centroids and (R_g, G_g, B_g) are the pixel points or data points. Most of the time is spent to calculate the square root, so it is basically time consuming. The Manhattan distance is the absolute sum of the horizontal and vertical components of the image data matrix. This is essentially a consequence of being forced to adhere to single-axis movement; one cannot move diagonally in more than one axis simultaneously and, is given in the following equation:

$$\text{Distance } d = \sum_{s_i \in c_k} \|s_i - \mu_k\|. \quad (13)$$

Whenever each pair is in nonempty intersection, there exists an intersection point for the whole collection; therefore, the Manhattan distance forms an injective metric space. In the proposed optimized distance method for the given query image only the distance related cluster is searched. The optimized distance metric is given as

$$D = \left\| \sum_{i=0}^{n} \left[(R_c - R_g)^3 + (G_c - G_g)^3 + (B_c - B_g)^3 \right] \right\|. \quad (14)$$

The optimized method computes the cube power of the distance between the centroid and color pixel points of the three colors and determines the summation of all the added values. The modulus of the whole summed values is calculated to get the distance value.

4. Results and Discussion

The field observations are carried out in the forest border areas in Sadivayal elephant camp. The hardware setup was arranged to capture the image of elephants. The wireless camera was mounted on a wood stick and the camera was battery powered with 12 V. Using RF receiver, the video received is converted to image frames using camcorder software in PC. The elephant image frames are stored in the PC memory and updated every 5 seconds. Figure 4 shows

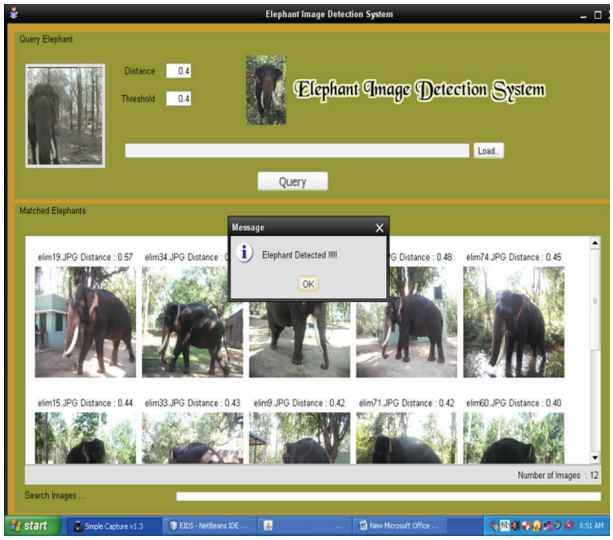


Figure 4: Retrieved online elephant image.

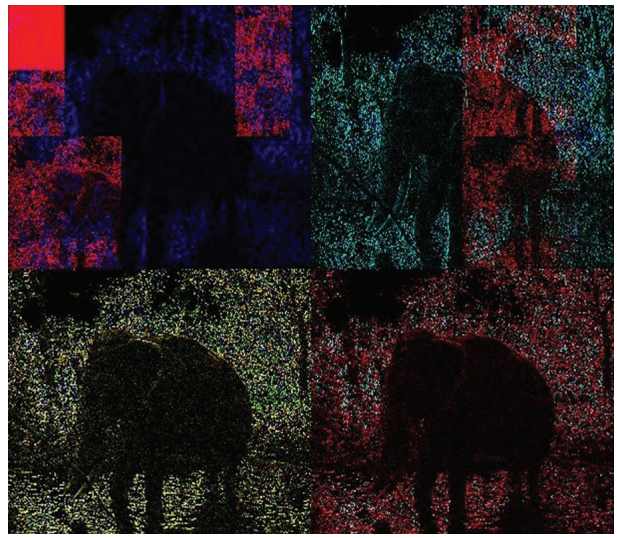


Figure 5: 3 Level decomposed query elephant image.

the online EIDS window in which 12 elephant images are retrieved in 6.33 seconds. The elephant images are arranged in the order of similarity value obtained. As the retrieved images are more than 5, a GSM message "Elephant Detected" is sent using the AT command.

In this work, we perform Haar wavelet decomposition on the raw elephant image, by determining the scaling coefficient and largest wavelet coefficients. The scaling coefficient is stored along with the difference and location (i, j) of each wavelet coefficient for every image [15]. The 3-level decomposed query elephant image is shown in Figure 5. Using the 3-level wavelet decomposition, the highest and informative elephant image features are extracted from the coefficients. These features are used during the process of the query and database image comparison of the elephants. The Haar wavelet transform of the elephant is calculated by passing it through a series of filters (high and low pass filters) and then downsampled, as we can see from Figure 6.

At each level, the elephant image is decomposed into low and high frequencies, and this decomposition halves the resolution since only half the number of samples is retained to characterize the entire image. The Haar wavelet leads to a decomposition of coefficients at level j in four components, and at level $j + 1$. Due to successive downsampling by 2, the image length must be a power of 2, or a multiple of a power of 2, and the length of the image determines the maximum levels into which the elephant can be decomposed.

The Haar wavelet coefficients of different species are plotted and shown in Figure 7. The Haar wavelet coefficient of each species varies with the elephant image and can be distinguished from other animals.

The value of wavelet coefficients of elephant and bison is closer because the major color is black for both species and the value lies in between 90 and 95. The coefficients for tiger and deer possess higher band in between 120 and 130. The obtained elephant coefficient is averaged to get the threshold value. We fix 0.6 as threshold value to obtain the similar

Figure 6: Downsampled decomposed elephant image.

elephant images from the database image for the given query image.

The EIDS system is tested offline with elephant and nonelephant images. All the images used in this work are in the dimensions of 3648 × 2763. We also tested the system for group of elephant images and it is shown in Figure 8.

The elephant normally moves in herds in the forest borders during the period of migration. For the given query image, the system retrieved 13 images in 6.462 seconds. We tested the system with group of elephant images of different sizes and postures. Nonelephant images like bison, bear, deer, monkey, and human which are the most commonly seen species around forest border areas are given as query, which produced zero image search result. We tested the case with a bison image as it is of similar color texture as that of elephants for which the system retrieved zero search images. On zero search result, no alert is made; hence, GSM message is not sent.

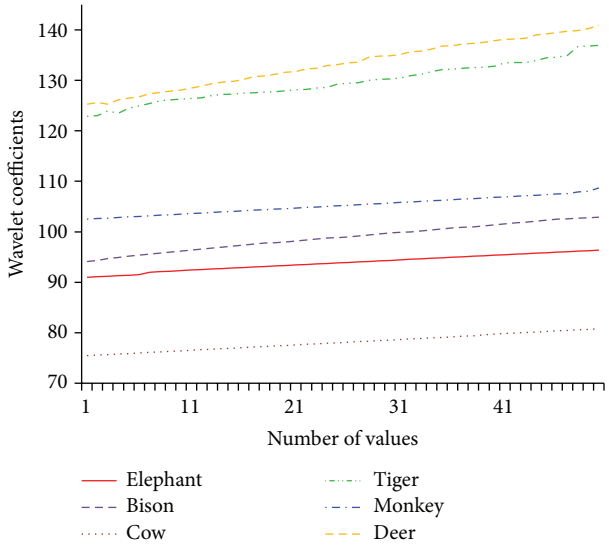


Figure 7: Wavelet Coefficients for Different Species.

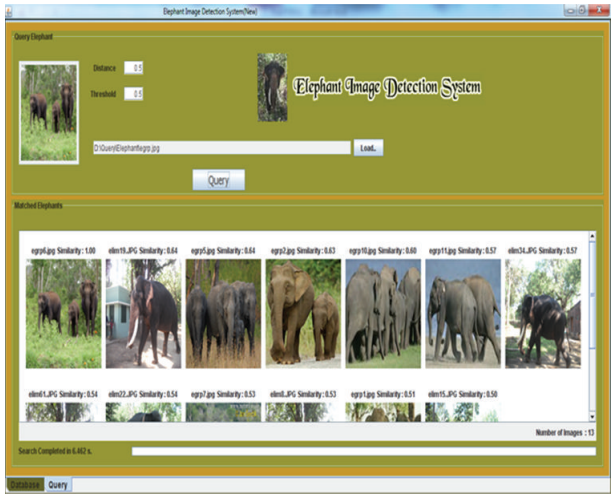


Figure 8: Retrieved elephant group image.

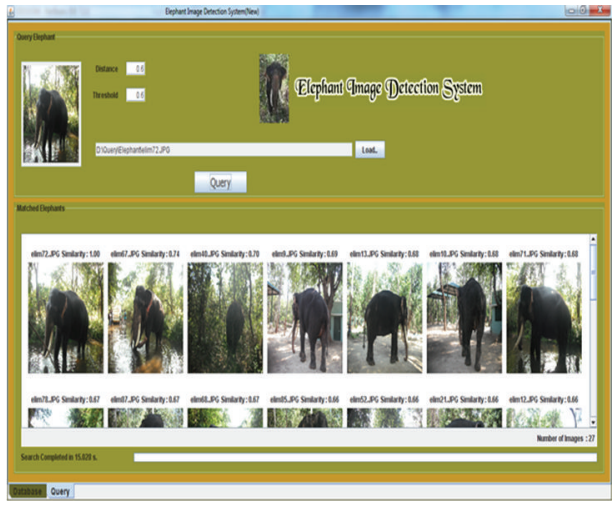


Figure 9: Retrieved images for optimized distance metric.

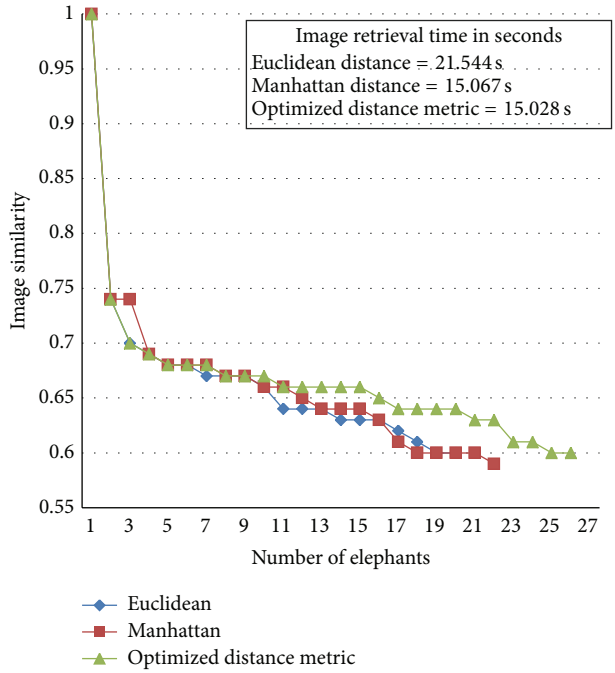


Figure 10: Image similarity metric comparison.

We compared the three distance measures with the number of images retrieved, retrieval time, and the retrieval rate per image. The Euclidean scheme produced 20 retrieved images with retrieval time of 21.544 seconds and the retrieval rate per image is 1.077 seconds. The Manhattan scheme produced 22 retrieved images with retrieval time of 15.067 seconds and the retrieval rate per image is 0.684 seconds. For the proposed optimized distance metric, the images retrieved were 27 as shown in Figure 9 with retrieval time of 15.028 seconds and the retrieval rate per image is 0.556 seconds.

A comparison between all the distance methods is carried out in this work and the results are shown in Figure 10. It is observed that the proposed method retrieves more images with lesser time compared to the other two methods.

The optimized retrieval rate improvement over another distance metric is 18%. To assess the retrieval effectiveness, 10 query elephant images are selected and tested. The average retrieval time per query elephant image is calculated for all the three distance measures and the results are shown in Figure 11. The retrieval time per image is less for the proposed distance metric compared to other methods as the metric retrieves more images which are closer to the cluster center. The images are also retrieved in lesser time interval. Due to reduction of computational time and higher image retrieval rate, the time to react for the elephant intrusion is improved using the proposed method.

The recall rate [21] is defined as the ratio of the number of relevant (same shape and position) retrieved images to the total number of images in the database

$$\text{Recall rate} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images in database}}. \quad (15)$$

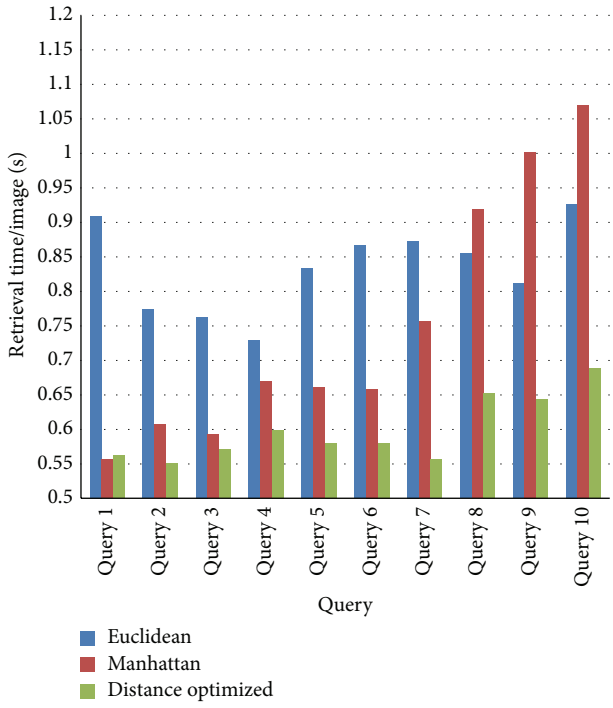


Figure 11: Image retrieval rate comparison.

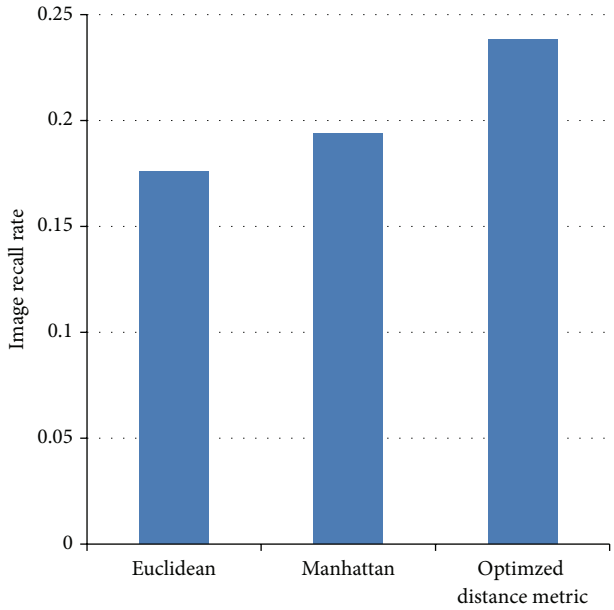


Figure 12: Elephant image recall rate comparison.

We compare image recall rate for the three methods as shown in Figure 12. It is observed that the optimized distance metric is 16% better than Manhattan and 18.5% better than Euclidean methods. We also tested the algorithm by varying the cluster formation. Table 1 shows the variation of number of images retrieved, retrieval time, and retrieval rate per image.

When we fix 2 clusters, 14 images were retrieved and the retrieval time for the 14 images was found to be 10.884 and the retrieval rate per image is 0.777.

Table 1: Database clustering for different K values.

K	Number of retrieved images	Retrieval time for images	Retrieval rate per image
2	14	10.884	0.777
3	12	7.604	0.6336
4	12	7.242	0.6032
5	9	4.949	0.5498

Table 2: Comparison between the online and offline observations.

Image threshold	Image distance		Retrieved images	
	Offline	Online	Offline	Online
0.4	1.0–0.49	0.57–0.36	22	12
0.5	1.0–0.49	0.44–0.34	22	6
0.6	1.0–0.59	0.47–0.38	17	5
0.7	1.0–0.74	0	2	0

We varied the cluster K value from 2 to 5 and recorded the corresponding number of retrieved images and the retrieval rate per image from the database for the query image given. It is seen that for small number of clusters more images were retrieved and for higher number of clusters the retrieved images were less in number as shown in Figure 13. So for our analysis, we fix 2 clusters in order to retrieve more images from the database.

A comparison was made between of offline and online elephant image detection system. In the online system for the threshold values 0.4 to 0.6, the number of images retrieved varied from 5 to 12 and for offline system, it produced 17 to 22 images. This variation is due to the camera posture and image capture in frame. The required number of 5 images is achieved in online system with the proposed optimized distance method that makes the system efficient and reliable. The result comparison between the online and offline detection system is recorded in the Table 2.

5. Conclusion

In conclusion, the findings of our work contribute to elephant conservation issues. The work provides solutions to human-elephant conflict. The study provides insights to protect elephants from human activities and reduces the work effort of forest officials. The real time elephant identification system provides solutions to the problem of human elephant conflict and provides solution for unsupervised process of individual species identification specifically for elephants. The system is completely automated; the strength of this approach stems from the ability to narrow down the collection of potential matches in the database with the query image. Optimal results for automated identification of individual elephants are obtained with the algorithm developed and is used to rank the most likely matches, followed by final supervised visual identifications and also with an early warning sent to the forest officials about the arrival of elephants from the forest borders into the human habitat. The real time automated

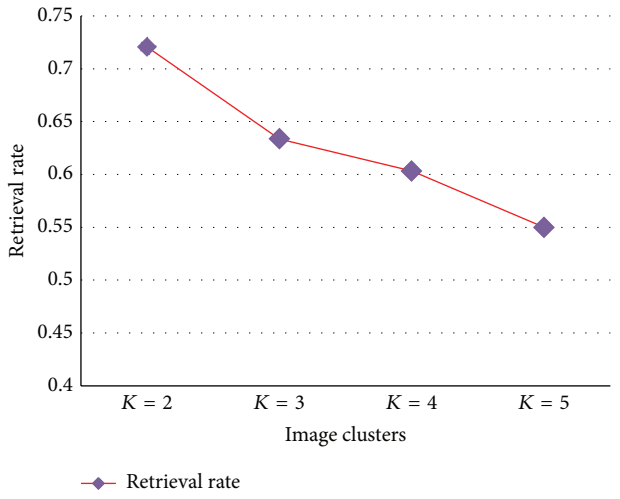


Figure 13: Image retrieval rate per cluster.

approach minimizes the manual work which is not possible all the time because it is difficult to monitor the presence of elephants manually when the herds march towards the forest borders. More importantly, our results demonstrate the importance of certainty in identifying approaching elephants in to human living areas and provide early warning about the elephant entry into the human habitat. We therefore recommend the use of the real time image processing technique to identify an approaching individual elephant as well as a group of elephants. The system can be implemented in forest border areas. We also solve the traits for distinguishing species where animals are differentiated on the basis of Haar wavelet color attributes. In the context of elephant conservation, the real time automated image processing system can be used for the elephant ecology learning, population monitoring, elephant habitat usage, commercial ivory poaching and human-elephant conflict. The system can also be deployed along forest border migration routes or at water holes and food plantation areas for elephant tracking and monitoring.

In this paper, we proposed and analyzed an elephant image detection system via wavelet decomposition of images, followed by feature extraction and similarity match under F-norm theory. We compared the retrieval performance of optimized distance metric based K-means clustering with the existing techniques like Euclidian distance and Manhattan distance. It turns out that optimized distance metric calculation has 18.5% improvement in recall rates. Field observations show that the proposed method can be used as an effective scheme to detect elephants in the forest border areas even in the presence of different species. This system has been rigorously tested through the various phases of the project and found to be efficient compared to the existing systems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] H. N. Kumara, S. Rathnakumar, M. Ananda Kumar, and M. Singh, "Estimating Asian elephant, elephasmaximus, density through distance sampling in the tropical forests of BiligiriRangaswamy Temple Tiger Reserve, India," *Journal Tropical Conservation Science*, vol. 5, no. 2, pp. 163–172, 2012.

[2] S. J. Sugumar and R. Jayaparvathy, "An early warning system for elephant intrusion along the forest border areas," *Current Science*, vol. 104, pp. 1515–1526, 2013.

[3] J. Lenin and R. Sukumar, "Action plan for the Mitigation of Elepant-Human conflict in India," Final Report to the U.S. Fish and Wildlife Service, Asian Nature Conservation Foundation, Bangalore, India, 2011.

[4] C. Arivazhagan and B. Ramakrishnan, "Conservation perspective of Asian Elephants (*Elephasmaximus*) in Tamil Nadu, Southern India," *International Journal of Biotechnology*, vol. 1, pp. 15–22, 2010.

[5] P. Fernando, M. A. Kumar, A. C. Williams, E. Wikramanayake, T. Aziz, and S. M. Singh, "Review of human-elephant conflict mitigation measures practiced in South Asia," AREAS Technical Support Document Submitted to World Bank, WWF-World Wide Fund for Nature, Vaud, Switzerland, 2008.

[6] L. E. King, A. Lawrence, I. Douglas-Hamilton, and F. Vollrath, "Beehive fence deters crop-raiding elephants," *African Journal of Ecology*, vol. 47, no. 2, pp. 131–137, 2009.

[7] S. R. Loarie, R. J. V. Aarde, and S. L. Pimm, "Fences and artificial water affect African savannah elephant movement patterns," *Biological Conservation*, vol. 142, no. 12, pp. 3086–3098, 2009.

[8] A. B. Venkataraman, R. Saandeep, N. Baskaran, M. Roy, A. Madhivanan, and R. Sukumar, "Using satellite telemetry to mitigate elephant-human conflict: an experiment in northern West Bengal, India," *Current Science*, vol. 88, no. 11, pp. 1827–1831, 2005.

[9] P. J. Venter and J. J. Hanekom, "Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings," *Biosystems Engineering*, vol. 106, no. 3, pp. 286–294, 2010.

[10] C. Vermeulen, P. Lejeune, J. Lisein, P. Sawadogo, and P. Bouche, "Unmanned aerial survey of elephants," *PLoS ONE*, vol. 8, no. 2, pp. 1–7, 2013.

[11] R. Dabarera and R. Rodrigo, "Vision based elephant recognition for management and conservation," in *Proceedings of the 5th International Conference on Information and Automation for Sustainability (ICIAfs '10)*, pp. 163–166, December 2010.

[12] A. Ardovini, L. Cinque, and E. Sangineto, "Identifying elephant photos by multi-curve matching," *Pattern Recognition*, vol. 41, no. 6, pp. 1867–1877, 2008.

[13] V. R. Goswami, M. V. Lauretta, M. D. Madhusudan, and K. U. Karanth, "Optimizing individual identification and survey effort for photographic capture-recapture sampling of species with temporally variable morphological traits," *Animal Conservation*, vol. 15, no. 2, pp. 174–183, 2012.

[14] P. Porwik and A. Lisowska, "The haar-wavelet transform in digital image processing: its status and achievements," *Machine Graphics and Vision*, vol. 13, no. 1-2, pp. 79–98, 2005.

[15] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1513–1521, 2003.

[16] M. Y. Latha, B. C. Jinaga, and V. S. K. Reddy, "Content based color image retrieval via wavelet transforms," *International*

Journal of Computer Science and Network Security, vol. 7, no. 12, p. 38, 2007.

[17] P. Manimegalai and K. Thanushkodi, "Content based color image retrieval using adaptive lifting," *International Journal of Advanced Networking and Applications*, vol. 1, no. 6, pp. 359–366, 2010.

[18] D. Doreswamy and K. S. Hemanth, "A novel design specification distance (DSD) based K-mean clustering performance evaluation on engineering materials' database," *International Journal of Computer Applications*, vol. 55, no. 15, pp. 26–33, 2012.

[19] G. F. Tzortzis and A. C. Likas, "The global kernel κ-means algorithm for clustering in feature space," *IEEE Transactions on Neural Networks*, vol. 20, no. 7, pp. 1181–1194, 2009.

[20] A. Vadivel, A. K. Majumdar, and S. Sural, "Performance comparison of distance metrics in content-based image retrieval applications," in *Proceedings of the International Conference on Information Technologies*, pp. 159–164, Bhubaneswar, India, 2003.

[21] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Information Retrieval*, vol. 11, no. 2, pp. 77–107, 2008.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 708075, 8 pages
http://dx.doi.org/10.1155/2014/708075

Research Article

MRI and PET Image Fusion Using Fuzzy Logic and Image Local Features

Umer Javed,[1,2] Muhammad Mohsin Riaz,[3] Abdul Ghafoor,[3] Syed Sohaib Ali,[3] and Tanveer Ahmed Cheema[2]

[1] *Faculty of Engineering and Technology, International Islamic University, Islamabad 44000, Pakistan*
[2] *School of Engineering and Applied Sciences, Isra University, Islamabad 44000, Pakistan*
[3] *College of Signals, National University of Sciences and Technology, Islamabad 44000, Pakistan*

Correspondence should be addressed to Abdul Ghafoor; abdulghafoor-mcs@nust.edu.pk

Received 7 August 2013; Accepted 29 October 2013; Published 19 January 2014

Academic Editors: S. Bourennane and J. Marot

An image fusion technique for magnetic resonance imaging (MRI) and positron emission tomography (PET) using local features and fuzzy logic is presented. The aim of proposed technique is to maximally combine useful information present in MRI and PET images. Image local features are extracted and combined with fuzzy logic to compute weights for each pixel. Simulation results show that the proposed scheme produces significantly better results compared to state-of-art schemes.

1. Introduction

Fusion of images obtained from different imaging systems like computed tomography (CT), MRI, and PET plays an important role in medical diagnosis and other clinical applications. Each imaging technique provides a different level of information. For instance, CT (based on X-ray principle) is commonly used for visualizing dense structures and is not suitable for soft tissues and physiological analysis. MRI on the other hand provides better visualization of soft tissues and is commonly used for detection of tumors and other tissue abnormalities. Likewise, information of blood flow in the body is provided by PET (a nuclear imaging technique) but it suffers from low resolution as compared to CT and MRI. Hence, fusion of images obtained from different modalities is desirable to extract sufficient information for clinical diagnosis and treatment.

Image fusion integrates (complementary as well as redundant) information from multimodality images to create a fused image [1–6]. It not only provides accurate description of the same object but also helps in required memory reduction by storing fused images instead of multiple source images. Different techniques are developed for medical image fusion which can be generally grouped into pixel, feature, and decision level fusion [7]. Compared to feature and decision, pixel level methods [1, 2] are more suited for medical imaging as they can preserve spatial details in fused images [1, 8].

Conventional pixel level methods (including addition, subtraction, multiplication, and weighted average) are simpler but are less accurate. Intensity Hue saturation (IHS)-based methods fuse the images by replacing the intensity component [1, 5, 9]. These methods generally produce high-resolution fused images but cause spectral distortion (due to inaccurate estimation of spectral information) [10]. Similarly, principal components analysis based methods fuse images by replacing certain principle components [11].

Multiresolution techniques including pyramids, discrete wavelet transform (DWT), contourlet, curvelet, shearlet, and framelet transform image into different bands for fusion (a comprehensive comparison is presented in [12]). DWT-based schemes decompose the input images into horizontal, vertical, and diagonal subbands which are then fused using additive or substitutive methods. Earlier DWT-based fusion schemes cannot preserve the salient features of the source images efficiently, hence producing block artifacts and inconsistency in the fused results [2, 3]. Human visual system is combined with DWT to fuse the low frequency bands using visibility and variance features, respectively. Local window

approach is used (to adjust coefficients adaptively) for noise reduction and maintaining homogeneity in fused image [4]. However, the method often produces block artifacts and reduced contrast [3, 5]. Consistency verification and activity measures combined with DWT can only capture limited directional information and hence are not suitable for sharp image transitions [13].

Texture features and visibility measure are used with framelet transform [5] to fuse high and low frequency components, respectively. Contourlet transform based methods use different and flexible directions to detect the intrinsic geometrical structures [13]. The common methods are variable weight using nonsubsampled contourlet transform [14]; and bio-inspired activity measurer using pulse-coded neural networks [15]. However, the down- and up-sampling in contourlet transform lack shift invariance and cause ringing artifacts [14]. Curvelet transform uses various directions and positions at length scales [16]; however, it does not provide a multiresolution representation of geometry [17]. Shearlet transform carries different features (like directionality, localization, and multiscale framework) and can decompose the image into any scale and direction to fuse the required information [17].

Prespecified transform matrix and learning techniques are used with kernel singular value decomposition to fuse images in sparse domain [18]. In [19], image fusion has been performed using redundancy DWT and contourlet transform. A pixel level neuro-fuzzy logic based fusion adjusts the membership functions (MFs) using backpropagation and least mean square algorithms [20]. A spiking cortical model is proposed to fuse different types of medical images [21]. However, these schemes are complex or work under certain assumptions/constraints.

A fusion technique for MRI and PET images using local features and fuzzy logic is presented. The proposed technique maximally combines the useful information present in MRI and PET images. Image local features are extracted and combined with fuzzy logic to compute weights for each pixel. Simulation results based on visual and quantitative analysis show the significance of the proposed scheme.

2. *À*-Trous-Based Image Fusion: An Overview

In contrast to conventional multiresolution schemes (where the output is downsampled after each level), *à*-trous or undecimated wavelet provides shift invariance, hence better suited for image fusion.

Let different approximations $I_{\mathrm{MRI},k}$ of MRI image I_{MRI} (having dimensions $M \times N$) be obtained by successive convolutions with a filter f, that is,

$$I_{\mathrm{MRI},k+1} = I_{\mathrm{MRI},k} * f, \tag{1}$$

where $I_{\mathrm{MRI},0} = I_{\mathrm{MRI}}$ and f is a bicubic B-spline filter. The kth wavelet plane $W_{\mathrm{MRI},k}$ of I_{MRI} is,

$$W_{\mathrm{MRI},k} = I_{\mathrm{MRI},k+1} - I_{\mathrm{MRI},k}. \tag{2}$$

The image I_{MRI} is decomposed into low $I_{\mathrm{MRI},L}$ and high $I_{\mathrm{MRI},H}$ frequency components as

$$\begin{aligned} I_{\mathrm{MRI}} &= I_{\mathrm{MRI},L} + I_{\mathrm{MRI},H} \\ &= I_{\mathrm{MRI},L} + \sum_{k=0}^{K} W_{\mathrm{MRI},k}, \end{aligned} \tag{3}$$

where K is the total number of decomposition levels. Similarly PET image I_{PET} in terms of low $I_{\mathrm{PET},L}$ and high $I_{\mathrm{PET},H}$ frequency components is

$$I_{\mathrm{PET}}(\beta) = I_{\mathrm{PET},L}(\beta) + \sum_{k=0}^{K} W_{\mathrm{PET},k}(\beta), \tag{4}$$

where $\beta \in \{R, G, B\}$, as PET images are assumed to be in pseudocolor [9].

Different methods are present in literature to fuse low and high frequency components which are generally grouped into substitute wavelet (SW) and additive wavelet (AW). The fused image I_{SW} using SW is

$$I_{\mathrm{SW}}(\beta) = I_{\mathrm{PET},L}(\beta) + \sum_{k=0}^{K} W_{\mathrm{MRI},k}. \tag{5}$$

Note that SW method fuses image by completely replacing the high frequency components of PET by high frequency components of MRI image, which can cause geometric and spectral distortion. SW and IHS (SWI) are combined to overcome the limitation in fused image I_{SWI}; that is,

$$I_{\mathrm{SWI}}(\beta) = I_{\mathrm{PET},L}(\beta) - \sum_{k=0}^{K} W_{\mathrm{INT},k} + \sum_{k=0}^{K} W_{\mathrm{MRI},k}, \tag{6}$$

where the intensity image I_{INT} is

$$I_{\mathrm{INT}} = \frac{1}{B} \sum_{\beta'} I_{\mathrm{PET}}(\beta'). \tag{7}$$

The substitution process in SWI method sometimes results in loss of information as the intensity component is obtained by simple averaging/weighting.

In AW method, the fused image I_{AW} is obtained by injecting high frequency components of I_{MRI} into I_{PET}:

$$I_{\mathrm{AW}}(\beta) = I_{\mathrm{PET}}(\beta) + \sum_{k=0}^{K} W_{\mathrm{MRI},k}. \tag{8}$$

AW method adds the same amount of high frequencies into low-resolution bands which causes redundancy of high frequency components (hence resulting in spectral distortion).

To cater the limitation, AW luminance proportional (AWLP) method injects the high frequencies in proportion to the intensity values [22]. Consider

$$I_{\mathrm{AWLP}}(\beta) = I_{\mathrm{PET}}(\beta) + \frac{I_{\mathrm{PET}}(\beta)}{(1/B)\sum_{\beta'} I_{\mathrm{PET}}(\beta')} \sum_{k=1}^{K} W_{\mathrm{MRI},k}, \tag{9}$$

where B are total number of bands. The fused image I_{AWLP} of AWLP preserves the relative spectral information amongst different bands. The fused image using improved additive wavelet proportional (IAWP) [23] method is

$$I_{\mathrm{IAWP}}(\beta) = I_{\mathrm{PET}}(\beta) + \frac{I_{\mathrm{PET}}(\beta)}{(1/B)\sum_{\beta'} I_{\mathrm{PET}}(\beta')}, \times \left[\sum_{k=1}^{K} W_{\mathrm{MRI},k} - \sum_{k=1}^{K} W_{\mathrm{MRIR},k}\right], \tag{10}$$

where $W_{\mathrm{MRIR},k}$ are wavelet planes of a low-resolution (a spatially degraded version of I_{MRI}) MRI image I_{MRIR}. The I_{MRIR} is obtained by filtering the high frequencies (by applying a smoothing filter). The major limitations of the above schemes includes induction of redundant high/low frequencies; and consequently spatial degradations.

3. Proposed Technique

The proposed scheme first decomposes the MRI and PET images into low and high frequencies using *à*-trous wavelet. High and low frequencies are then fused separately according to defined criterion. The overall fused image I_F in terms of high $I_{F,H}$ and low $I_{F,L}(\beta)$ frequencies is

$$I_F(\beta) = I_{F,L}(\beta) + I_{F,H}. \tag{11}$$

3.1. Fusion of Low Frequencies. Fusion of low frequencies $I_{\mathrm{MRI},L}$ and $I_{\mathrm{PET},L}$ is critical and challenging task. Various schemes utilize different criterions for fusion of low frequencies. For instance, one choice is to totally discard the low frequencies of one image, another choice is to take average or weighted average of both and so forth. However, the schemes provide limited performance as they do not cater the spatial properties of image. We have proposed fusion of low frequency using different weighting average for each pixel location. The weights are computed based on the amount of information contained in vicinity of each pixel.

3.1.1. Local Features. Local variance (LV) and local blur (LB) features are used with fuzzy inference engine to compute the desired weights for fusing low frequencies.

LV [24] is used to evaluate the regional characteristics of $I_{\mathrm{PET},L}$ image and is defined as I_{LV}:

$$I_{\mathrm{LV}}(\beta,m,n) = \frac{1}{(2m_1+1)(2n_1+1)} \times \sum_{m_2=m-m_1}^{m+m_1} \sum_{n_2=n-n_1}^{n+n_1} \left(I_{\mathrm{PET}}(\beta,m_2,n_2) - \bar{I}_{\mathrm{PET}}(\beta)\right)^2, \tag{12}$$

where $\bar{I}_{\mathrm{PET}}(\beta)$ is the mean value of $m_1 \times n_1$ window centered at (m,n) pixel. Note that image containing sharp edges results in higher value (and vice versa).

LB I_{LB} is computed using local Rényi entropy [25] of $I_{\mathrm{PET},L}$ image. Let $P_{\beta mn}(k)$ be the probability (or normalized histogram) having intensity values $k = 1, 2, \ldots, K$ within a local window (of size $m_1 \times n_1$) centered at (β,m,n) pixel. I_{LB} is defined as [25]

$$I_{\mathrm{LB}}(\beta,m,n) = -\frac{1}{2}\ln\left(\sum_{k}^{K} P_{\beta mn}^3(k)\right). \tag{13}$$

High values of I_{LV} and I_{LB} show that $I_{\mathrm{PET},L}$ contain more information and need to be assigned more weight as compared to $I_{\mathrm{MRI},L}$ image.

3.1.2. Fuzzy Inference Engine. Let high $\zeta_{\mathrm{LV},1}(u)$ and low $\zeta_{\mathrm{LV},2}(u)$ Gaussian Membership functions (MFs) having means $\overline{u}^{(1)}$, $\overline{u}^{(2)}$ and variances $\sigma_u^{(1)}$, $\sigma_u^{(2)}$ for LV be [26]

$$\zeta_{\mathrm{LV},1}(u) = e^{-((u-\overline{u}^{(1)})/\sigma_u^{(1)})^2}, \qquad \zeta_{\mathrm{LV},2}(u) = e^{-((u-\overline{u}^{(2)})/\sigma_u^{(2)})^2}. \tag{14}$$

Similarly let high $\zeta_{\mathrm{LB},1}(v)$ and low $\mu_{\mathrm{LB},2}(v)$ Gaussian MFs having means $\overline{v}^{(1)}$, $\overline{v}^{(2)}$ and variances $\sigma_v^{(1)}$, $\sigma_v^{(2)}$ for LB be

$$\zeta_{\mathrm{LB},L}(v) = e^{-((v-\overline{v}^{L})/\sigma_v^{L})^2}, \qquad \zeta_{\mathrm{LB},H}(v) = e^{-((v-\overline{v}^{H})/\sigma_v^{H})^2}. \tag{15}$$

The inputs $I_{\mathrm{LV}}(\beta,m,n)$ and $I_{\mathrm{LV}}(\beta,m,n)$ are mapped into fuzzy set using Gaussian fuzzifier [27] as

$$\zeta_{\mathrm{LV,LB}}(u,v) = e^{-((u-I_{\mathrm{LV}}(\beta,m,n))/\varsigma_1)^2} \times e^{-((v-I_{\mathrm{LB}}(\beta,m,n))/\varsigma_2)^2}, \tag{16}$$

where ς_1 and ς_2 are noise suppression parameters. The inputs are then processed by fuzzy inference engine using pre defined IF-THEN rules [26, 27] as follows.

$Ru^{(1)}$: IF $I_{\mathrm{LV}}(\beta,m,n)$ is high and $I_{\mathrm{LB}}(\beta,m,n)$ is high THEN $I_{\mathrm{WT}}(\beta,m,n)$ is high.

$Ru^{(2)}$: IF $I_{\mathrm{LV}}(\beta,m,n)$ is low and $I_{\mathrm{LB}}(\beta,m,n)$ is high THEN $I_{\mathrm{WT}}(\beta,m,n)$ is medium.

$Ru^{(3)}$: IF $I_{\mathrm{LV}}(\beta,m,n)$ is high and $I_{\mathrm{LB}}(\beta,m,n)$ is low,THEN $I_{\mathrm{WT}}(\beta,m,n)$ is medium.

$Ru^{(4)}$: IF $I_{\mathrm{LV}}(\beta,m,n)$ is low and $I_{\mathrm{LB}}(\beta,m,n)$ is low THEN $I_{\mathrm{WT}}(\beta,m,n)$ is low.

The output MFs for high (having mean $\overline{y}^{(1)}$ and variance $\sigma_y^{(1)}$), medium (having mean $\overline{y}^{(2)}$ and variance $\sigma_y^{(2)}$), and low (having mean $\overline{y}^{(3)}$ and variance $\sigma_y^{(3)}$) are defined as

$$\zeta_{W,1}(y) = e^{-((y-\overline{y}^{(1)})/\sigma_y^{(1)})^2}, \qquad \zeta_{W,2}(y) = e^{-((y-\overline{y}^{(2)})/\sigma_y^{(2)})^2}, \qquad \zeta_{W,3}(y) = e^{-((y-\overline{y}^{(3)})/\sigma_y^{(3)})^2}. \tag{17}$$

The output of fuzzy inference engine is

$$\zeta'_{W,L}(y) = \max_{\{c,d,e\}}\left[\sup_{\{u,v\}}\zeta_{\mathrm{LV,LB}}(u,v)\,\zeta_{\mathrm{LV},c}(u)\,\zeta_{\mathrm{LV},d}(v)\,\zeta_{W,e}(y)\right], \tag{18}$$

Table 1: Quantitative measures for fused PET-MRI images.

Scenario	Techniques	Entropy [29]	MI [29]	SSIM [30]	Xydeas and Petrovic [31]	Piella [32]
Normal brain	DWT [12]	5.403	1.6607	0.6083	0.4944	0.7558
	GIHS [6]	5.381	1.7017	0.7095	0.5362	0.8014
	GFF [33]	5.115	1.7479	0.6803	0.4825	0.6741
	IAWP [23]	5.152	1.7753	0.6735	0.3233	0.3331
	Proposed	5.738	1.7912	0.6788	0.5746	0.8469
Grade II astrocytoma	DWT [12]	3.4820	1.3817	0.7287	0.6495	0.8566
	GIHS [6]	3.4679	1.3848	0.8149	0.6227	0.8779
	GFF [33]	3.5558	1.3758	0.8120	0.6417	0.8561
	IAWP [23]	3.6351	1.3770	0.8018	0.3757	0.5405
	Proposed	3.5762	1.4292	0.8133	0.6674	0.9125
Grade IV astrocytoma	DWT [12]	5.4140	1.7487	0.6775	0.5727	0.8434
	GIHS [6]	5.7868	1.7084	0.6207	0.5697	0.8547
	GFF [33]	5.6628	1.7883	0.6819	0.5112	0.7917
	IAWP [23]	5.6831	1.8298	0.6718	0.3584	0.5642
	Proposed	5.8204	1.8683	0.6739	0.5885	0.8755

where $\{c, d\} \in \{1, 2\}$ and $e \in \{1, 2, 3\}$. The weights $I_{\mathrm{WT}}(\beta, m, n)$ are obtained by processing fuzzy outputs using center average defuzzifier [27].

The $I_{F,L}(\beta)$ image is obtained by weighted sum of $I_{\mathrm{PET},L}$ and $I_{\mathrm{MRI},L}$ as

$$I_{F,L}(\beta, m, n) = I_{\mathrm{WT}}(\beta, m, n) I_{\mathrm{PET},L}(\beta, m, n) + (1 - I_{\mathrm{WT}}(\beta, m, n)) I_{\mathrm{MRI},L}(m, n). \quad (19)$$

3.2. Fusion of High Frequencies. Let $W_{\mathrm{MRI\text{-}MRIR},k}$ represent a wavelet plane of the resultant image $I_{\mathrm{MRI}} - I_{\mathrm{MRIR}}$. This ensures that only those high frequency components are used for image fusion, which are not already present in I_{MRI}. By the virtue of this, the proposed scheme not only avoids redundancy of information but also results in improved fusion results as compared to early techniques. The fused high frequency image $I_{F,H}$ is

$$I_{F,H} = \sum_{k=1}^{K} W_{\mathrm{MRI\text{-}MRIR},k}. \quad (20)$$

Note that $I_{F,H}$ is not dependent on the bands β because I_{MRI} is gray-scale image.

4. Results and Discussion

The simulations of proposed and existing schemes are performed on PET and MRI images obtained from Harvard database [28]. The fusion database for brain images is classified into normal, grade II astrocytoma, and grade IV astrocytoma images. The MRI and PET images are coregistered with 256×256 spatial resolution. The proposed fusion scheme is compared visually and quantitatively (using entropy [29], mutual information (MI) [29], structural similarity (SSIM) [30], Xydeas and Petrovic [31] metric, and Piella [32] metric) with DWT [12], GIHS [6], IAWP [23], and GFF [33] schemes.

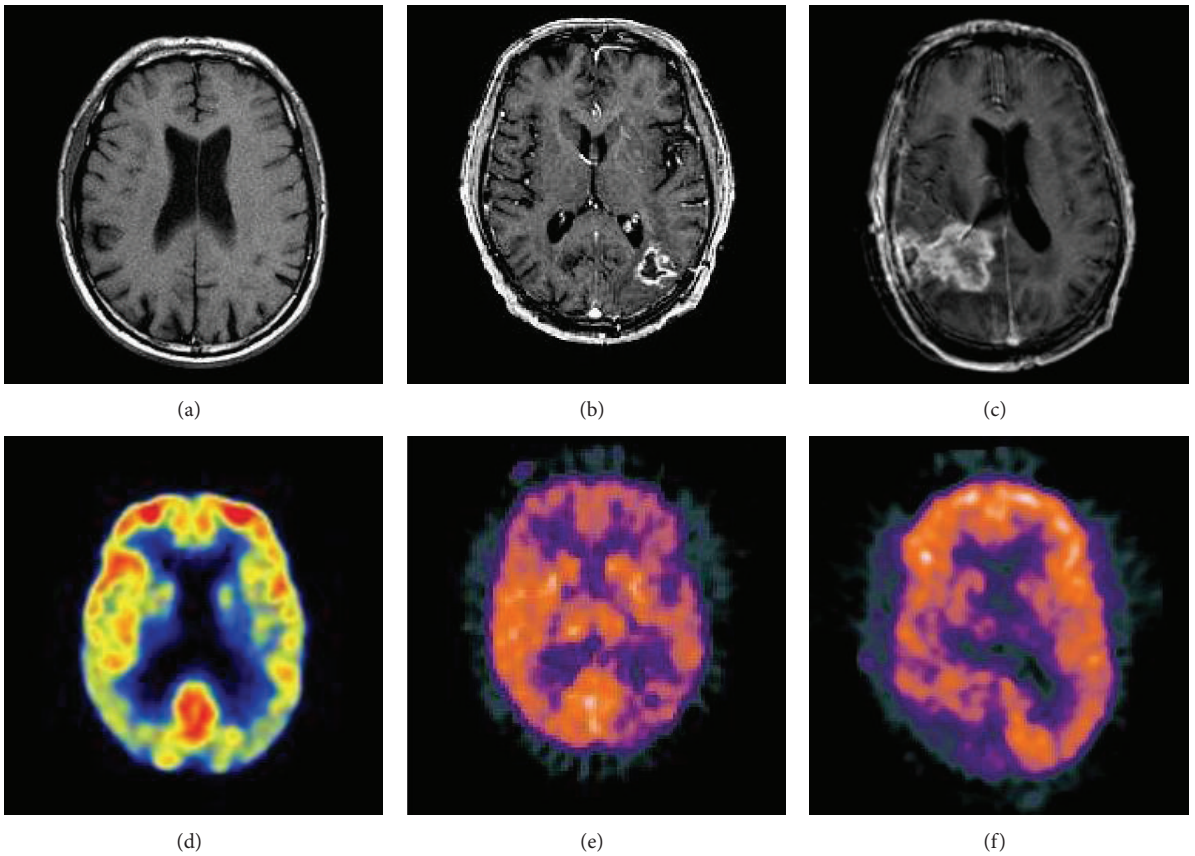
The original MRI images belonging to normal brain, grade II astrocytoma, and grade IV astrocytoma are shown in Figures 1(a)–1(c), respectively. Fluorodeoxyglucose (FDG) is a radiopharmaceutical commonly used for PET scans. The PET-FDG images of normal, grade II, and grade IV astrocytoma are shown in Figures 1(d)–1(f), respectively. It can be seen that different imaging modalities provide complementary information for the same region.

Figure 2 shows fused images (of normal brain) obtained by using different techniques. It can be seen from Figure 2(e) that the proposed technique has preserved the complementary information of both modalities and the fuzzy based weight assessment has enabled offering less spectral information loss as compared to other state-of-art techniques.

Figure 3 shows fused images (of grade II astrocytoma class) obtained by using different techniques. From Figure 3(e), it can be observed that the proposed technique provides complementary information contained in both modalities and the fuzzy based weight assessment has enabled offering less spectral information loss as compared to other state of art techniques. The improvement in fused images is more visible in the tumorous region (bottom right corner).

Figure 4 shows fused images (of Grade IV astrocytoma) obtained by using different techniques. Similar improvement (as that of Figures 2(e) and 3(e)) can be observed in Figure 4(e). It is easy to conclude that the proposed scheme provides better visual quality compared to the existing schemes.

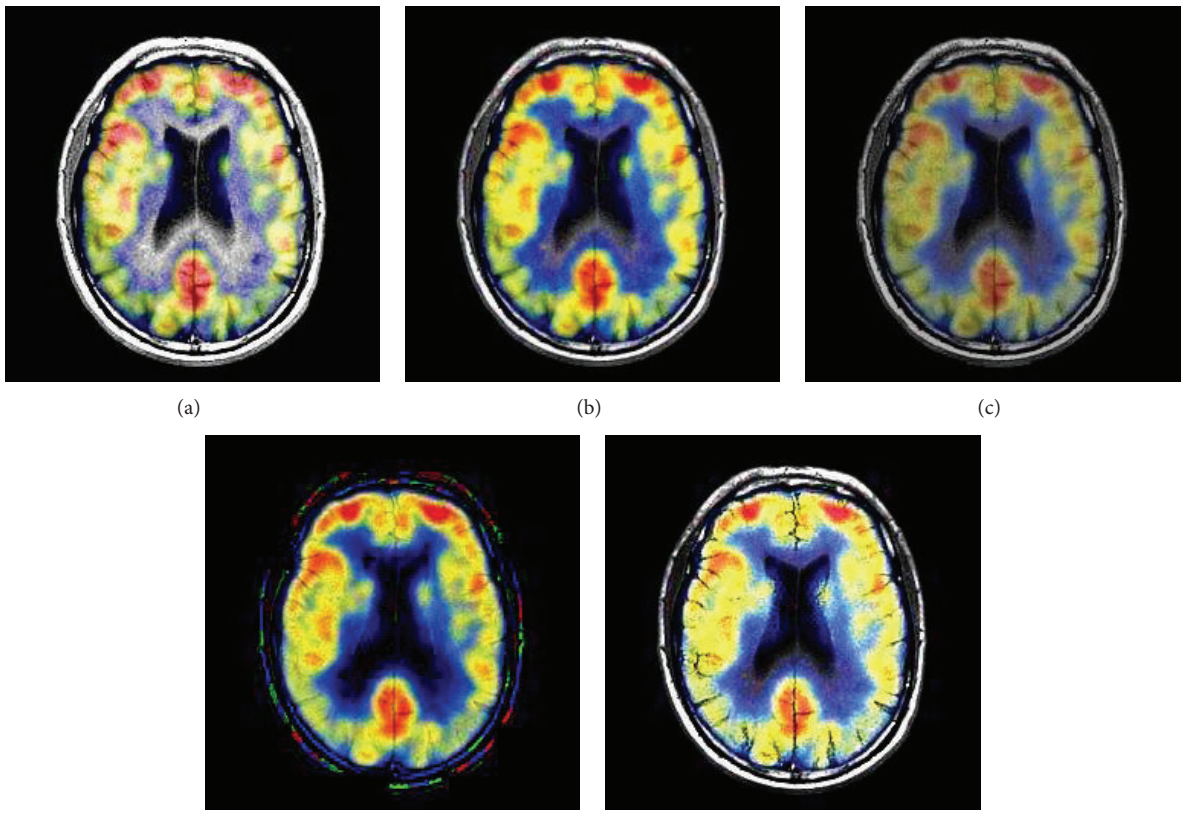
Table 1 shows the quantitative comparison of different fusion techniques. Note that a higher value of the metric represents better quality. The fused images obtained using proposed technique provide better quantitative results in



(a) (b) (c)

(d) (e) (f)

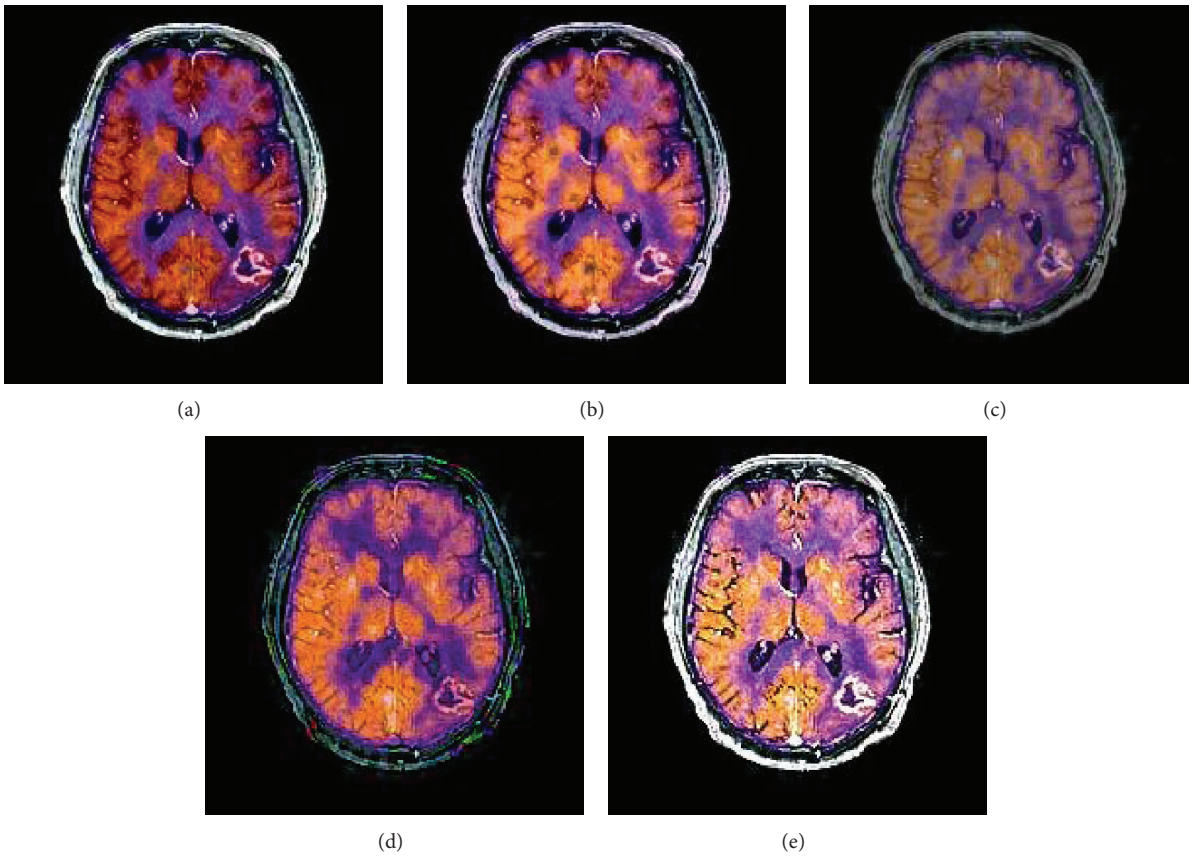
Figure 1: Original MRI and PET images: (a)–(c) MRI; (d)–(f) PET.



(a) (b) (c)

(d) (e)

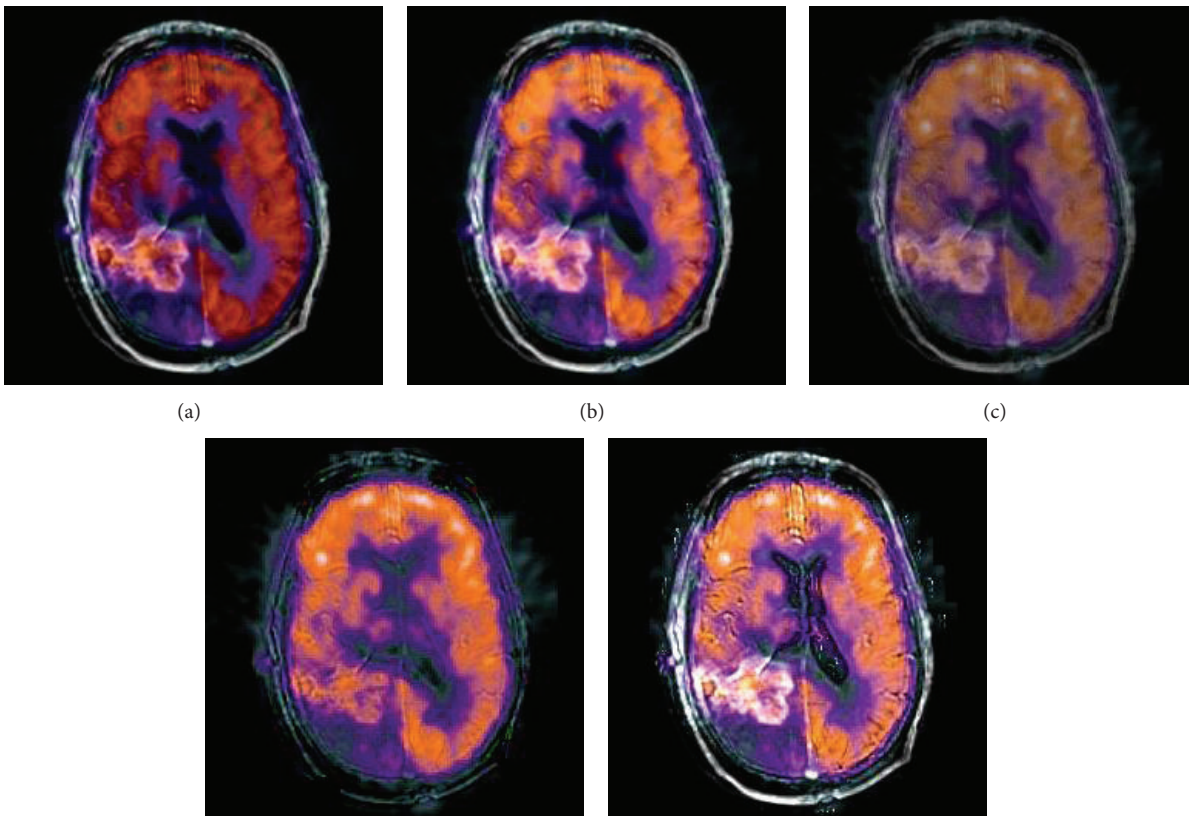
Figure 2: Image fusion results for normal images: (a) DWT [12]; (b) GIHS [6]; (c) GFF [33]; (d) IAWP [23]; (e) proposed technique.



(a) (b) (c)

(d) (e)

Figure 3: Image fusion results for grade II astrocytoma images: (a) DWT [12]; (b) GIHS [6]; (c) GFF [33]; (d) IAWP [23]; (e) proposed technique.



(a) (b) (c)

(d) (e)

Figure 4: Image fusion results for grade IV astrocytoma images: (a) DWT [12]; (b) GIHS [6]; (c) GFF [33]; (d) IAWP [23]; (e) proposed technique.

terms of entropy [29], MI [29], SSIM [30], Xydeas and Petrovic [31], and Piella [32] metrics.

5. Conclusion

An image fusion technique for MRI and PET using local features and fuzzy logic is presented. The proposed scheme maximally combines the useful information present in MRI and PET images using image local features and fuzzy logic. Weights are assigned to different pixels for fusing low frequencies. Simulation results based on visual and quantitative analysis show that the proposed scheme produces significantly better results compared to state of art schemes.

6. Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "Human visual system inspired multi-modal medical image fusion framework," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1708–1720, 2013.

[2] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.

[3] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques—an introduction, review and comparison," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 4, pp. 249–263, 2007.

[4] Y. Yang, D. S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 579341, 13 pages, 2010.

[5] G. Bhatnagar and Q. M. J. Wu, "An image fusion framework based on human visual system in framelet domain," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 10, no. 1, Article ID 1250002, 2012.

[6] T. Li and Y. Wang, "Biological image fusion using a NSCT based variable-weight method," *Information Fusion*, vol. 12, no. 2, pp. 85–92, 2011.

[7] S. T. Shivappa, B. D. Rao, and M. M. Trivedi, "An iterative decoding algorithm for fusion of multimodal information," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, Article ID 478396, 10 pages, 2008.

[8] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Information Fusion*, vol. 13, no. 1, pp. 10–19, 2012.

[9] S. Daneshvar and H. Ghassemian, "MRI and PET image fusion by combining IHS and retina-inspired models," *Information Fusion*, vol. 11, no. 2, pp. 114–123, 2010.

[10] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1391–1402, 2005.

[11] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.

[12] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.

[13] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[14] D. Li and H. Chongzhao, "Fusion for CT image and MR image based on nonsubsampled transformation," in *Proceedings of the IEEE International Conference on Advanced Computer Control (ICACC '10)*, vol. 5, pp. 372–374, March 2010.

[15] X.-B. Qu, J.-W. Yan, H.-Z. Xiao, and Z.-Q. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1508–1514, 2008.

[16] E. Candès, L. Demanet, D. Donoho, and L. X. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[17] Q.-G. Miao, C. Shi, P.-F. Xu, M. Yang, and Y.-B. Shi, "A novel algorithm of image fusion using shearlets," *Optics Communications*, vol. 284, no. 6, pp. 1540–1547, 2011.

[18] N. N. Yu, T. S. Qiu, and W. H. Liu, "Medical image fusion based on sparse representation with KSVD," in *Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, vol. 39, pp. 550–553, 2013.

[19] S. Rajkumar and S. Kavitha, "Redundancy Discrete Wavelet Transform and Contourlet Transform for multimodality medical image fusion with quantitative analysis," in *Proceedings of the 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET '10)*, pp. 134–139, November 2010.

[20] J. Teng, S. Wang, J. Zhang, and X. Wang, "Neuro-fuzzy logic based fusion algorithm of medical images," in *Proceedings of the 3rd International Congress on Image and Signal Processing (CISP '10)*, vol. 4, pp. 1552–1556, October 2010.

[21] R. Wang, Y. Wu, M. Ding, and X. Zhang, "Medical image fusion based on spiking cortical model," in *Medical Imaging 2013: Digital Pathology*, vol. 8676 of *Proceedings of SPIE*, 2013.

[22] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, 2007.

[23] Y. Kim, C. Lee, D. Han, Y. Kim, and Y. Kim, "Improved additive-wavelet image fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 263–267, 2011.

[24] D.-C. Chang and W.-R. Wu, "Image contrast enhancement based on a histogram transformation of local standard deviation," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 518–531, 1998.

[25] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B42–B51, 2007.

[26] M. M. Riaz and A. Ghafoor, "Fuzzy logic and singular value decomposition based through wall image enhancement," *Radioengineering Journal*, vol. 22, no. 1, p. 580, 2012.

[27] L. X. Wang, *A Course in Fuzzy Systems and Control*, Prentice Hall, New York, NY, USA, 1997.

[28] Harvard Medical Atlas Database, http://www.med.harvard.edu/AANLIB/home.html.

[29] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[31] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[32] G. Piella, "Image fusion for enhanced visualization: a variational approach," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 1–11, 2009.

[33] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 2864–2875, 2013.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 262356, 8 pages
http://dx.doi.org/10.1155/2014/262356

Research Article

Detecting Image Splicing Using Merged Features in Chroma Space

Bo Xu, Guangjie Liu, and Yuewei Dai

School of Automation, Nanjing University of Science & Technology, Nanjing 210094, China

Correspondence should be addressed to Bo Xu; xubo.njust@gmail.com

Received 29 August 2013; Accepted 29 October 2013; Published 16 January 2014

Academic Editors: S. Bourennane and J. Marot

Image splicing is an image editing method to copy a part of an image and paste it onto another image, and it is commonly followed by postprocessing such as local/global blurring, compression, and resizing. To detect this kind of forgery, the image rich models, a feature set successfully used in the steganalysis is evaluated on the splicing image dataset at first, and the dominant submodel is selected as the first kind of feature. The selected feature and the DCT Markov features are used together to detect splicing forgery in the chroma channel, which is convinced effective in splicing detection. The experimental results indicate that the proposed method can detect splicing forgeries with lower error rate compared to the previous literature.

1. Introduction

Image splicing is a forgery manner to copy and paste regions within separate image sources, usually carried out by digital image editing tools such as Photoshop. It is often used as an initial step of photomontage, which is very popular in digital image content editing. The splicing tampered image could be used in news reports, photography contest, key proof in the academic papers, and so on, which could bring certain negative influences. As a result, it is an important issue to develop reliable splicing detection methods.

In the forgery process, the manually introduced transitions of edges and corners are different from those in the natural images. The differences are commonly described by the inconsistency and abnormality, and they are used for splicing detection. Farid [1] discussed how to detect unnatural higher-order correlations introduced into the signal by the tampering process based on bispectral analysis. In the frequency domain, a "natural" signal usually has weak higher-order statistical correlations, and certain "unnatural" correlations will be introduced if this signal is passed through a nonlinearity (which would almost surely occur in tampering). Based on it, Ng et al. [2] proposed a method to detect the abrupt splicing discontinuity using the bicoherence features. They studied the effects of image splicing on magnitude and phase characteristics of bicoherence (the normalized bispectrum). The difference between means of magnitudes of a test image's bi-coherence and its estimating authentic vision's bi-coherence and the difference between negative phase entropy of those two were used as features. Gopi et al. [3, 4] proposed to detect forgeries using an artificial neural network, independent component analysis (ICA), and autoregressive coefficients. In [5, 6], Hsu and Chang proposed a method based on camera response function estimated from geometry invariants. Wang and Zhang [7] employed the Sobel edge detector, derivative operation, and Hough transform to detect image splicing. Lint et al. [8] proposed a method based on computing the inverse camera response functions by analyzing the edges in different patches of the image and verifying their consistency.

Shi et al. proposed to use statistics of 2D phase congruency [9], a natural image model [10], and Hilbert-Huang transform [11] to separate spliced images from authentic ones. They believed that on one hand, steganography and splicing had different goals and strategies causing different statistical artifacts on images, on the other hand, both of them made the touched (stego and spliced) image different from corresponding original ones. Therefore, they built a natural image model using steganalysis features including statistical moments of characteristic functions and Markov

transition probabilities from image 2D array and multisize block discrete cosine transform 2-D array. Wang et al. [12] proposed an image splicing detection method based on GLCM of the edge image in the chroma space. Zhang et al. [13] proposed a splicing detection scheme based on moment features extracted from the discrete cosine transform and image quality features. Dong et al. [14] proposed a method by analyzing the discontinuity of image pixel correlation and coherency caused by splicing, and they also proposed to detect image splicing in chroma space [15], which is claimed to be very efficient. Bayram et al. [16] fused several single tools together to detect tampering. The feature vectors BSM, IQM, and HOW, which were initially proposed for steganalysis, were used in this paper, and SFFS (sequential forward floating search) is used for feature selection.

Existing methods exploit features from various respects to detect splicing forgeries. Recently, Fridrich and Kodovský [17] proposed to build plenty of submodels as rich model to detect image steganography, which achieved good results. The SRM exploits the spatial correlations of neighborhood pixels from the image residuals which are hard to be preserved during image steganography. It is believed that the case would be the same for image splicing. In this paper, the image splicing detection method based on SRM is studied. AT first, the performance of the submodels is evaluated and analyzed by the detailed experimental results in the luminance channel. Moreover, one selected submodel with feature from DCT is used to detect splicing forgery in chroma channel.

The rest of the paper is organized as listed in following. The SRM is described in Section 2. Section 3 discusses the detection method. And experiments are given in Section 4. In Section 5, a conclusion is drawn to the whole paper.

2. Feature Discriptions

In [17], Fridrich and Kodovský proposed a general methodology for steganalysis of digital images based on the concept of the SRM consisting of a large number of diverse submodels. The submodels consider various types of relationships among neighboring samples of noise residuals obtained by linear and nonlinear filters with compact supports.

2.1. SRM. The SRM captures a large number of different types of dependencies among neighboring pixels. By employing the submodels, better results can be achieved beyond enlarging a single model, which is unlikely to produce good results as the enlarged model will have too many underpopulated bins. By forming a model by merging many smaller submodels, this problem can be avoided.

The main formation processes of SRM are residuals, truncation, and quantization and cooccurrence matrix. The description of all residuals is given here, for other detail description, please refer to [17]. For example, in the first order residual $R_{ij} = X_{i,j+1} - X_{i,j}$, the central pixel $X_{i,j}$ is predicted as its immediate neighbor, $\widehat{X}_{ij} = X_{i,j+1}$, while the predictor in the second order residual $R_{ij} = X_{i,j-1} + X_{i,j+1} - 2X_{i,j}$ assumes that the image is locally linear in the horizontal direction, $2\widehat{X}_{i,j} = (X_{i,j-1} + X_{i,j+1})$.

2.1.1. Residual Classes. The residuals are divided into six classes depending on the central pixel predictor they are built from. The classes will be mentioned by the following descriptive names: 1st, 2nd, 3rd, SQUARE, EDGE3 × 3, and EDGE5 × 5. The predictors in class "1st" estimate the pixel as the value of its neighbor, while those from class "2nd" ("3rd") incorporate a locally linear (quadratic) model. The class "SQUARE" makes use of more pixels for the prediction. The class "EDGE3 × 3" predictors, derived from the 3 × 3 square kernel S3a, were included to provide better estimates at spatial discontinuities (edges).

The larger 5 × 5 predictor in S5a was obtained as a result of optimizing the coefficients of a circularly-symmetrical 5 × 5 kernel. The "EDGE5 × 5" residuals E5a–E5d are built from S5a in an analogical manner as E3a–E3d are built from S3a.

2.1.2. Residual Symmetries. Each residual exhibits symmetries that will later allow us to reduce the number of submodels and make them better populated. If the residual does not change after being computed from the image rotated by 90 degrees, it is nondirectional; otherwise, it is directional. For instance, 1a, 1b, 2a, 2e, and E3c are directional while 1e, 2b, 2c, S3a, and E3d are nondirectional. Two cooccurrence matrices are computed for each residual, one for the horizontal scan and the other for the vertical scan. We call a residual *hv*-symmetrical if its horizontal and vertical cooccurrences can be added to form a single matrix (submodel) based on the argument that the statistics of natural images do not change after rotating the image by 90 degrees.

2.1.3. Syntax. The syntax of the names follows this convention:

$$\text{name} = \{\text{type}\}\,\{f\}\,\{\sigma\}\,\{\text{scan}\}, \quad (1)$$

where type ∈ {spam, minmax}, f is the number of filters, σ is the symmetry index, and the last symbol scan ∈ $\{\emptyset, h, v\}$ may be missing (for *hv*-symmetrical residuals) or it is either h or v, depending on the cooccurrence scan that should be used with the residual. For more detailed descriptions and definitions of the SRM, please refer to [17].

2.2. Markov Matrices in DCT. Splicing operation changes the local relationships in pixel neighborhood, and Markov process could be used to model the changes, which is commonly used in image processing. In this paper, the Markov feature is extracted in DCT of Chrominance space. As shown in Figure 1, the test image is transformed using 8 × 8 block DCT at first, and after preprocessing to the coefficient matrix, the 2D differential matrices and probability transition

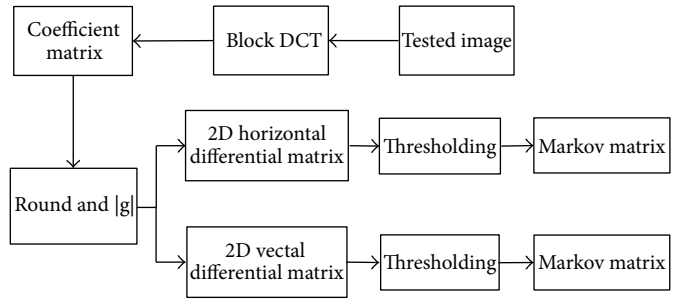


Figure 1: Feature extraction in DCT of Chrominance.

Table 1: Detection results of the first order submodels.

Model	P_{FA}	P_{MD}	P_E (total error)
s1_minmax22h	1	0	0.5
s1_minmax22v	0.1401	0.1457	0.1429
s1_minmax24	0.1549	0.2219	0.1884
s1_minmax34	0.2097	0.2175	0.2136
s1_minmax34h	1	0	0.5
s1_minmax34v	0.1334	0.1799	0.1567
s1_minmax41	1	0	0.5
s1_minmax48h	0.1556	0.1996	0.1776
s1_minmax48v	0.1488	0.1915	0.1702
s1_minmax54	0.1286	0.1786	0.1536
s1_spam14hv	0.1661	0.1876	0.1768

Table 2: Detection results of the second order submodels.

Model	P_{FA}	P_{MD}	P_E (total error)
s2_minmax21	**0.1233**	**0.1247**	**0.124**
s2_minmax24h	0.1276	0.181	0.1543
s2_minmax24v	**0.1178**	**0.138**	**0.1279**
s2_minmax32	0.13	0.1591	0.1446
s2_minmax41	***0.1122***	***0.1214***	***0.1168***
s2_spam12hv	0.1193	0.1565	0.1379

Table 3: Detection results of the third order submodels.

Models	P_{FA}	P_{MD}	P_E (total error)
s3_minmax22h	1	0	0.5
s3_minmax22v	0.1439	0.1353	0.1396
s3_minmax24	0.1551	0.1422	0.1487
s3_minmax34	0.1637	0.16	0.1618
s3_minmax34h	1	0	0.5
s3_minmax34v	0.1639	0.152	0.1579
s3_minmax41	1	0	0.5
s3_minmax48h	0.1772	0.1564	0.1668
s3_minmax48v	0.1694	0.1684	0.1689
s3_minmax54	0.1806	0.1697	0.1751
s3_spam14hv	***0.1328***	***0.1325***	***0.1327***

matrices are obtained using the following (Equations (2)–(3)):

$$\begin{aligned}
\mathrm{CM}_v(i,j) &= \left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\left[\delta\left(F(m,n)=i, F(m,n+1)=j\right)\right]\right) \\
&\quad\times\left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\delta\left(F(m,n)=i\right)\right)^{-1}, \\
\mathrm{CM}_v(i,j) &= \left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\left[\delta\left(F(m,n)=i, F(m+1,n)=j\right)\right]\right) \\
&\quad\times\left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\delta\left(F(m,n)=i\right)\right)^{-1}, \\
\mathrm{CM}_d(i,j) &= \left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\left[\delta\left(F(m,n)=i,\right.\right.\right. \\
&\qquad\left.\left.\left. F(m+1,n+1)=j\right)\right]\right) \\
&\quad\times\left(\sum_{m=1}^{M-2N-2}\sum_{n=1}^{}\delta\left(F(m,n)=i\right)\right)^{-1},
\end{aligned} \tag{2}$$

where $i, j \in [0, 8]$, and

$$\delta(x=i, y=j) = \begin{cases} 1, & \text{if } x=i \text{ and } y=j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

If the added splicing image block is regarded as the additive noise to the original image, the calculated probability transition matrices would reveal the clue of splicing.

The dimension of the feature vector is adjustable by setting threshold T. In this paper, T is set to 3, and the dimension of DCT feature is 98.

3. Detection Method

The SVM is employed in the detection procedure. The LibSVM is introduced and profiled at first, followed by the training and testing steps.

3.1. LIBSVM. LIBSVM is a popular machine learning tool [16]. In this paper, the RBF kernel of LibSVM is employed, and C and gamma are automatically selected using fivefold cross-validation.

As different feature sets will be merged together in the detection algorithm, feature normalization is a necessary step before training and testing. The specific scaling method is very simple. Firstly, the mean and standard deviation for each feature set are calculated. Secondly, by subtracting the mean in each feature and dividing the values of each feature by its standard deviation, the normalized features are obtained. In favor of making comparison, all the training and testing processing are done after the feature normalization.

3.2. Training Steps. The detection method consists of the following steps.

(1) Read an image from training image set and label it, 0 for authenticate image and 1 for spliced image. Covert it into 8 bit gray scale image if necessary.

(2) Let $q = 1$, and compute the feature of submodel $s\{i\}$ as described in Section 2.

(3) Repeat 1-2 until all the images from the training image set have been processed.

(4) Train the models for individual submodel or merged ones using LIBSVM as described in the previous parts.

3.3. Detection Steps

(1) Read an image from the testing image set. Covert it into 8 bit gray scale image if necessary.

(2) Let $q = 1$, and compute its feature by the individual submodel or the merged ones as the training steps.

(3) Test the feature using the corresponding trained SVM.

(4) Repeat steps (1)–(3) until all images have been tested.

4. Experiments

According to the detection method discussed in Section 3, the experiments are performed. At first, the general detection performance of the individual submodel is tested on both image datasets in luminance channel. And then, one submodel is chosen for the chroma channel experiment and comparison to the state of the art result is given.

4.1. Experiment Setup

4.1.1. Image Dataset. In the experiment, two splicing image datasets are used. The Columbia uncompressed image splicing detection evaluation dataset [5] is commonly used as a benchmark. The database contains totally 183 authenticate images and 180 spliced images. The image sizes range from 757×568 to 1152×768 and are all uncompressed. The image detection evaluation database (CASIA TIDE) v2 consists of 7,491 authentic and 5,123 sophisticatedly tampered color images of different sizes varying from 240×160 to 900×600. This database is with larger size and more realistic and challenging tampered images with complex splicing as well as blurring. The authentic images are collected from the Corel image dataset, websites, and so on. The tampered images are generated with resizing, rotation, or other distortion, and postprocessing (such as blurring) is performed after cut-and-past operation, which make the dataset more realistic and challenging.

Table 4: Detection results of the EDGE3 × 3 submodels.

Model	P_{FA}	P_{MD}	P_E (total error)
s3 × 3_minmax22h	0.1366	0.1962	0.1664
s3 × 3_minmax22v	0.11	0.1802	0.1451
s3 × 3_minmax24	***0.1302***	***0.1213***	***0.1258***
s3 × 3_minmax41	0.1591	0.1735	0.1663
s3 × 3_spam14hv	**0.1168**	**0.1323**	**0.1245**

Table 5: Detection results of the EDGE5 × 5 submodels and the square submodel.

Model	P_{FA}	P_{MD}	P_E (total error)
s5 × 5_minmax22h	0.1656	0.1622	0.1639
s5 × 5_minmax22v	**0.1146**	**0.1367**	**0.1256**
s5 × 5_minmax24	0.1639	0.1291	0.1465
s5 × 5_minmax41	0.1882	0.2073	0.1977
s5 × 5_spam14hv	***0.1131***	***0.114***	***0.1135***
s35_spam11	0.1719	0.1424	0.1572

Table 6: Detection results of the merger submodels.

Model	P_{FA}	P_{MD}	P_E (total error)
2-best-submodels	0.0959	0.1016	0.0987
4-best-submodels	0.1121	0.1078	0.1100
8-best-submodels	0.1173	0.1162	0.1168

Table 7: Detection results on the CASIA V2 dataset.

Model	P_{FA} %	P_{MD} %	P_E %
YSRM	16.97	9.11	13.04
CrSRM	0.43	0.19	0.31
CbSRM	0.38	0.19	0.28
CrDCT	0.57	0.15	0.36
Cb + DCT	0.16	0.18	0.17
Cr + DCT	0.23	0.19	0.21
[15]	6.70	2.10	4.40

4.1.2. Experiment Settings. The detection results of the experiment are evaluated by the detection error with the form as

$$P_E = \frac{1}{2}\left(P_{FA} + P_{MD}\right), \tag{4}$$

where P_{FA} and P_{MD} are the probabilities of false alarm and missed detection within the model whose parameters C and

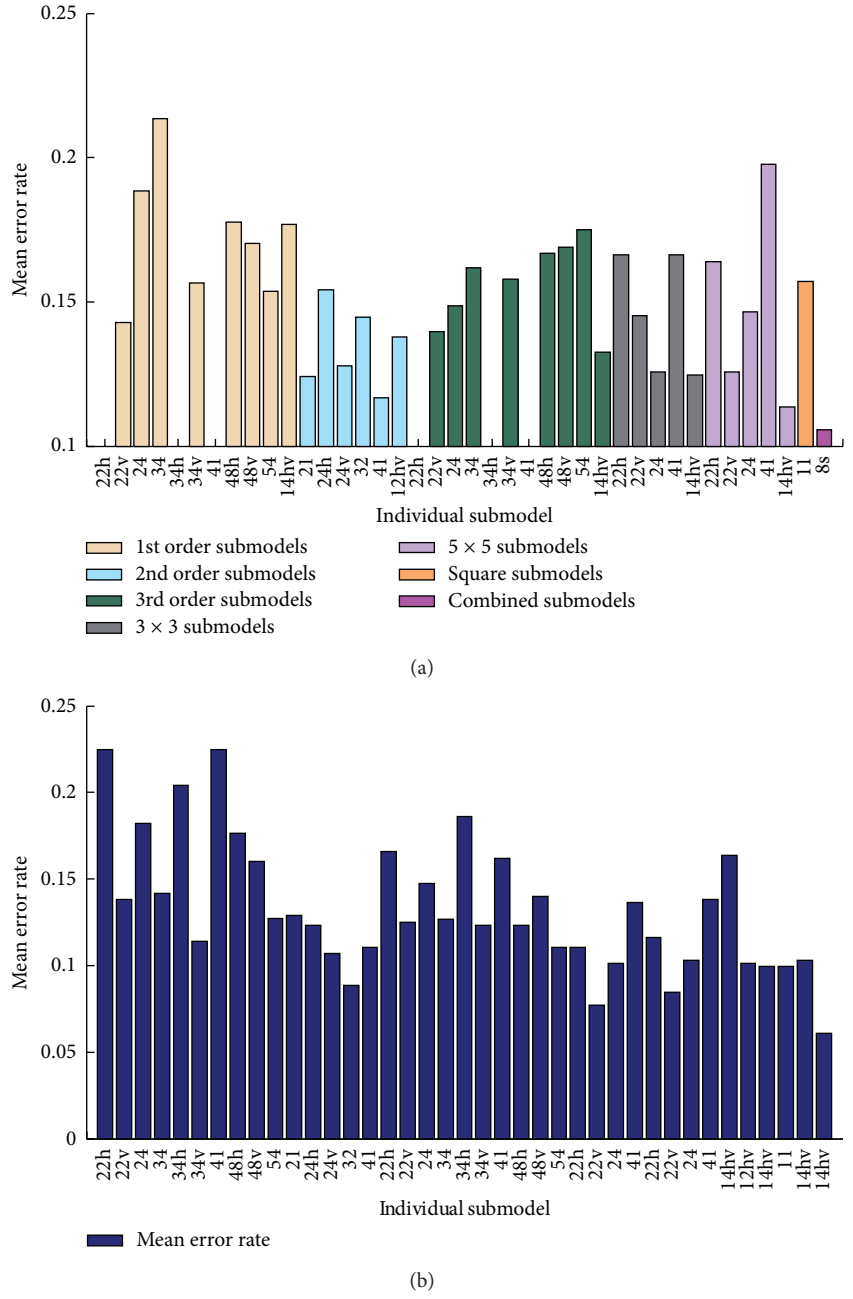


FIGURE 2: The mean errors of individual submodels for (a) Columbia dataset, (b) CASIA V2 dataset. Blank bars represent that the mean errors of the models are not available.

γ are both obtained from the cross validation. We repeat the experiments of each model for 10 times and take the average of them as the final results, which is shown in Figure 2.

4.2. Analysis of Individual Models in Luminance Channel. Figure 2 shows the mean errors of the valid individual models, in both datasets. Generally speaking, all the submodels are useful in the splicing detection, and the top results are observed by using higher order features, the second order submodel, the EDGE3 × 3 models, and the EDGE5 × 5 models, rather than first order submodel.

The detailed results of the submodels are listed in Tables 1, 2, 3, 4, 5 and 6 and the ROC curves of them are given in Figures 3, 4, 5, 6, 7, and 8. As for the merged models, the top eight submodels with best detection results are highlighted in boldface in Tables 2–5, the top four submodels are highlighted in boldface and italic, and the top 2 submodels are s5 × 5_spam14hv and s2_minmax41. The ROC curve of them is given in Figure 9.

4.3. Experiment in Chroma Channel. Based on the results from the previous section, a submodel from EDGE3 × 3,

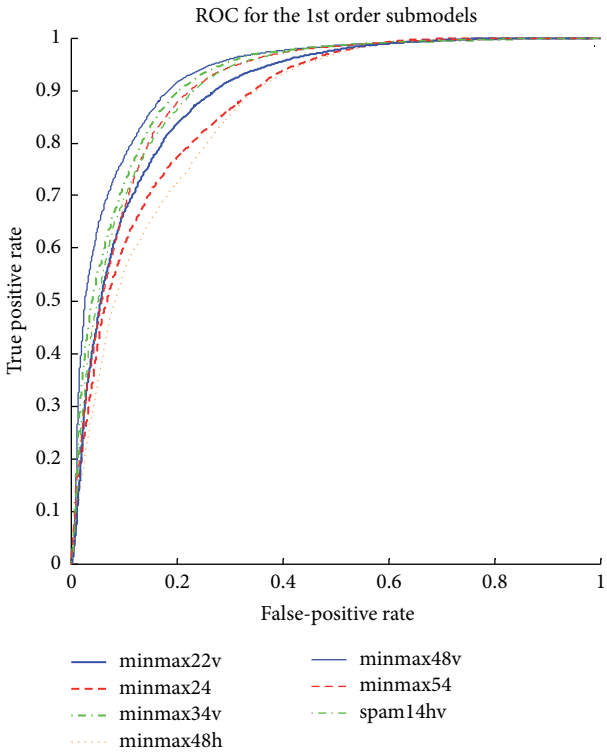


Figure 3: ROC curves for first order features.

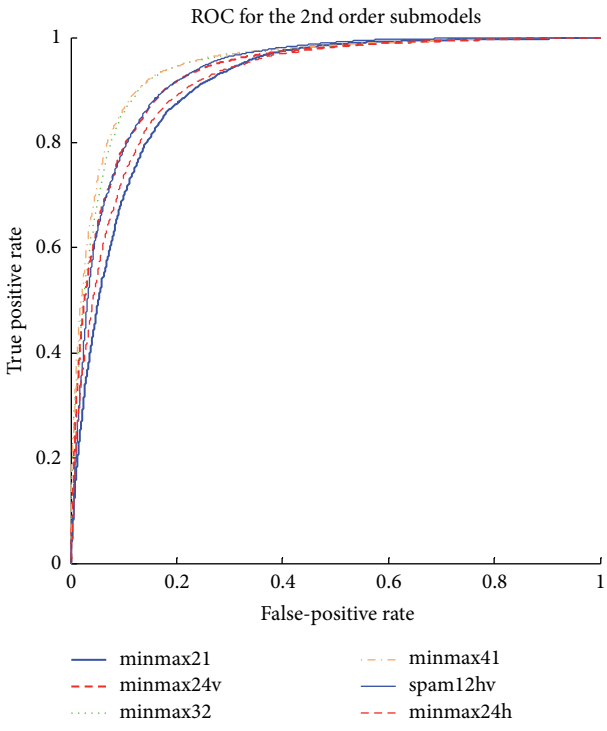


Figure 4: ROC curves for second order features.

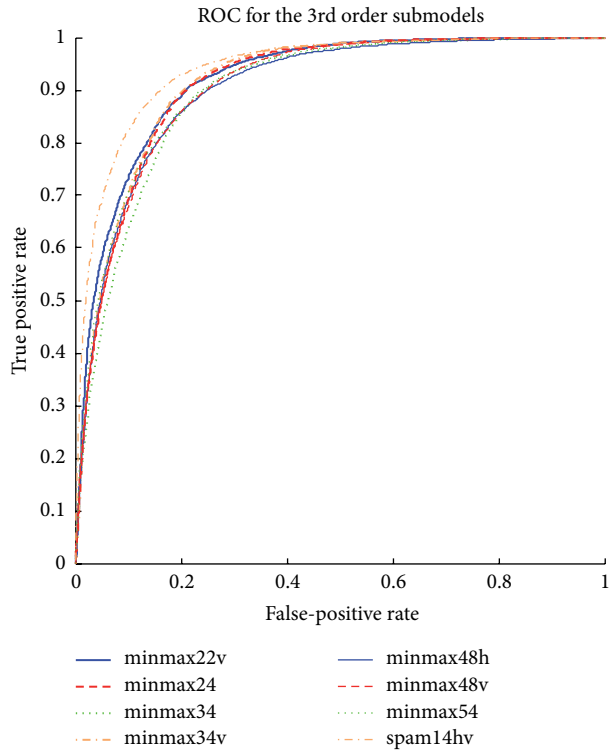


Figure 5: ROC curves for third order features.

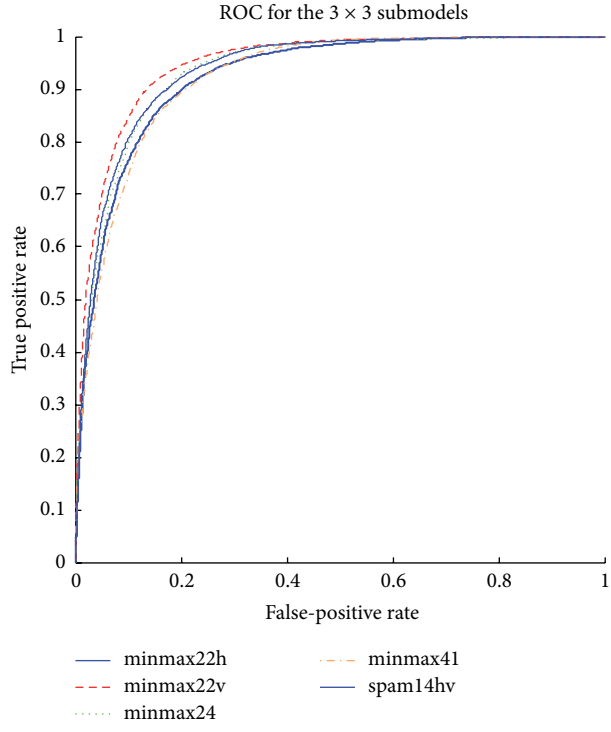


Figure 6: ROC curves for EDGE3 × 3 features.

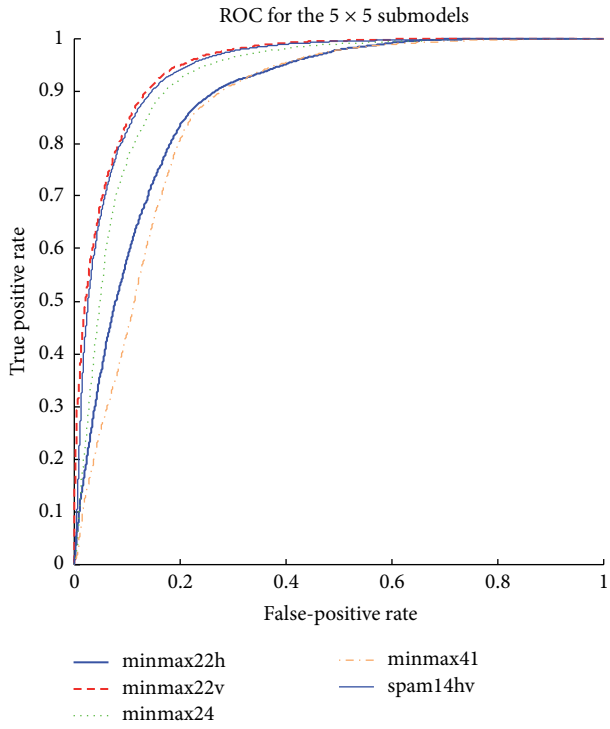


Figure 7: ROC curves for EDGE5 × 5 features.

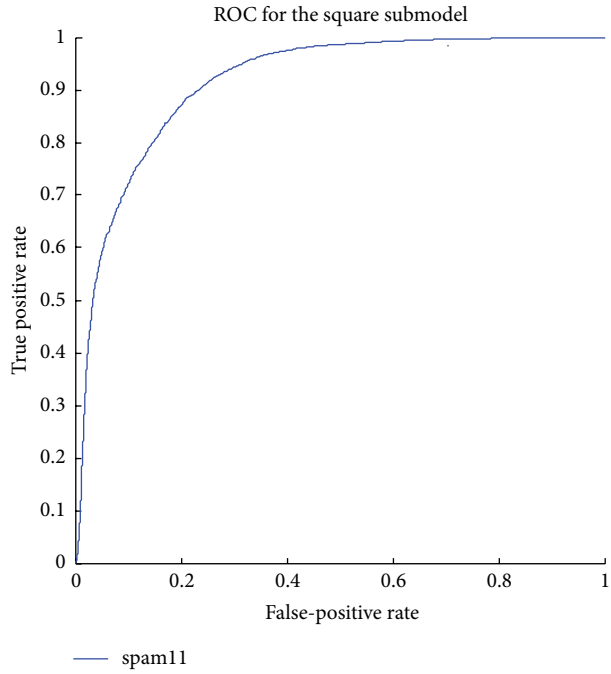


Figure 8: ROC curves for the square features.

s3 × 3_14hv, is selected for further experiment. It is shown in [15] that chorma channel is more effective than luminance channel in detecting image splicing forgeries. The CASIA V2 dataset, which is larger and more challenging, is used for the experiment. The results are listed in Table 7. We can find that the P_{FA} drops about 15% when changing feature from Y channel to Cb, Cr channels, and the P_E drops about 13%. The SRM feature from Cb channel combined with the DCT feature from Cr channel achieved the best P_E. The ROC curve is shown in Figure 10.

5. Conclusions

SRM exploits plenty of statistical features from neighbors of image spatial domain, which could capture abnormality caused by steganography. In this paper, we try to detect the splicing introduced abnormality using SRM. The 39 submodels are exploited to detect image splicing in luminance channel at first. Then, the submodel s 3 × 3_14hv is selected and merged with DCT feature to detect splicing in chorma channel on CASIA V2 dataset, which is larger and more challenging. The experimental results indicate that the proposed method can detect splicing forgeries with much higher accuracy than in luminance channel.

In the near future, we will try to study the nonlearning based detention method in chorma channel, which could be helpful in revealing the underlying mechanic of the method, and could be better for practical application.

Conflict of Interests

The authors declare that there is no conflict of interests.

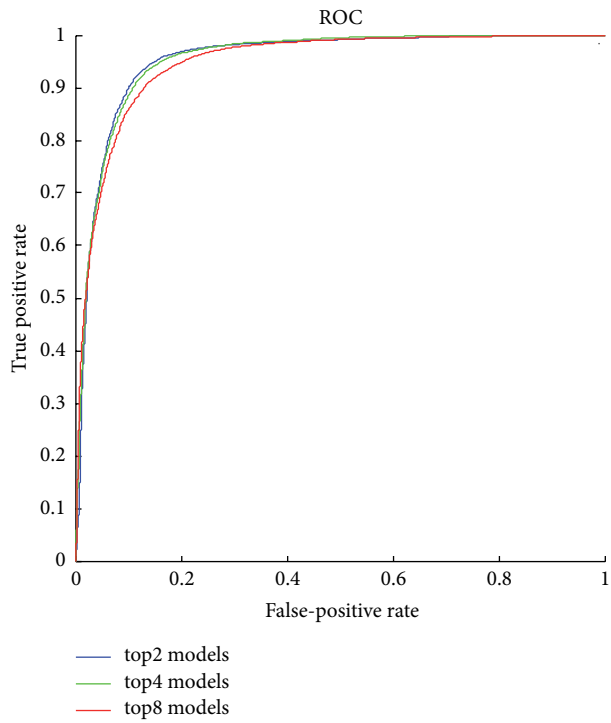


Figure 9: ROC curves for combined features.

Acknowledgments

This paper is supported by the NSF of Jiangsu province (Grant no. BK2010484), and the NSF of China (Grant no. 61170250 and no. 61103201).

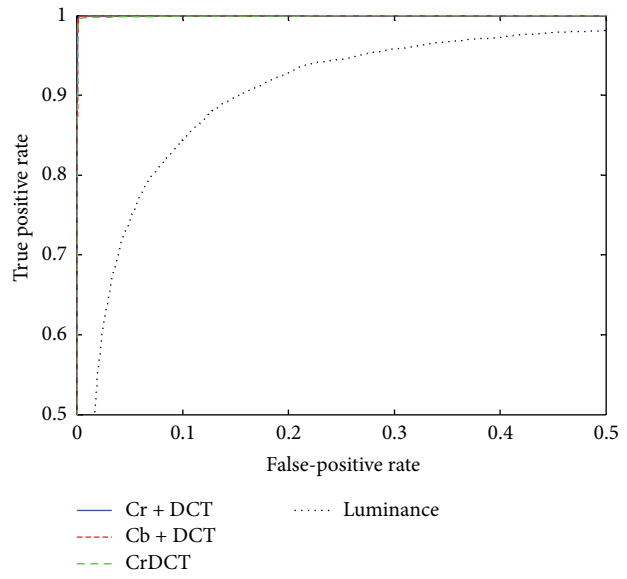


Figure 10: ROC curves for detection results of YCbCr channel on the CASIA v2 dataset.

References

[1] H. Farid, "Detecting digital forgeries using bispectral analysis," Technical Report AIM-1657, AI Lab, Massachusetts Institute of Technology, 1999.

[2] T. T. Ng, S. F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 5, pp. 688–691, 2004.

[3] E. S. Gopi, "Digital image forgery detection using artificial neural network and independent component analysis," *Applied Mathematics and Computation*, vol. 194, no. 2, pp. 540–543, 2007.

[4] E. S. Gopi, N. Lakshmanan, T. Gokul, S. K. Ganesh, and P. R. Shah, "Digital image forgery detection using artificial neural network and auto regressive coefficients," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '06)*, pp. 194–197, Ottawa, Canada, May 2006.

[5] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 549–552, Toronto, Canada, July 2006.

[6] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '07)*, pp. 28–31, July 2007.

[7] J. Wang and H. Zhang, "Exposing digital forgeries by detecting traces of image splicing," in *Proceedings of the 8th International Conference on Signal Processing (ICSP '06)*, Beijing, China, November 2006.

[8] Z. Lint, R. Wang, X. Tang, and H.-Y. Shum, "Detecting doctored images using camera response normality and consistency," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, pp. 1087–1092, usa, June 2005.

[9] W. Chen, Y. Q. Shi, and W. Su, "Image splicing detection using 2-d phase congruency and statistical moments of characteristic function," in *Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents*, Proceedings of SPIE, San Jose, Calif, USA, January 2007.

[10] Y. Q. Shi, C. Chen, and W. Chen, "A natural image model approach to splicing detection," in *Proceedings of the 9th Multimedia and Security Workshop (MM&Sec '07)*, pp. 51–62, New York, NY, USA, September 2007.

[11] D. Fu, Y. Q. Shi, and W. Su, "Detection of image splicing based on Hilbert-Huang transform and moments of characteristic functions with wavelet decomposition," in *Proceedings of the International Workshop on Digital Watermarking*, pp. 177–187, Jeju, Korea, November 2006.

[12] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," in *Proceedings of the IEEE International Conference on Image Processing, ICIP 2009*, pp. 1257–1260, Cairo, Egypt, November 2009.

[13] Z. Zhang, J. Kang, and Y. Ren, "An effective algorithm of image splicing detection," in *Proceedings of the International Conference on Computer Science and Software Engineering*, pp. 1035–1039, Washington, DC, USA, 2008.

[14] J. Dong, W. Wang, T. Tan, and Y. Shi, "Run-length and edge statistics based approach for image splicing detection," in *Proceedings of the International Workshop on Digital-forensics and Watermarking (IWDW '08)*, pp. 76–87, Busan, Korea, November 2008.

[15] W. Wang, J. Dong, and T. Tan, "Image tampering detection based on stationary distribution of Markov chain," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2101–2104, September 2010.

[16] S. Bayram, I. Avcibaş, B. Sankur, and N. Memon, "Image manipulation detection," *Journal of Electronic Imaging*, vol. 15, no. 4, Article ID 041102, 2006.

[17] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transaction on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 628494, 15 pages
http://dx.doi.org/10.1155/2014/628494

Research Article

Color Face Recognition Based on Steerable Pyramid Transform and Extreme Learning Machines

Ayşegül Uçar

Mechatronics Engineering Department, Engineering Faculty, Firat University, 23119 Elazig, Turkey

Correspondence should be addressed to Ayşegül Uçar; agulucar@firat.edu.tr

Received 12 August 2013; Accepted 7 October 2013; Published 16 January 2014

Academic Editors: S. Bourennane and J. Marot

This paper presents a novel color face recognition algorithm by means of fusing color and local information. The proposed algorithm fuses the multiple features derived from different color spaces. Multiorientation and multiscale information relating to the color face features are extracted by applying Steerable Pyramid Transform (SPT) to the local face regions. In this paper, the new three hybrid color spaces, *YSCr*, Z^nSCr, and B^nSCr, are firstly constructed using the *Cb* and *Cr* component images of the *YCbCr* color space, the *S* color component of the *HSV* color spaces, and the Z^n and B^n color components of the normalized *XYZ* color space. Secondly, the color component face images are partitioned into the local patches. Thirdly, SPT is applied to local face regions and some statistical features are extracted. Fourthly, all features are fused according to decision fusion frame and the combinations of Extreme Learning Machines classifiers are applied to achieve color face recognition with fast and high correctness. The experiments show that the proposed Local Color Steerable Pyramid Transform (LCSPT) face recognition algorithm improves seriously face recognition performance by using the new color spaces compared to the conventional and some hybrid ones. Furthermore, it achieves faster recognition compared with state-of-the-art studies.

1. Introduction

Color information of face images is very important for face recognition [1]. In [1–3], it was demonstrated that facial color features could drastically improve recognition performance compared with gray based cues. The *RGB* color space consists of a combination of the red, green, and blue components of images. The other color spaces are derived from *RGB* color spaces by linear or nonlinear transformations. Many recent works about face recognition have used the different color spaces in order to improve the recognition performance [1–8].

Two normalized hybrid color space methods were developed in [5]. In [6], the conventional color spaces such as *HSV*, *RGB*, and *YCbCr* were evaluated comparatively with respect to each other and with respect to gray space by using Principal Component Analysis (PCA). In [7], a question of what kind of color space is suitable for color face recognition was surveyed and a set of optimal coefficients to combine the *R*, *G*, and *B* color components by a discriminant criterion was found. In [8], a new hybrid color space combining the *RGB* and *YIQ* color spaces was proposed. The results revealed that the hybrid color space is more powerful than gray space and even more than *RGB* color space. Some authors generated a new color space as *RCrQ* by taking the *R*, *Cr*, and *Q* color components of *RGB*, *YCbCr*, and *YIQ* color spaces into consideration sequentially in [9]. In this approach, Gabor Transform, Local Binary Patterns (LBP), and Discrete Cosine Transform (DCT) were applied to the *R*, *Cr*, and *Q* chromatic component images, respectively. All information obtained from the three color component images was fused by weighed sum rule. In [10], a new Discriminative Color Features (DCF) method was presented on the *RGBr* color space obtained by means of subtraction of the primary *G* and *B* color component images from primary *R* component image. In [1], Canonical Correlation Analysis (CCA) was presented for face feature extraction and recognition. In [11], Gabor wavelet and LBP were individually applied to *RCrQ*

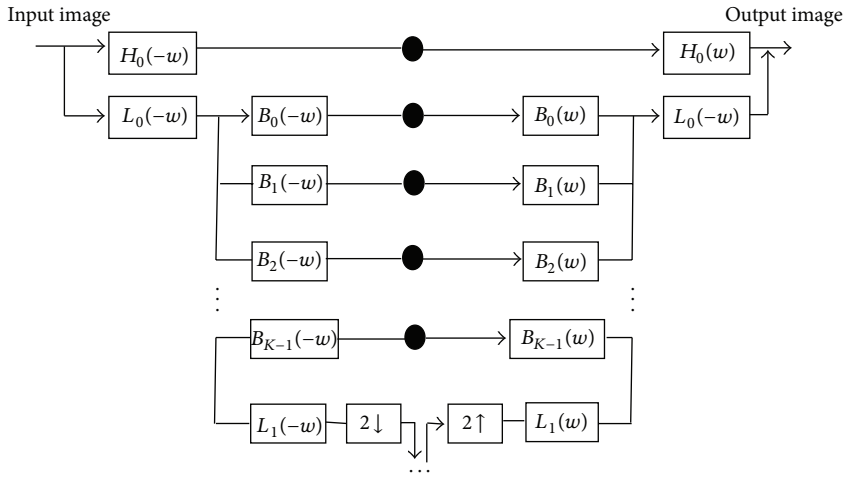


Figure 1: Block diagram of pyramid decomposition [12].

color space and normalized *ZRG* color space proposed in [5] and the outputs of each classifier were combined by the feature fusion and the decision fusion. Although there are many studies, to determine the best hybrid color space is still a challenging problem for face recognition.

This paper evaluates different hybrid color spaces for improving face recognition performance by proposing Local Color Steerable Pyramid Transform (LCSPT) algorithm. Steerable Pyramid Transform (SPT) is a linear multiscale, multiorientation image decomposition technique [12]. SPT aims to represent the original image at different resolutions. Thus, the face image is analyzed by enhancing and isolating image features. SPT was successfully applied on gray images for face recognition in [13].

In color face recognition, the novelties presented in this paper are three fold.

(1) Firstly, an effective color feature extraction algorithm that increases the performance of face recognition by using SPT is proposed. In the algorithm, the features relating to each color component of color face images are extracted by using SPT at different angles and different scales.

(2) Secondly, three novel hybrid color spaces are proposed.

(3) Thirdly, a group of classifiers is applied to the efficient feature set obtained by using SPT with respect to decision frame. In this study, the Extreme Learning Machines (ELMs) for Single Layer Feed-forward Neural Networks (SLFNNs) are developed as an efficient classification method in color face recognition area. SLFNNs have been widely used in face recognition due to their approximation capabilities for nonlinear mappings using input samples. The weights and biases parameters of SLFNNs are usually iteratively adjusted by gradient-based learning algorithms. The past studies in the field of face recognition show that they are generally very slow due to improper learning steps or may easily converge to local minima and they need a number of iterative learning steps in order to obtain better learning performance [14–18]. To get rid of these limitations of SLFNNs for color face recognition in this paper, ELM proposed by Huang et al. [14] is suggested and the combination of ELM and SPT are used. In ELM, the weights of hidden nodes and biases are randomly chosen and output weights are analytically determined. ELM reaches to a good generalization performance in an extremely fast period [18]. ELM has been successfully applied to face recognition area [19, 20]. Also ELM has not been applied together with SPT in the face recognition literature. In this study, ELM and SPT were applied to color face recognition the first time. Comparative and extensive experiments have been illustrated to present the effectiveness of a new algorithm on the color FERET database [21] and the AR database [22].

The rest of this paper is organized as follows. In Section 2, SPT is shortly introduced. The basic architecture of ELM classifier is presented in Section 3. In Section 4, the types of color spaces are introduced. Section 5 describes the proposed LCSPT face recognition algorithm. In Section 6, the comparative experimental results are illustrated. The paper is concluded in Section 7.

2. Steerable Pyramid Transform

The Steerable Pyramid is a multiorientation, multiscale image decomposition method proposed by Freeman and Adelson as an alternative to wavelet transform [12]. In SPT, an image is decomposed into noncorrelated frequency subbands localized in different orientations at different scales. The transform has steerable orientation subbands and a tight frame referred to as self-inverting. Steerable function means that it can be represented as a linear combination of rotated versions of itself. Self-inverting transform means that the synthesis function is the same as the analytical function. The combination of these two properties results in that the subbands become invariant from translation and rotation.

As shown in Figure 1, the input image is firstly decomposed into a highpass subband using a nonoriented highpass filter $H_0(w)$ and then into a lowpass subband using a narrow band lowpass filter $L_0(w)$. Afterwards this lowpass subband is decomposed into K-oriented portions using the bandpass

filters $B_k(w)$ ($k = 0, 1, \ldots, K-1$) and into a lowpass subband L_1 [12]. The decomposition is done recursively by subsampling the lower lowpass subband. The small black boxes represent decomposed subband images. $2\downarrow$ and $2\uparrow$ indicate downsampling and upsampling by a multiplier of 2 along the rows and columns. Recursive steps extract different directional information at a given scale J.

The lowpass filters and highpass filters are defined in the Fourier domain by [12]

$$L_0(r,\theta) = L\left(\frac{r}{2},\theta\right), \quad H_0(r,\theta) = H\left(\frac{r}{2},\theta\right), \tag{1}$$

where r and θ are the polar frequency coordinates. Consider

$$L(r,\theta) = \begin{cases} 2\cos\left(\frac{\pi}{2}\log_2\left(\frac{4r}{\pi}\right)\right), & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 2 & r \geq \frac{\pi}{4} \\ 0 & r \leq \frac{\pi}{2} \end{cases} \tag{2}$$

$$B_k(r,\theta) = H(r)\,Q_k(\theta), \quad k \in [0, K-1],$$

where $B_k(r,\theta)$ represents the K-directional bandpass filters used in the recursive steps with radial and angular parts, defined as

$$H(r) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{2r}{\pi}\right)\right), & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 1 & r \geq \frac{\pi}{2} \\ 0 & r \leq \frac{\pi}{4} \end{cases} \tag{3}$$

$$Q_k(\theta) = \frac{(K-1)!}{\sqrt{K\,[2\,(K-1)]!}}\left[2\cos\left(\theta - \frac{\pi k}{K}\right)\right]^{K-1}.$$

Figure 2 shows all filtered images at 3 scales (128×128, 64×64, and 32×32) and 4 orientation subbands ($-\pi/4$, 0, $\pi/4$, and $\pi/2$) on R component image of a cropped original FERET image. The SPT can locally detect the multiscale edges of facial images [13]. Detected features are noticeable by the first visual area of human visual cortex. In SPT, the lowest spatial-frequency subbands include distinctive edge information, whereas the higher spatial-frequency subbands contain finer edge information. The SPT coefficients consist of much redundant or irrelevant information. A suitable combination of these subbands can provide superior results. In [23], facial expression recognition was carried out by using only one subband.

3. Extreme Learning Machine

The architecture of a simple conventional ELM proposed by Huang et al. [14–18] which is shown in Figure 3 is similar to SLFNNs with M hidden neurons and common activation functions.

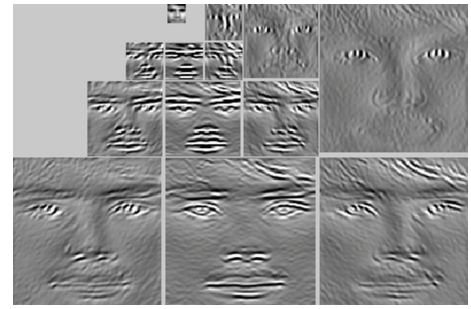


Figure 2: Steerable Pyramid Transform ($K = 4$ and $J = 3$) on R component of a cropped original color FERET image.

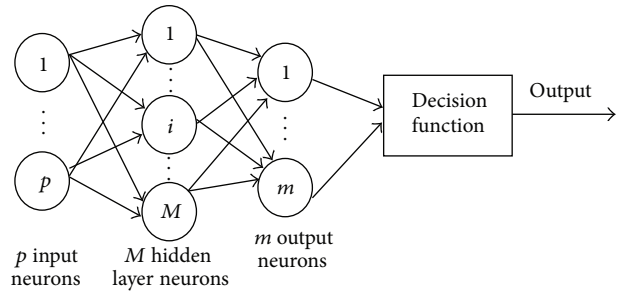


Figure 3: A simple ELM architecture.

Suppose we have a set of N arbitrary distinct samples (x_i, y_i) where $x_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]^T \in R^p$ is a p-dimensional input vector and $y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^T \in R^m$ is an m-dimensional output vector. The input-output relation of a conventional ELM with M hidden nodes and activation function $g(x)$ has the form

$$\sum_{i=1}^{M}\beta_i g(x_k) = \sum_{i=1}^{M}\beta_i g(w_i \cdot x_k + b_i) = o_k, \quad k = 1, \ldots, N, \tag{4}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{ip}]^T \in R^p$ means the weights between the inputs nodes and an ith hidden node, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T \in R^m$ means the weights between an ith hidden node and output nodes, and b_i means the threshold of an ith hidden node.

The conventional ELM in (4) can approximate N samples with zero error by satisfying

$$\sum_{k=1}^{N}\|o_k - y_k\| = 0 \tag{5}$$

or

$$\sum_{i=1}^{M}\beta_i g(w_i \cdot x_k + b_i) = y_k, \quad k \in \{1, 2, \ldots, N\}. \tag{6}$$

The output of ELM can be written more compactly in matrix form as

$$G\beta = Y, \tag{7}$$

where

$$G(w_1,\ldots,w_M,b_1,\ldots,b_M,x_1,\ldots,x_N) = \begin{bmatrix} g(w_1\cdot x_1+b_1) & \cdots & g(w_M\cdot x_1+b_M) \\ \vdots & \cdots & \vdots \\ g(w_1\cdot x_N+b_1) & \cdots & g(w_M\cdot x_N+b_M) \end{bmatrix}_{N\times M},$$

$$\beta_i = [\beta_{i1},\beta_{i2},\ldots,\beta_{im}]^T{}_{M\times m}, \qquad Y_i = [y_1,y_2,\ldots,y_N]^T{}_{N\times m}. \tag{8}$$

In (8), G is the hidden layer output matrix of ELM and g is infinitely differentiable activation function; the number of hidden nodes is chosen as $M \ll N$. Here, the $\widehat{w}_i, \widehat{b}_i, \widehat{\beta}$ $(i = 1,\ldots,M)$ parameters of conventional SLFNN are adjusted by solving the primal optimization problem:

$$\left\|G\left(\widehat{w}_1,\ldots,\widehat{w}_K,\widehat{b}_1,\ldots,\widehat{b}_K\right)\widehat{\beta}-Y\right\| = \min_{w_i,b_i,\beta}\left\|G(w_1,\ldots,w_M,b_1,\ldots,b_M)\beta-Y\right\|. \tag{9}$$

The objective function for optimization problem in (9) is expressed as

$$E = \sum_{k=1}^{N}\left(\sum_{i=1}^{M}\beta_i g(w_i\cdot x_k+b_i)-y_k\right)^2. \tag{10}$$

The parameters are optimized by calculating the negative gradients of objective function in (10) with respect to w_i, b_i, β_i. Consider

$$w_j = w_{j-1} - \tau\frac{\partial E}{\partial W}. \tag{11}$$

The accuracy and learning speed of gradient based method particularly depend on the learning rate, τ. Small learning rate provides very slow convergence, whereas a larger learning rate exhibits the bad local minima effect. ELM uses minimum norm least-square solution to get rid of these limitations. The weight and bias values of ELM are randomly assigned unlike SLFNNs. Output weights of ELM are analytically determined through a generalized inverse operation of the hidden layer weight matrices, since the learning problem is converted into a simple linear system. So it is obtained extremely fast with better generalization performance than those of traditional SLFNN for hidden layers with infinitely differentiable activation functions. The final ELM achieves not only the smallest training error but also the smallest generalization error thanks to the obtained smallest norm of output weights similar to Support Vector Machines (SVM) [24].

For randomly fixed weights in the hidden nodes, the learning of ELM is equal to a least-square solution in (9). G is a nonsquare matrix for $M \ll N$. ELM represented by the linear system in (7) gives a norm least-square solution as $\widehat{\beta} = G^*Y = (G^TG)^{-1}G^TY$, where G^* is the Moore-Penrose generalized inverse of matrix G. The smallest training error is achieved by using

$$\left\|G\widehat{\beta}-Y\right\| = \left\|GG^*Y-Y\right\| = \min_{\beta}\left\|G\beta-Y\right\|. \tag{12}$$

The performance of ELMs having different activation functions has been presented for both regression and classification in [14–18]. In this paper, we are interested in the ELM classifiers with a sigmoid activation function.

4. Hybrid Color Spaces for Face Recognition

The hardware-oriented models including digital image processing commonly use *RGB* color space. Each pixel of a color image is represented in the hardware as binary values for the red, green, and blue color components. Different color spaces are used for different applications. Hence, the *RGB* color space can be converted to the desired color space by using the values in some formulation with respect to the application. The color components of *RGB* color space have largely correlation with each other. Hence the conventional color spaces such as *YIQ*, *YCbCr*, $L^*a^*b^*$, and *HSV* are more effective than original *RGB* color space at face recognition. In this paper, we investigated the color spaces in the literature and their hybrid color component combinations and their color component combinations for improving face recognition performance.

The *HSV* (hue, saturation value) color space is defined as follows [25]:

$$H = \begin{cases}\theta & \text{if } B\le G \\ 360-\theta & \text{if } B>G\end{cases}$$
$$S = 1-\frac{3}{(R+G+B)}\left[\min(R,G,B)\right], \tag{13}$$
$$V = \max(R,G,B),$$

where

$$\theta = \cos^{-1}\left\{\frac{0.5\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{1/2}}\right\}. \tag{14}$$

In this color space, hue (H) is a measure of the spectral composition of a color, saturation (S) shows the relative purity or the amount of white light mixed with a hue, and value (V) refers to the luminance of the image. This model is commonly used for face detection and skin detection [26–28].

The *YCbCr* color space is given by [29]

$$\begin{bmatrix}Y\\Cb\\Cr\end{bmatrix} = \begin{bmatrix}0.257 & 0.504 & 0.098\\-0.148 & -0.291 & 0.439\\0.439 & -0.368 & -0.071\end{bmatrix}\begin{bmatrix}R\\G\\B\end{bmatrix} + \begin{bmatrix}16\\128\\128\end{bmatrix}, \tag{15}$$

where Cb and Cr are chrominance components and Y is separating luminance component. This space is effective for skin color segmentation and face detection [26–28, 30].

The *YIQ* color space is computed by

$$\begin{bmatrix}Y\\I\\Q\end{bmatrix} = \begin{bmatrix}0.299 & 0.5870 & 0.1140\\0.5957 & -0.2745 & -0.3213\\0.2115 & -0.5226 & -0.3111\end{bmatrix}\begin{bmatrix}R\\G\\B\end{bmatrix}, \tag{16}$$

where I stands for in-phase and Q stands for "quadrature", which is based on quadrature amplitude modulation. Consider

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{17}$$

The $L^*a^*b^*$ color spaces are defined based on the XYZ tristimulus values by the following equations:

$$\begin{aligned} L^* &= \begin{cases} 116\left(\dfrac{Y}{Y_n}\right)^{1/3} - 16 & \text{if } \dfrac{Y}{Y_n} > 0.008856 \\ 903.3\left(\dfrac{Y}{Y_n}\right) & \text{if } \dfrac{Y}{Y_n} \le 0.008856 \end{cases} \\ a^* &= 500 * \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right), \\ b^* &= 200 * \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right), \end{aligned} \tag{18}$$

where X_n, Y_n, and Z_n are the tristimulus values of the reference white point. Consider

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > 0.008856 \\ 7.787 * t + \dfrac{16}{116} & \text{if } t \le 0.008856. \end{cases} \tag{19}$$

The $L^*a^*b^*$ color space corresponds to brightness ranging from black (0) to white (100). a^* component corresponds to the measurement of redness (positive values) or greenness (negative values). b^* component corresponds to the measurement of yellowness (positive values) or blueness (negative values). This color space was effectively used for color face expression recognition in [31].

The normalized XYZ and RGB color spaces are obtained by using the across-color-component normalization technique in [4]. In this paper, the normalized XYZ and RGB color spaces are names as XYZ-n and RGB-n, respectively. The normalized color components, X, Y, Z, R, G, and B are named as X^n, Y^n, Z^n, R^n, G^n, and B^n, respectively. XYZ-n and RGB-n color spaces are defined as

$$\begin{aligned} \begin{bmatrix} X^n \\ Y^n \\ Z^n \end{bmatrix} &= \begin{bmatrix} 0.6070 & 0.1740 & 0.2000 \\ -0.0901 & 0.3631 & -0.2730 \\ -0.4600 & -0.1986 & 0.6586 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \\ \begin{bmatrix} R^n \\ G^n \\ B^n \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ -0.5774 & 0.7887 & -0.2113 \\ -0.5774 & -0.2113 & 0.7887 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \end{aligned} \tag{20}$$

In [10], a simple effective model was generated by means of the subtraction of primary colors with respect to Ockham's razor principle. In this paper, the color space is named as RGB-r. The color components, G and B, are named as Gr and Br, respectively. Consider

$$\begin{bmatrix} Rr \\ Gr \\ Br \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{21}$$

Generally, the HSV and $YCbCr$ color spaces are the best color spaces used for skin detection and face detection [3, 26–28, 30, 31]. The R component images have fine face region [32]. The Cr and Cb component images contain partial face contour information [9]. The S and V components are the powerful component images for color face recognition [6, 29], whereas RIQ color space [8], $RCrQ$ color space [9], RGB-r color space [10], and ZRG-n color space consisting of the color components of XYZ-n and RGB-n color spaces [11] have been powerful color spaces for face recognition.

5. Feature Extraction for Color Face Recognition

This section details the novel color feature extraction and multiple feature combination methods for the proposed LCSPT face recognition algorithm. The algorithm incorporates features such as local spatial information and color information for improving face recognition performance. The color information is obtained by using novel hybrid color spaces derived from six conventional color spaces, RGB, YIQ, $YCbCr$, XYZ, HSV, and $L^*a^*b^*$ and three hybrid color spaces, the RGB-n [5], XYZ-n [5], and the RGB-r [10]. The hybrid spaces in this paper are constructed by 3 components as in [33]. So the dominant features of each component image are merged.

Illustration of the proposed LCSPT face recognition algorithm is given in Figure 4. The algorithm is applied in five steps.

(1) In this paper, new three color spaces, $YSCr$, Z^nSCr, and B^nSCr are constructed. The new hybrid color spaces consist of the Cb and Cr component images of the $YCbCr$ color space, the S color component of the HSV color spaces, and the Z^n and B^n color components of the normalized XYZ color space.

(2) Each component contrast is enhanced and divided into local partitions by an efficient pixel number. An efficient pixel number is determined by taking the resolution of face image into consideration.

(3) By applying SPT at a specific scale and a specific orientation to each local image portion, the statistical features such as mean, entropy, and variance of the local face images are extracted.

(4) The group of ELM classifiers is employed to classify the statistical features relating to the color component images of each subband. A decision fusion system combines local decisions from each classifier in the group into a single decision. The combination is implemented by a product decision rule to generate a fused decision vector [34].

6. Experiments and Results

This section evaluates the effectiveness of the proposed LCSPT algorithm on possibly the most representative examples of color face recognition. We used the color FERET database [21] and the AR database for experiments [22].

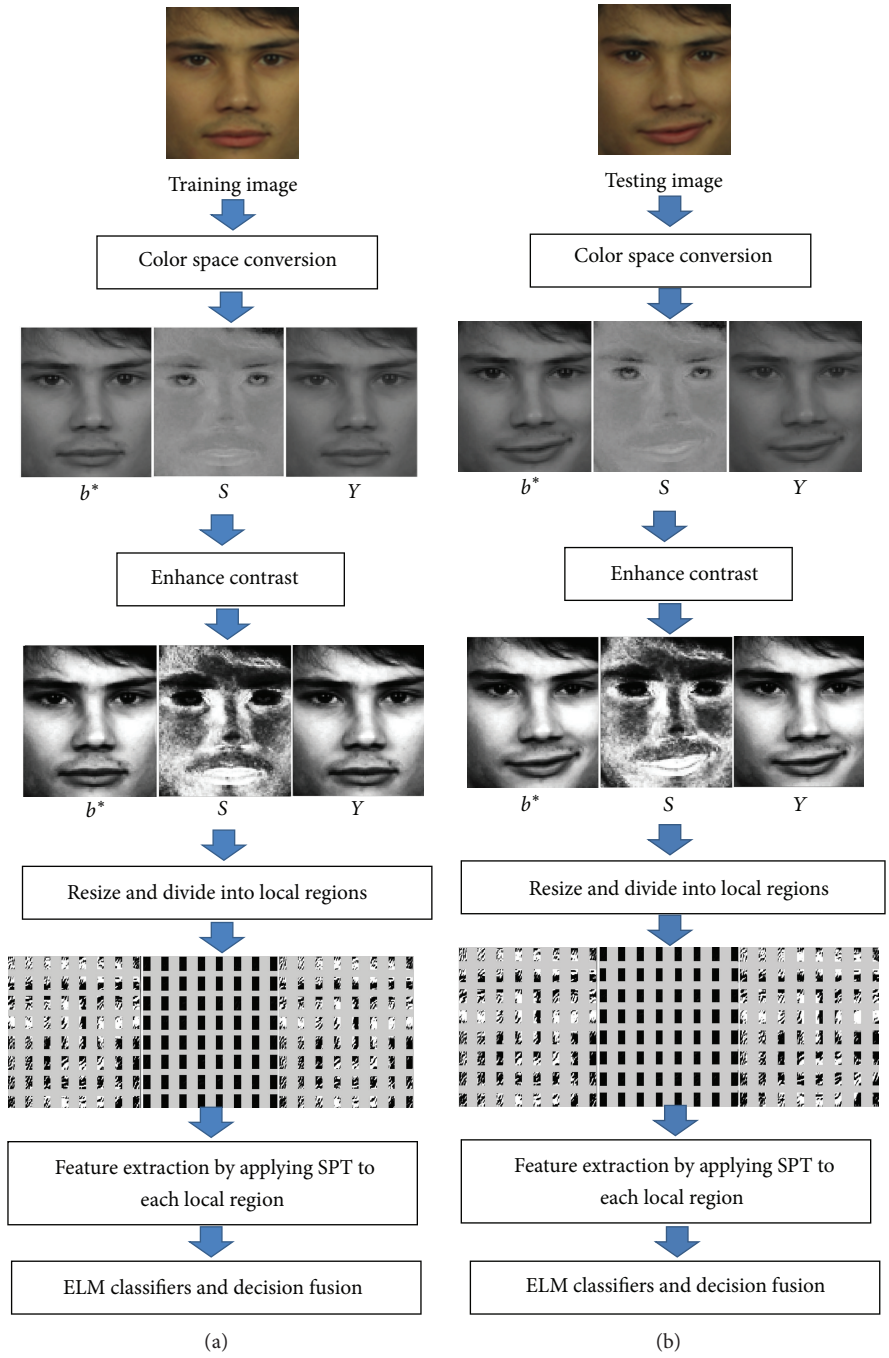


Figure 4: The block scheme of the proposed LCSPT algorithm.

The experiments cover a wide range of facial variability and moderately controlled capturing conditions: facial expression (AR and color FERET), illumination changes (AR and color FERET), aging (color FERET), and slight changes in pose (color FERET). Experiments performed were using a single sample per class for color FERET database as well as more than one sample per class for the AR database.

In color FERET database, we centered all face images with respect to their ground truth face coordinates in [21] cropped and scaled to 128×128 pixels resolution. We used the cropped AR face images in [35]. In particular, we applied SPT at 3 scales (128×128, 64×64, 32×32) and 4 orientation subbands ($-\pi/4$, 0, $\pi/4$, $\pi/2$) to the cropped images. The highpass subband is labeled as *HS* in all tables. In experiments, all subbands relating to only the first scale were used since the results of the first scale were better than the others. If the other scales were used for the face recognition, the recognition performance might increase too. However, the computation complexity will increase since the input space dimension will grow. We tried many efficient pixel numbers such as 4 × 4,

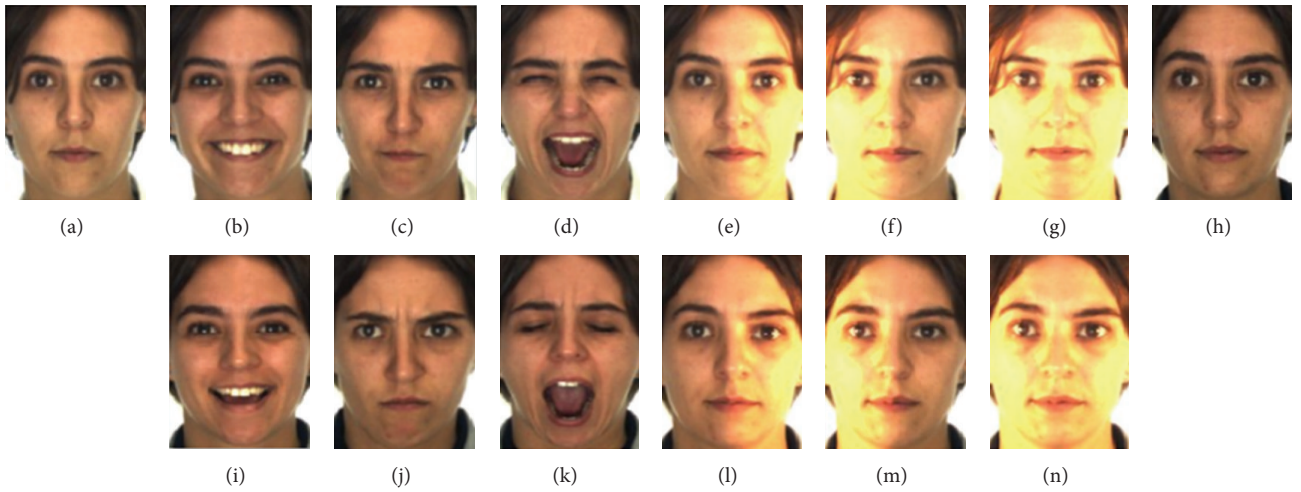


Figure 5: Sample images for one subject of the AR database. The images of (a) neutral expression, (b) smile, (c) anger, (d) scream, (e) left light on, (f) right light on, (g) all sides with light on, and the images of (h)–(n) under the second session.

16 × 16, and 32 × 32. We obtained the best performances for both datasets by using an efficient pixel number of 8 × 8.

All the experiments are run on a personal notebook computer with 2.4-GHz Intel(R) Core(TM)2 Duo processor, 3 GB memory, and Windows 7 operation system. Comparative studies of ELM, SVM, k-Nearest-Neighbors (k-NN), and Feed-forward Neural Networks (FNNs) for the proposed LCSPT face recognition algorithm are carried out. In order to validate both the classification accuracy and the training and testing speeds of SVM, MATLAB interface LIBSVM 2.83 software implementing Sequential Minimal Optimization algorithm, decomposing the overall QP problem into QP subproblems, http://www.csie.ntu.edu.tw/~cjlin/libsvm, was used [36]. The values of kernel and regularization parameters were selected taken as $1/(2\sigma^2) = [2^4, 2^3, 2^2, \ldots, 2^{-10}]$ and $C = [2^{12}, 2^{11}, 2^{10}, \ldots, 2^{-2}]$, respectively. $15 \times 15 = 225$ combinations of the parameters were generated. The best combination was searched [36, 37]. The parameters exhibiting the best 10-fold cross-validation accuracy on the training dataset were accepted as optimal ones as in [24, 36, 37]. 10-fold cross-validation divides the training set into 10 subsets of equal size, and sequentially one subset is tested using the classifier that was trained on the remaining 9 subsets.

ELM, having fast learning and testing speed, allows us to repeat the experiments several times. We changed hidden neuron number of ELM with sigmoid activation function to find the best number. We firstly took as 10 and then increased to the input sample size by the increasing step of 2. We searched ELM with the best correctness. ELM gives better results for large hidden neuron number [15]. On the other hand, an FNN can give good results for small hidden neuron number. The hidden neuron number of the FNNs with sigmoid function was determined in a range from 10 to the input sample size by steps of 2. The FNNs were trained using the conjugate gradient learning algorithm for 500 epochs. For k-NN, we changed the neighbor number from 1 to 5. We run every experiment for each classifier 10 times. Average results are reported in tables.

In addition, we compared the performance of our LCSPT face recognition algorithm with those of color Local Binary Decision (LBD) method in [38] and Local Color Vector Binary Patterns (LCVBP) method in [39]. We used the MATLAB source codes available in [38, 39]. For LBD, we used the local standard deviation filter with a window size of 7 × 5 pixels for a normalization window size of 80 × 90 with respect to the recommendations in [38]. For LCVBP, we rescaled to the size of 112 × 112 pixels and then divided into the local regions with the size of 18 × 21 pixels as in [39]. Our method performed the best recognition performance in all experiments.

We also tried the feature fusion frame for our algorithm. Decision fusion has slightly better performance with respect to the feature fusion in many subbands. The results relating to the feature fusion frame are not included in the tables in order not to corrupt the completeness of the paper. In addition, the feature fusion frame is used together with the dimension reduction techniques in general because it has a large number of features. This also means an additional computational cost. If the dimension reduction techniques are not used, the feature fusion frame requires large computational time.

6.1. Evaluation of Proposed LCSPT on AR Database. The AR database [22] contains over 4,000 frontal view color face images of 126 subjects (76 men and 56 women). Each subject has up to 26 images taken in two sessions, separated by two weeks. Each session contains 13 images with different facial expressions, lighting conditions, and occlusions. The images of 100 subjects were used in our experiments [35]. Figure 5 shows the image samples relating to one person in the AR database used in our experiments. The images consist of neutral expression, smile, anger, scream, left light on, right light on, and all sides light on for both sessions under the same conditions. We planned two experiments on the

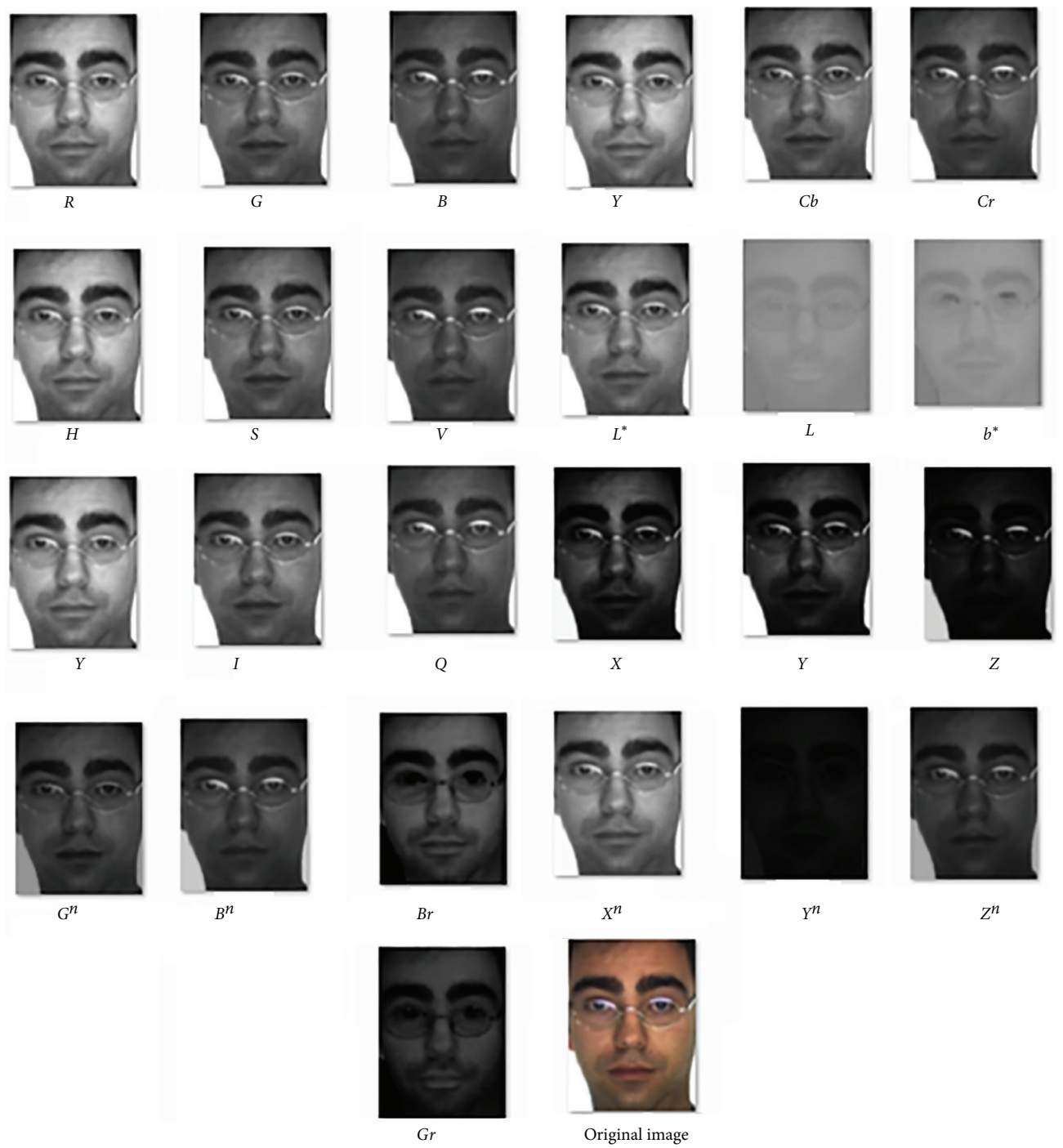


Figure 6: A subject of the AR database and its color component images.

AR database in order to study the robustness of the LCSPT face recognition algorithm. In the first AR experiment, we contained only the variation of facial expressions. We used four images (a)–(d) of session 1 for training and three images (i)–(k) of session 2 for testing. In the second AR experiment, we included both illumination variations and expression variations. We used seven images (a)–(g) of session 1 for training and seven images (h)–(n) of session 2 for testing. We did not consider the occluded face recognition in each session.

Firstly, the performance of our algorithm on 24 color components (R, G, B, H, S, V, Y, Cb, Cr, Y, I, Q, X, Z, L^*, a^*, b^*, X^n, Y^n, Z^n, B^n, G^n, Gr, and Br) of the 9 color spaces (RGB, HSV, $YCbCr$, YIQ, XYZ, $L^*a^*b^*$, RGB-n [5], XYZ-n [5], and RGB-r [10]) is constructed. The color component images of the evaluated color spaces are shown on an image belonging

Table 1: Performance comparison in color component images of LCSPT-ELM face recognition algorithm: the AR database.

	ex1					ex2				
	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$
R	90.97	95.13	94.44	87.50	94.44	91.96	91.07	91.36	86.30	91.36
G	91.66	94.44	93.05	90.97	93.05	90.77	91.66	91.66	85.71	91.07
B	91.66	91.66	93.13	**92.64**	94.11	93.75	93.15	93.75	**91.07**	**94.04**
Y	**93.75**	94.44	93.05	91.66	**94.44**	**94.04**	**93.75**	**94.64**	90.47	**94.04**
Cb	86.11	90.97	90.97	86.11	91.67	85.71	89.28	90.17	87.20	92.26
Cr	86.80	92.36	93.75	88.19	93.05	86.90	92.55	91.66	88.39	92.26
S	89.58	91.66	95.83	90.97	94.44	90.17	94.34	94.94	89.58	94.94
H	85.41	82.63	84.02	75.00	90.27	87.20	85.41	86.30	77.08	86.90
V	90.97	93.05	90.97	87.50	90.97	88.09	86.60	86.90	83.63	88.09
Y	90.97	93.05	93.75	87.50	92.36	91.36	92.26	92.85	88.69	92.26
I	88.89	90.27	95.13	90.27	91.66	87.20	94.04	93.75	88.09	91.36
Q	74.30	77.77	82.63	74.30	81.25	71.42	74.10	80.95	74.70	76.78
L^*	91.66	**95.13**	93.05	90.97	93.75	91.07	94.04	93.15	89.88	93.75
a^*	82.63	90.27	92.36	86.80	90.97	83.63	88.98	91.66	84.52	87.50
b^*	93.62	90.68	92.15	92.15	92.15	85.71	89.28	91.36	84.82	90.47
G^n	90.97	92.36	93.75	88.88	94.44	91.36	92.55	92.26	88.39	92.26
B^n	93.05	92.36	**96.52**	91.66	93.75	92.85	93.45	**94.64**	90.47	93.15
X^n	92.36	94.44	94.44	90.27	**94.44**	**94.04**	**93.75**	94.04	90.77	93.75
Y^n	87.50	89.58	90.97	86.80	88.19	89.28	88.98	91.66	86.90	89.88
Z^n	90.97	92.36	**96.52**	92.36	**94.44**	93.45	93.15	94.04	90.47	**94.04**
X	89.58	94.44	93.75	86.80	90.97	90.47	92.55	92.26	88.69	92.85
Z	92.36	93.05	95.13	88.19	93.75	92.85	92.55	92.26	88.09	93.45
Gr	85.41	92.36	94.44	87.50	92.36	82.73	93.75	90.17	86.30	89.28
Br	89.58	92.36	92.36	88.88	94.44	87.79	91.36	93.45	87.79	94.94

Bold values mean the color components with correctness over 90% for all subbands and the subbands with the highest correctness.

to the AR database as in Figure 6. Table 1 presents our results on 24 color component images. The left side of Table 1 lists the results of AR experiment 1. These results indicate that the correctness at all subbands for G, B, Y, L^*, B^n, X^n, and Z^n color components are over 90%. However, the correctness of the other color components is below 90% in one or many subbands. Taking into consideration all the subbands for the G, B, Y, L^*, B^n, X^n, and Z^n color components, we observe the best results on the Y color component for HS subband, the L^* color component for $-\pi/4$ subband, the Z^n and B^n color components for 0 subband, B for $\pi/4$ subband, and the Y, Z^n, and X^n color components for $\pi/2$ subband. From these results, we conclude that the results of the Y, L^*, Z^n, B^n, and B color components are better than the G color component and even the others for facial expression experiment. In the case of an experiment including the variation of facial expression, Z^n and B^n color components are the best in the color components with respect to the correctness in the their subbands since the fusion of the subbands with high correctness will provide a higher correctness.

The right side of Table 1 lists the results of AR experiment 2. These results indicate that the correctness at all subbands on only B, Y, B^n, X^n, and Z^n color components is over 90%. Taking all the subbands for B, Y, B^n, X^n, and Z^n color components into consideration, we observe the best results on Y and X^n color components for HS subband, Y and X^n color components for $-\pi/4$ subband, Y and B^n color components for 0 subband, B for $\pi/4$ subband, and Y, B and Z^n color components for $\pi/2$ subband. From these results, we conclude that Y color component is specifically better than the others for the experiment including illumination experiment. The results correlate with the literature [26–28, 30]. From all the results in Table 1, we infer that the color components, Y, Z^n, B^n, X^n, and B can be used to obtain an acceptable good performance in both experiments. Hence, we can give our priority to these components in an experiment including the variation of illumination.

We compared our new three hybrid color spaces to 9 color spaces of *RGB*, *HSV*, *YCbCr*, *YIQ*, *XYZ*, $L^*a^*b^*$, *RGB-n* [5], *XYZ-n* [5], and *RGB-r* [10] and the hybrid color space of *RIQ* [8], *RCrQ* [9], and *ZRG-n* [5] and presented advantages with respect to the conventional RGB color space. In order to make our results clearer, we give only the best hybrid color spaces although we tried all combinations of 24 color components. Table 2 shows the recognition correctness of all hybrid color spaces. Specifically, we obtained that the hybrid color spaces generated by using the S and Cr color components combined together with the Y, B^n, X^n, B, and Z^n color components improve more effectively the face recognition performance. For two AR experiments, the results of the $YSCr$, B^nSCr, and Z^nSCr hybrid color spaces outperform those of the powerful conventional color spaces, the other color spaces

Table 2: Performance comparison in different hybrid color spaces of LCSPT-ELM face recognition algorithm: the AR database.

	ex1					ex2				
	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$
RGB	97.72	98.41	97.02	96.33	97.02	96.78	97.08	96.28	94.40	97.08
HSV	95.25	96.61	94.72	95.64	94.69	93.51	94.58	97.58	95.59	93.98
YCbCr	95.63	98.41	97.71	94.94	98.41	95.88	98.26	97.07	94.39	97.07
YIQ	96.33	97.02	97.72	95.64	96.33	95.88	97.07	97.67	95.88	96.78
XYZ	94.24	95.63	96.33	90.77	96.33	93.50	94.99	94.39	90.23	94.99
$L^*a^*b^*$	94.02	99.01	97.02	97.41	98.41	96.78	98.16	97.37	95.98	98.08
RGB-n [5]	93.75	96.82	97.71	95.64	95.69	94.99	97.85	97.37	94.99	97.07
XYZ-n [5]	93.55	97.02	97.72	95.64	97.72	94.99	95.59	95.59	92.91	95.29
RGB-r [10]	94.25	97.02	97.72	95.64	97.72	94.69	98.56	97.37	94.10	97.97
RIQ [8]	94.94	97.71	97.71	96.33	98.41	94.39	96.18	95.88	93.80	95.88
RCrQ [9]	96.33	97.72	97.72	95.64	98.41	94.69	96.28	96.18	93.80	96.18
ZRG-n [5]	97.02	97.72	97.02	94.94	97.02	96.18	96.48	97.37	94.10	98.26
YSCr	96.33	99.11	**99.80**	97.72	99.11	96.78	98.86	96.26	96.78	**99.45**
B^nSCr	95.64	99.11	**99.80**	97.02	99.11	97.07	**99.16**	98.56	96.78	**99.16**
Z^nSCr	95.64	99.11	**99.80**	98.41	99.11	96.18	**99.16**	98.56	96.78	**99.16**

Bold values specify the hybrid color spaces and their subbands with the highest correctness.

Table 3: Comparison of training time, testing time, and correctness of LCVBP, LBD, and LCSPT-ELM: the AR database.

	ex1			ex2		
	Training time (s)	Testing time (s)	Correctness rate (%)	Training time (s)	Testing time (s)	Correctness rate (%)
LCVBP [39]	5850.12	5320.45	98.01	10350	7750	97.56
LBP [38]	3826.01	—	98.23	3795.91	—	98.12
LCSPT-ELM	544.86	360.27	99.80	602.52	358.08	99.16

Table 4: Comparison of parameter adjusting time and testing time of k-NN, FNN, SVM, and ELM: the AR database.

	ex1			ex2		
	Parameter adjusting time (s)	Testing time (s)	Best parameters	Parameter adjusting time (s)	Testing time (s)	Best parameters
k-NN	4.3	0.12	$k = 2$	10.27	0.38	$k = 2$
SVM	119	0.28	$(2^1, 2^{-10})$	508.12	1.157	$(2^1, 2^{-10})$
ELM	12.79	0.034	$N = 104$	82.54	0.080	$N = 122$
FNN	1875	0.028	$N = 44$	11494	0.032	$N = 54$

and the individual color components such as R, G, Y, and Z^n. Moreover, we obtained the best results with the correctness of 99.45 in the *YSCr* hybrid color space. On the other hand, if one wants to have the higher correctness for each color component or each color space in Table 1 and Table 2, then all their subbands could be fused by the decision or feature fusion method [13, 20] in terms of more computational complexity.

In Table 3, we compared our results on the *YSCr* hybrid color space to LBD method in [38] and LCVBP method [39] in terms of the training time and the testing time and the recognition correctness. As can be seen from these results, our LCSPT-ELM face recognition algorithm is the best one in computing time especially. Moreover, the correctness of LCSPT-ELM outperforms the others.

We also compared the parameter adjusting time, the testing time, and correctness of SVM, k-NN, FNN, and ELM. Figure 7 shows the correctness of all classifiers. ELM outperforms the others in terms of the correctness. In Table 4, the results on the Y color component image are given. FNN is the most time consuming method in with regard to the parameter adjusting time, but FNN has the shortest testing time due to the high compact network architecture [15]. The parameter adjusting time of k-NN is the fastest, however, its performance and testing time are worse than that of the ELM for both AR experiments. The advantage of ELM is obviously seen by taking both the correctness and training time into consideration. ELM runs around 6 times faster than SVM and 130 times faster than FNN. After the parameter adjusting process we obtained the optimum parameters for the AR

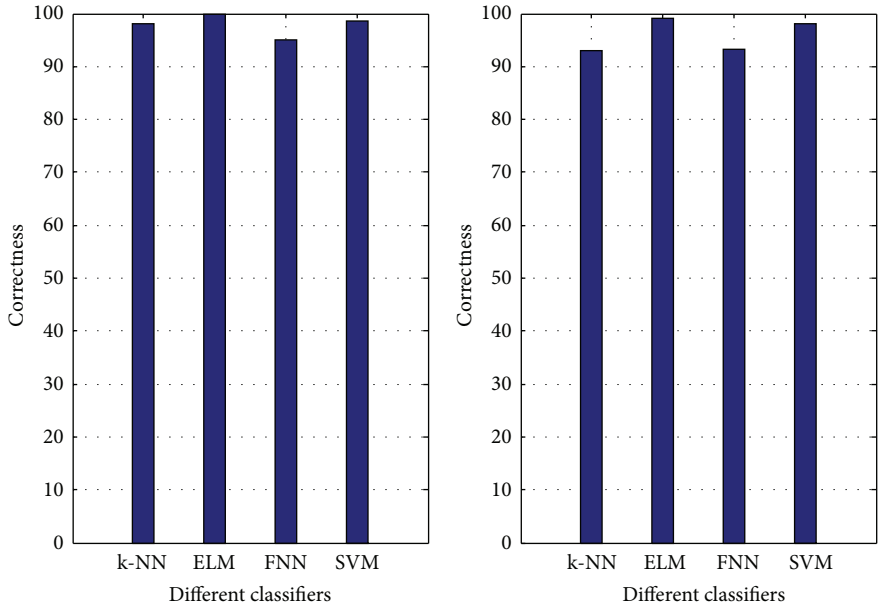


Figure 7: Performance comparison for k-NN, ELM, FNN, and SVM classifiers: AR database: (a) Experiment 1 and (b) Experiment 2.

experiments. The hidden neuron number of FNN and ELM, neighbor number of k-NN, and $(C, 1/(2\sigma^2))$ parameters of SVM are given in Table 4.

6.2. Evaluation of Proposed LCSP on Color FERET Database. The FERET database consists of 11,388 color facial images obtained from 994 subjects being captured in the course of 15 sessions. The images have a resolution of 512 × 768 pixels. The database is very challenging due to significant appearance changes in the individual subjects in terms of aging, facial expressions, glasses, hair, moustache, nonuniform illumination variations, and slight changes in pose. The database is divided into five subsets: fa, fb, fc, dup1, and dup2. fa subset contains one frontal view per subject and in total 1196 subjects. We conducted single-sample-per-class face recognition experiments on the color FERET database [21]. The FERET evaluation methodology requires that the training processing be carried out using only the fa subset. We selected a subset of color face images of 204 persons. Their ground truth face coordinates are available in the color FERET database. We generated the training set by only single image portrait images consisting of 204 people from fa subset, whereas testing set was from the fb subset and the dup 1 subset. Figure 8 shows some sample images relating to the subsets of fa, fb, and dup1 of the color FERET database used in our experiments.

The left side of Table 5 shows the results on the fb subset of color FERET database. We observe the best results on the B_r color component for HS subband, the S color component for $-\pi/4$ subband, the Y, X, and I color components for 0 subband, Cb color component for $\pi/4$ subband, and the B and Z^n color components for $\pi/2$ subband. From these results, we conclude that Br, S, Y, X, I, Cb, B, and Z^n color components are especially better than the others. The right side of Table 5 shows the results on the dup I subset of the color FERET database. As we can see from the right side of Table 5, the best color components are the G, I, Y, and Z^n. If we search the best color components in Table 5 for both experiments on color FERET database, we can observe that the Y, Z^n, and I color components can achieve good correctness for expression, illumination, and aging experiments.

From Table 6, it can be observed that the best hybrid color spaces for both the fb subset and the dup1 subset are $YSCr$, B^nSCr, and Z^nSCr. Taking both the AR database and the color FERET database into consideration, we infer a result that the Y and Z^n color components and the $YSCr$, B^nSCr, and Z^nSCr hybrid color spaces are very effective for our face recognition algorithm.

The comparison results for the $YSCr$ hybrid color space are described in Table 7. It is shown that the proposed LCSPT-ELM outperforms in terms of computing time and recognition correctness. In this paper, the extracted feature number is very small because only 3 features for each local portion of the color component of each image are used. Naturally, training and testing time are very small. On the other hand, the feature number can be reduced by the dimension reduction techniques such as PCA and LDA. Therefore, the computational complexity can be more reduced.

Table 8 shows the average results of LCSPT face recognition algorithm using k-NN, FNN, SVM, and ELM on the Y color component image. Figure 9 depicts the results of LCSPT face recognition algorithm on the $YSCr$ hybrid color space. From Table 8 and Figure 9, we observe that the ELM outperforms the other classifiers.

Table 5: Performance comparison in the color component images of LCSPT-ELM face recognition algorithm: the color FERET database.

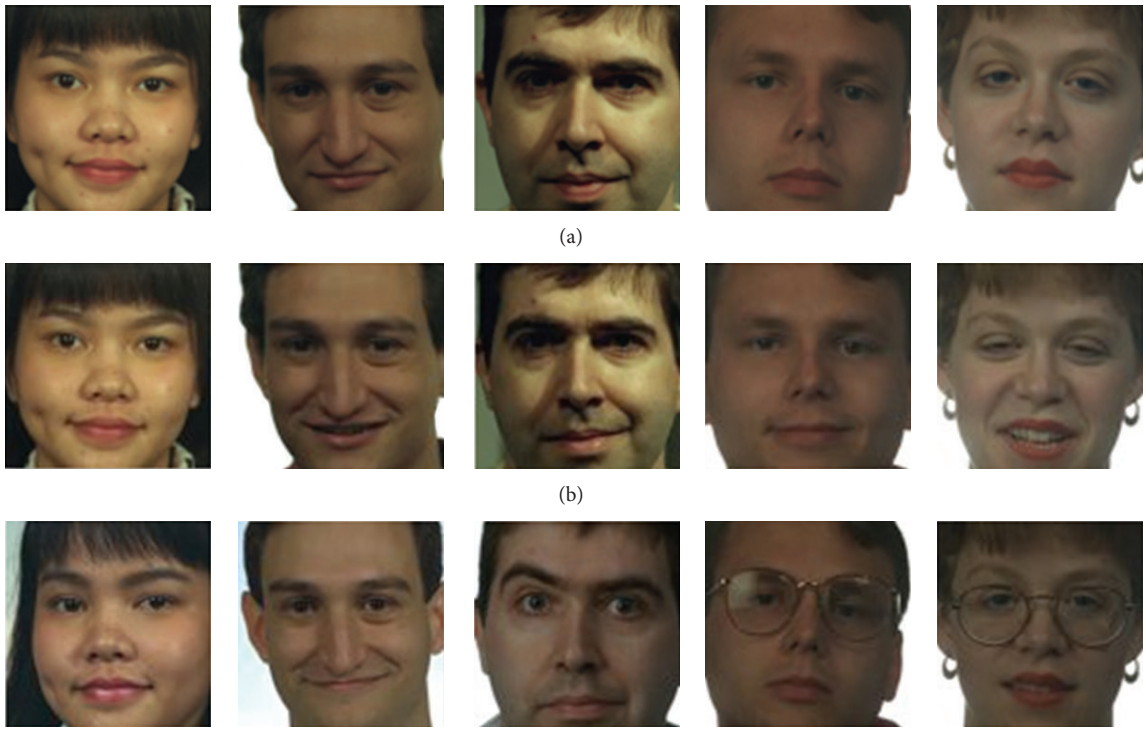
	fb set					dup1 set				
	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$
R	91.19	93.15	91.68	89.72	91.68	73.35	71.88	74.82	72.37	73.35
G	91.19	91.68	94.13	92.66	93.64	**78.25**	72.37	77.27	69.43	73.84
B	92.66	92.66	94.13	93.64	**95.11**	70.90	71.39	71.39	68.45	74.82
Y	91.70	92.66	**95.64**	91.19	93.64	72.86	70.90	**84.39**	**75.39**	74.82
Cb	94.62	92.17	94.62	**95.11**	94.62	69.92	64.03	73.35	61.09	73.84
Cr	93.64	93.15	93.15	89.72	92.66	75.80	75.31	79.72	67.47	77.27
S	90.70	**95.60**	95.11	94.13	96.09	70.41	69.43	70.90	72.86	77.27
H	92.66	91.19	91.68	87.27	93.15	69.43	62.56	69.92	55.70	65.50
V	87.76	92.66	92.17	90.21	91.68	69.43	73.84	74.82	71.88	74.82
Y	89.72	92.66	94.62	91.19	94.13	71.39	72.37	76.29	74.82	76.78
I	92.66	93.64	**95.64**	93.15	94.62	76.29	**77.76**	80.21	72.37	82.17
Q	73.05	74.03	76.98	75.50	76.98	49.82	53.25	56.19	52.76	55.21
L^*	92.17	92.17	93.64	91.68	93.64	73.35	71.88	70.90	70.41	73.35
a^*	93.64	91.68	90.21	90.70	93.64	74.82	69.43	76.78	63.05	76.78
b^*	94.62	91.68	93.15	93.15	93.15	66.49	65.50	71.88	66.49	72.86
G^n	93.15	92.66	93.64	91.19	94.13	70.41	69.43	74.33	66.49	71.88
B^n	91.70	92.66	94.13	92.17	93.15	68.94	69.92	70.41	68.94	75.31
X^n	91.68	93.15	93.64	91.19	92.17	73.84	70.90	69.92	69.43	74.82
Y^n	91.68	89.23	89.23	89.72	91.19	71.39	65.01	70.41	60.60	69.92
Z^n	92.17	91.68	94.62	92.17	**95.11**	69.92	68.45	73.35	68.45	**83.29**
X	90.21	93.64	**95.64**	91.19	92.66	69.43	72.37	76.78	75.37	76.78
Z	88.74	92.66	94.13	90.21	93.15	66.00	67.47	73.84	69.92	73.84
Gr	92.66	92.17	92.66	90.70	94.13	74.82	72.86	76.78	65.50	77.27
Br	**96.58**	95.11	94.13	94.13	93.64	72.85	74.33	75.80	71.88	77.76

Bold values specify the color components and their subbands with high correctness.

Table 6: Performance comparison in different hybrid color spaces of LCSPT-ELM face recognition algorithm: the color FERET database.

	fb set					dup1 set				
	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$	HS	$-\pi/4$	0	$\pi/4$	$\pi/2$
RGB	97.79	97.790	97.30	96.81	96.81	76.86	73.92	74.41	73.43	75.88
HSV	90.42	96.850	97.71	93.34	95.92	80.32	80.34	80.24	78.11	85.17
YCbCr	99.26	97.30	98.77	98.28	98.77	82.74	84.70	84.70	81.27	84.21
YIQ	96.32	97.30	98.77	96.32	97.79	75.88	77.35	82.25	76.86	80.29
XYZ	91.42	94.85	96.81	93.38	96.32	70.00	72.94	76.86	75.88	78.33
$L^*a^*b^*$	92.12	97.88	97.30	97.17	98.81	80.74	83.23	83.41	82.43	84.84
RGB-n [5]	93.89	94.85	97.83	93.87	95.32	80.68	84.70	83.23	80.80	79.25
XYZ-n [5]	92.89	94.85	96.81	93.87	96.32	74.90	73.43	75.88	75.88	79.31
RGB-r [10]	99.70	96.81	97.30	95.83	97.79	80.78	84.70	84.21	79.80	82.25
RIQ [8]	95.83	96.81	98.28	96.81	98.28	75.39	80.29	82.74	79.31	79.80
RCrQ [9]	98.28	97.30	97.79	97.79	96.81	76.37	79.80	81.76	79.31	81.76
ZRG-n [5]	94.36	94.85	95.83	93.87	96.32	75.88	75.88	77.35	74.90	77.84
YSCr	96.81	98.28	98.77	98.28	**99.75**	85.19	84.70	84.70	83.23	**89.11**
B^nSCr	97.79	98.28	98.77	99.26	**99.75**	82.74	84.21	83.72	81.27	**89.11**
Z^nSCr	97.79	98.28	98.77	99.26	**99.75**	83.72	83.72	85.19	82.74	**89.11**

Bold values specify the hybrid color spaces and their subbands with the highest correctness.

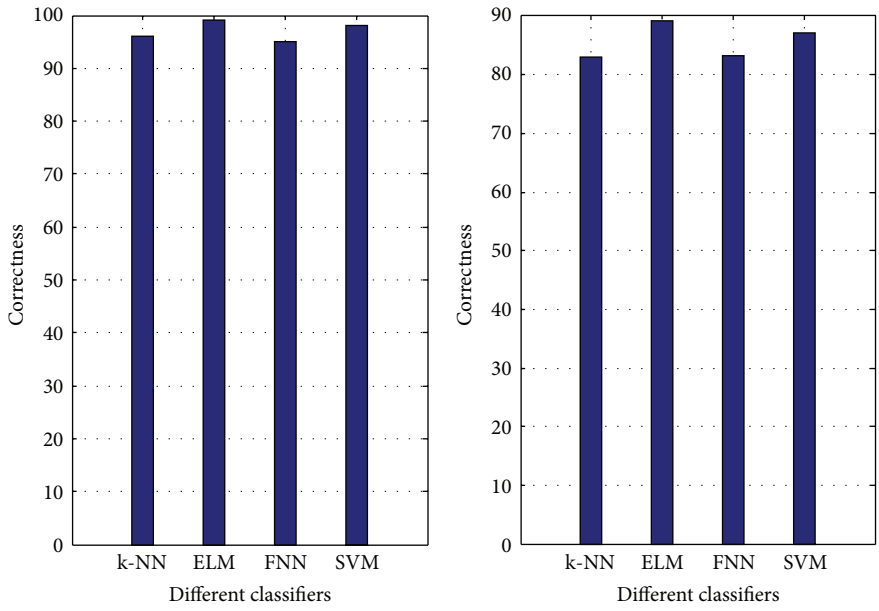


(a)

(b)

(c)

Figure 8: Some sample images of the subsets of (a) fa, (b) fb, and (c) dup1 of the color FERET database.



(a) (b)

Figure 9: Performance comparison for k-NN, ELM, FNN, and SVM classifiers: the color FERET database: (a) fb subset and (b) dup1 subset.

7. Conclusions

This paper presents a novel face recognition algorithm by means of fusing color and local spatial information (see Supplementary Material available online at http://dx.doi.org/10.1155/2014/628494)). The effectiveness of the proposed algorithm is assessed on 6 conventional color spaces, *RGB*, *HSV*, *YCbCr*, *YIQ*, *XYZ*, and $L^*a^*b^*$, and 6 powerful hybrid color spaces developed in the literature, *RGB-r* [10], *XYZ-n* [5], *RGB-n* [5], *ZRG-n* [5], *RIQ* [8], and *RCrQ* [9]. In addition, 3 new hybrid color spaces are constructed in this paper. In particular, the proposed hybrid color spaces,

Table 7: Comparison of training time, testing time, and correctness of LCVBP, LBD, and LCSPT-ELM: the color FERET database.

	fa set			dup1 set		
	Training time (s)	Testing time (s)	Correctness rate (%)	Training time (s)	Testing time (s)	Correctness rate (%)
LCVBP [39]	8021.77	7406.13	97.12	7676.57	8338.85	82.43
LBP [38]	3700.91	—	98.03	3091.71	—	88.49
LCSPT-ELM	544.86	360	99.75	602.52	358	89.11

Table 8: Comparison of parameter adjusting time and testing time of k-NN, FNN, SVM, and ELM: the color FERET database.

	fb set			dup1 set		
	Parameter adjusting time (s)	Testing time (s)	Best parameters	Parameter adjusting time (s)	Testing time (s)	Best parameters
k-NN	16.98	0.1512	$k = 2$	12.37	0.15	$k = 2$
SVM	99.12	1.23	$(2^{-2}, 2^{-10})$	91.32	1.12	$(2^{-2}, 2^{-10})$
ELM	16.38	0.06	$N = 120$	15.28	0.032	$N = 114$
FNN	2325.80	0.02	$N = 52$	2313.62	0.02	$N = 58$

YSCr, Z^nSCr, and B^nSCr are configured as the combination of the *Cb* and *Cr* component images of the *YCbCr* color space, the *S* color component of the *HSV* color spaces, and the Z^n and B^n color components of the normalized *XYZ* color space. Experiments are constructed using the most challenging color FERET database and AR database.

The novelty of this paper is based on the following aspects: (i) a novel color feature extraction method is introduced by applying SPT algorithm to each color component of color face images; (ii) new hybrid color spaces are presented for improving the color face recognition performance being used together with SPT algorithm; (iii) in decision frame, the fusion of ELM classifiers is developed for fast color face recognition.

Experimental results show that SPT is an effective tool for extracting information from the color face images. The proposed LCSPT-ELM algorithm has very short training and testing time and a good recognition correctness. It is illustrated that the new hybrid color spaces of *YSCr*, B^nSCr, and Z^nSCr have the best performance on our algorithm. LCSPT-ELM algorithm can be used for real time face recognition applications thanks to short testing time and parameter adjusting time.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper was supported by the Firat University Scientific Research Projects Foundation (no. MF.12.33). The author would like to thank Professor Guang-Bin Huang for his personal help about ELM and for his codes.

References

[1] X. Jing, S. Li, C. Lan, D. Zhang, J. Yang, and Q. Liu, "Color image canonical correlation analysis for face feature extraction and recognition," *Signal Processing*, vol. 91, no. 8, pp. 2132–2140, 2011.

[2] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 39, no. 5, pp. 1217–1230, 2009.

[3] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002.

[4] M. Villegas, R. Paredes, A. Juan, and E. Vidal, "Face verification on color images using local features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08)*, pp. 1–6, Anchorage, Alaska, USA, June, 2008.

[5] J. Yang, C. Liu, and L. Zhang, "Color space normalization: enhancing the discriminating power of color spaces for face recognition," *Pattern Recognition*, vol. 43, no. 4, pp. 1454–1466, 2010.

[6] S. Yoo, D. G. Sim, Y. G. Kim, and R. H. Park, "Performance comparison of principal component analysis-based face recognition in color space," in *Advanced Biometric Technologies*, G. Chetty and J. Yang, Eds., pp. 281–296, InTech, Rijeka, Croatia, 2011.

[7] J. Yang, C. Liu, and J.-Y. Yang, "What kind of color spaces is suitable for color face recognition?" *Neurocomputing*, vol. 73, no. 10-12, pp. 2140–2146, 2010.

[8] Z. Liu and C. Liu, "A hybrid color and frequency features method for face recognition," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1975–1980, 2008.

[9] Z. Liu and C. Liu, "Fusion of color, local spatial and global frequency information for face recognition," *Pattern Recognition*, vol. 43, no. 8, pp. 2882–2890, 2010.

[10] C. Liu, "Extracting discriminative color features for face recognition," *Pattern Recognition Letters*, vol. 32, no. 14, pp. 1796–1804, 2011.

[11] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color local texture features for color face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1366–1380, 2012.

[12] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[13] M. El Aroussi, M. El Hassouni, S. Ghouzali, M. Rziza, and D. Aboutajdine, "Local appearance based face recognition method using block based steerable pyramid transform," *Signal Processing*, vol. 91, no. 1, pp. 38–50, 2011.

[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Real-time learning capability of neural networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 863–878, 2006.

[15] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[16] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.

[17] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16-18, pp. 3056–3062, 2007.

[18] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42, no. 2, pp. 513–529, 2012.

[19] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognition*, vol. 45, no. 8, pp. 2868–2883, 2012.

[20] A. A. Mohammed, R. Minhas, Q. M. Jonathan Wu, and M. A. Sid-Ahmed, "Human face recognition based on multidimensional PCA and extreme learning machine," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2588–2597, 2011.

[21] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[22] A. Martinez and R. Benavente, "The AR face database," CVC Technical Report 24, Universitat Autònoma de Barcelona, Barcelona, Spain, 1998.

[23] A. Uçar, "Facial expression recognition based on significant face components using steerable pyramid transform," in *Proceedings of the International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV '13)*, vol. 2, pp. 687–692, Las Vegas, Nev, USA, July, 2013.

[24] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, Ny, USA, 1st edition, 1995.

[25] K. Jack, *Video Demystified: A Handbook for the Digital Engineer*, LLH Technology Publishing, Eagle Rock, Va, USA, 3rd edition, 2001.

[26] J. M. Chaves-González, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, and J. M. Sánchez-Pérez, "Detecting skin in face recognition systems: a colour spaces study," *Digital Signal Processing*, vol. 20, no. 3, pp. 806–823, 2010.

[27] S. V. Tathe and S. P. Narote, "Face detection using color models," *World Journal of Science and Technology*, vol. 2, no. 4, pp. 182–185, 2012.

[28] C. Garcia, G. Zikos, and G. Tziritas, "Face detection in color images using wavelet packet analysis," in *Proceedings of the IEEE Multimedia Computing and Systems (ICMCS '99)*, vol. I, pp. 703–708, Florence, Italy, June, 1999.

[29] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551–564, 1999.

[30] K.-W. Wong, K.-M. Lam, and W.-C. Siu, "A robust scheme for live detection of human faces in color images," *Signal Processing: Image Communication*, vol. 18, no. 2, pp. 103–114, 2003.

[31] S. M. Lajevardi and H. R. Wu, "Facial expression recognition in perceptual color space," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3721–3733, 2012.

[32] P. Shih and C. Liu, "Comparative assessment of content-based face image retrieval in different color spaces," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 873–893, 2005.

[33] A. Uçar, "Color face recognition based on curvelet transform," in *Proceedings of the International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV '12)*, vol. 2, pp. 561–566, Las Vegas, Nev, USA, July, 2012.

[34] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[35] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[36] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Tech. Rep., National Taiwan University, Department of Computer Science, Taipei, Taiwan, 2003, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[37] A. Uçar, Y. Demir, and C. Güzeliş, "A Penalty function method for designing efficient robust classifiers with input-space optimal separating surfaces," *Turkish Journal of Electrical Engineering and Compute Sciences*. In press.

[38] A. P. James and S. Dimitrijev, "Face recognition using local binary decisions," *IEEE Signal Processing Letters*, vol. 15, pp. 821–824, 2008.

[39] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Local color vector binary patterns from multichannel face images for face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2347–2353, 2012.

Research Article

Flood Detection/Monitoring Using Adjustable Histogram Equalization Technique

Fakhera Nazir,[1] Muhammad Mohsin Riaz,[2] Abdul Ghafoor,[2] and Fahim Arif[1]

[1] *Department of Computer Software Engineering, College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan*
[2] *Department of Electrical Engineering, College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan*

Correspondence should be addressed to Abdul Ghafoor; abdulghafoor-mcs@nust.edu.pk

Received 6 August 2013; Accepted 9 October 2013; Published 2 January 2014

Academic Editors: S. Bourennane and J. Marot

Flood monitoring technique using adjustable histogram equalization is proposed. The technique overcomes the limitations (overenhancement, artifacts, and unnatural look) of existing technique by adjusting the contrast of images. The proposed technique takes pre- and postimages and applies different processing steps for generating flood map without user interaction. The resultant flood maps can be used for flood monitoring and detection. Simulation results show that the proposed technique provides better output quality compared to the state of the art existing technique.

1. Introduction

Remote sensing technology has played an important role in flood monitoring in recent years. This development (optical/aerial to radar remote sensing) provides all weather capability as compared to the optical sensors for the purpose of flood mapping. Flood mapping [1–3] is one of the techniques used for flood monitoring in which pre- and postflood images are compared to classify undated (nonflooded) and inundated (flooded) areas.

Initially flood monitoring was limited to satellite [4] and aerial images [5]. However due to the development of radar remote sensing, the issue of limited performance in bad weather conditions (like clouds, lightening, etc.) [6] is resolved. The use of Synthetic Aperture Radar (SAR) imagery has solved the problem of flood monitoring due to its all weather capability [7]. Broadly the flood mapping techniques are divided into supervised (which requires operator involvement), semisupervised, and unsupervised techniques.

Some popular flood mapping techniques are visual interpretation [8], segmentation [9], thresholding [10], texture matching [11], and analysis of dynamic and physical characteristic of region of interest [12]. Visual interpretation [8] is the commonly used supervised approach for flood mapping. Besides consistent results of visual interpretation, user involvement is not always practically feasible.

Segmentation (semisupervised) technique [9] was proposed to minimize the involvement of user. The technique generates a connectivity map using fuzzy logic by selecting a seed point from user. Still it depends upon correct identification of seed point (chosen by user). A fast ready flood map (without user involvement) and a detailed flood map based on seed growing mechanism were proposed in [13] to overcome empirical settings. However, the detailed flood map still needs user ability to locate the points for segmentation.

Thresholding uses certain thresholds for unsupervised floods monitoring [10]. The thresholds are selected on the basis of the output of three electromagnetic scattering models to generate fast ready maps. However, these thresholds do not work under complex environmental conditions (in that case users involvement is required for reliable results) [14]. Moreover, a universal threshold cannot be not justified for flood detection [15].

Texture matching is also used to identify water areas from images [11]. United and homogenous regions of water are segmented; chromatic and texture features are then compared

to predefined samples of water. Main limitations are heavy computation time and overlapping texture features.

Different flood monitoring techniques are combined to generate inundation map [16]. The map represents the degree of belief for each pixel. However, reliable calibration and verification are not always possible. In [12], complex coherence map is used to perform an analysis on SAR data for flood monitoring and receding. This technique is not only confined to flood damage assessment but also other areas can be monitored (like earthquake). However, it depends upon the availability of optical images for observed events.

Recently, a chain of processing-based method [17] was proposed for better visual representation of an event. This method applied different processing chains (adaptive histogram clipping (AHC), histogram remapping (HR), and histogram equalization (HE)) to improve visualization. RGB image is then generated by combining the processed pre-, post- and difference images. The chain of events is performed to preserve the important information (in SAR image) [9]. However this process sometimes highlights extra details in the difference image which degrades the quality. Moreover the equalization process results in excessive contrast enhancement, which in turn gives the processed image an unnatural look.

To resolve these issues of [17], we propose a contrast enhancement-based technique to improve the visibility of resultant flood maps. The technique follows the three chains for processing proposed by Dellepiane and Angiati [17]. However, the chains are applied on the pre- and postimages for the generation of difference image only. A fast ready flood map is generated by combining the difference image with the equalized pre- and postimages. In contrast to Dellepiane and Angiati [17], we have used Adjustable HE (AHE) [18] with a low percentile value to improve the visibility. Results are evaluated using different data sets which show the significance of proposed technique.

2. Proposed Methodology

An AHE-based flood monitoring technique is proposed which is composed of three chains of processing. Let $I_{X(l,m)}$ be pre-, $I_{Y(l,m)}$ post-, and $I_{Z(l,m)}$ difference images, where $l \in [0, \ldots, L-1]$ and $m \in [0, \ldots, M-1]$.

The first step is AHC, which is used to preserve the information content. The histograms of (pre-, post-, and difference) images are clipped/truncated at a specific percentile value (to remove the intensities which contain much less amount of information). The histograms of preimage I_X are clipped using a specific percentile value q. Let $h_X(i)$ be the histogram of image I_X, where i represents the intensity levels [0–255]. The cumulative histogram $C_X(i)$ is

$$C_X(i) = \sum_{j=1}^{i} h_X(i). \tag{1}$$

The clipping is performed as

$$q = \arg\left[C_X(i) \mid C_X(i) = q \times M \times L\right]. \tag{2}$$

Dellepiane and Angiati [17] have used the same approach in which they used $q = 0.98$. However the issue in [17] is the excessive amount of details present in the final RGB map generated using clipped pre-, post- and difference image at proposed q percentile which finally contribute to flooding. To resolve this issue we used different q percentile value in the first step to generate the difference image. It is observed that at low percentile values required details are removed, whereas, at higher percentile values, unwanted details get more prominent, thus degrading the quality. Therefore, we have used $q = 0.30$ because it preserves the required intensity values which contribute to flooding.

In the second step (HR), the clipped histogram is remapped to the original intensity range using linear scaling. The histogram of image I_{X_1} (adaptive histogram clipped image I_X) is remapped to full intensity range [0–255]. Let $\min(I_{X_1})$ represent the minimum of all intensities and let $\max(I_X)$ represent the maximum of all intensities in the image. I_{X_2} (histogram remapped image I_{X_1}) is given as [17]

$$I_{X_2} = \frac{I_{X_1} - \min\left(I_{X_1}\right)}{\max\left(I_{X_1} - \min\left(I_{X_1}\right)\right)} \times 255. \tag{3}$$

In the third step (AHE), we use adaptive histogram equalization to enhance the image. Reference [17] uses traditional HE which sometimes overenhances the image and produces unwanted artifacts (roughness, etc.) of processed images (pre-, post- and difference). Furthermore, the processed images sometimes may not reveal all the details or merge the details which results in degradation of image quality. A contrast enhancement technique is required to maintain the smoothness and natural aspect of an image, for visually pleasing results. To achieve the proposed outcome, we use a new framework for histogram modification [18] to improve the visualization by preserving its details.

HE usually maps the input intensity levels i to the output level X_i according to

$$X_i = (N-1) \times C_{X(i)}, \tag{4}$$

where N represents the total intensity levels in image and $C_{X(i)}$ represents the cumulative histogram. This mapping is suitable for images with continues intensity levels where it perfectly equalizes the histogram. However, for digital images, traditional HE is not useful because of their discrete intensity levels [19].

In order to make it suitable for digital images, input histogram can be modified without compromising its contrast enhancement. The modified histogram can then be used as a mapping function for HE. The issues of HE are addressed by using the proposed Arici et al. [18] framework for histogram modification. The technique introduces specifically designed penalty terms which can be used to adjust the level of contrast enhancement.

Once the intensity range is remapped, AHE [18] is used to minimize the effects (like overenhancement, unusual artifacts, and unnatural look). The principle of AHE is to minimize the difference between modified h_{X_m} and current

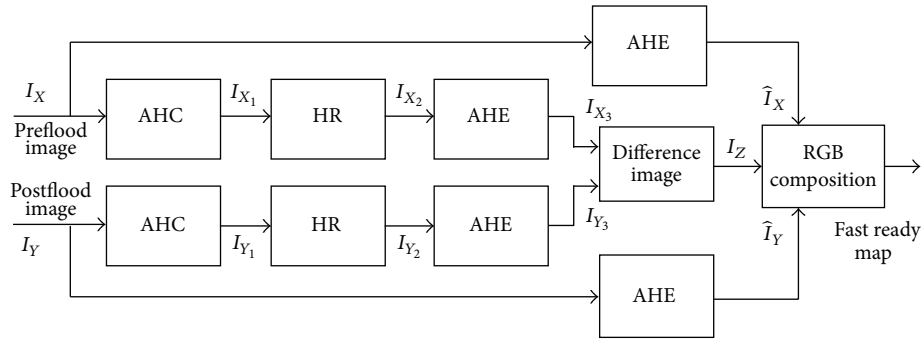
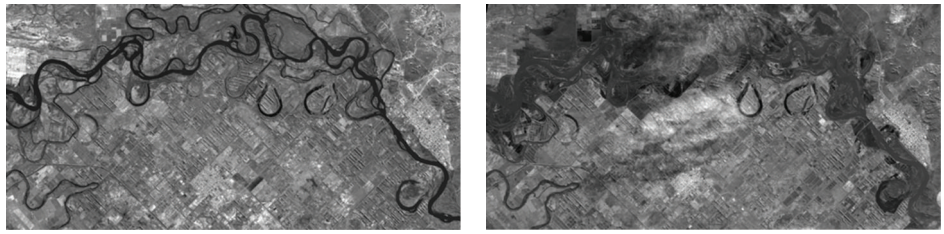


Figure 1: Flow chart of the proposed algorithm.



(a) (b)

Figure 2: Original images of Choele Choel City, Argentina, observed by "Daichi" (ALOS). (a) Preflooded image acquired on April 29, 2006. (b) Postflooded image acquired on July 30, 2006.

histogram h such that the modified histogram is also closer to the uniform histogram h_{X_u}; that is,

$$\min \left\| h - h_X \right\| + \alpha \left\| h - h_{X_u} \right\|, \tag{5}$$

where α is used to adjust the contribution of current and uniform histogram. The modified histogram by solving (5) [18] is

$$h_{X_m} = \left(\frac{1}{1+\alpha} \right) \times h_X + \left(\frac{\alpha}{1+\alpha} \right) \times h_{X_u}. \tag{6}$$

The modified histogram h_{X_m} is used to produce images I_{X_3} and I_{Y_3}.

Note that conventional histogram used in [17] produces unwanted artifacts, overenhancement, and unnatural look. This is due to the fact that the goal of traditional HE is to match the input histogram with uniform distribution. However, AHE also minimizes the difference between modified and input histogram (along with the input and uniform histogram). Hence, AHE produces more reliable results for flood monitoring.

Difference image I_Z is then generated using I_{X_3} and I_{Y_3} [17]:

$$I_Z(l,m) = 128 + \frac{I_{X_3}(l,m) - I_{Y_3}(l,m)}{2}. \tag{7}$$

Fast ready flood map is generated finally by combining adaptive histogram equalized pre- and postimages with the difference image. In [17] RGB map is generated by applying all chains of processing on pre- and postimages which are then combined with difference image. However, the processing of all images through same chains does not preserve intensity values in pre- and postimages. Hence, in our case, I_X and I_Y are only passed through the third chain of processing (AHE) to produce $\widehat{I}_X$ and $\widehat{I}_Y$, respectively.

Finally I_Z, $\widehat{I}_X$, and $\widehat{I}_Y$ are combined to generate fast ready map by assigning blue, green, and red bands to pre-, post-, and difference images, respectively. The level of red color is high for pixels whose prevalue dominates and vice versa. In RGB image, medium to dark red color represents permanent water like rivers and dark blue color represents the flooded areas.

The reason for using only third step AHE for RGB generation is to preserve intensity values of pre- and postimages that maintain the details. The purpose of using processed pre- and postflooded images for difference image generation is to remove the intensities which contribute very low in flooded areas.

Figure 1 shows the block diagram of proposed technique.

3. Simulation and Results

For evaluation of existing and proposed techniques, flood-occurring areas in Choele Choel City, Argentina, are considered. The images are observed by "Daichi," Advance land observing satellite on April 29 (preflooded image, shown in Figure 2(a)), and July 30, 2006 (postflooded image, shown in Figure 2(b)), respectively. Second data set includes the images of Tomakomai, Japan, acquired by Phased Array Type L-band SAR (PALSAR) using H/V polarization on August 19, 2006,

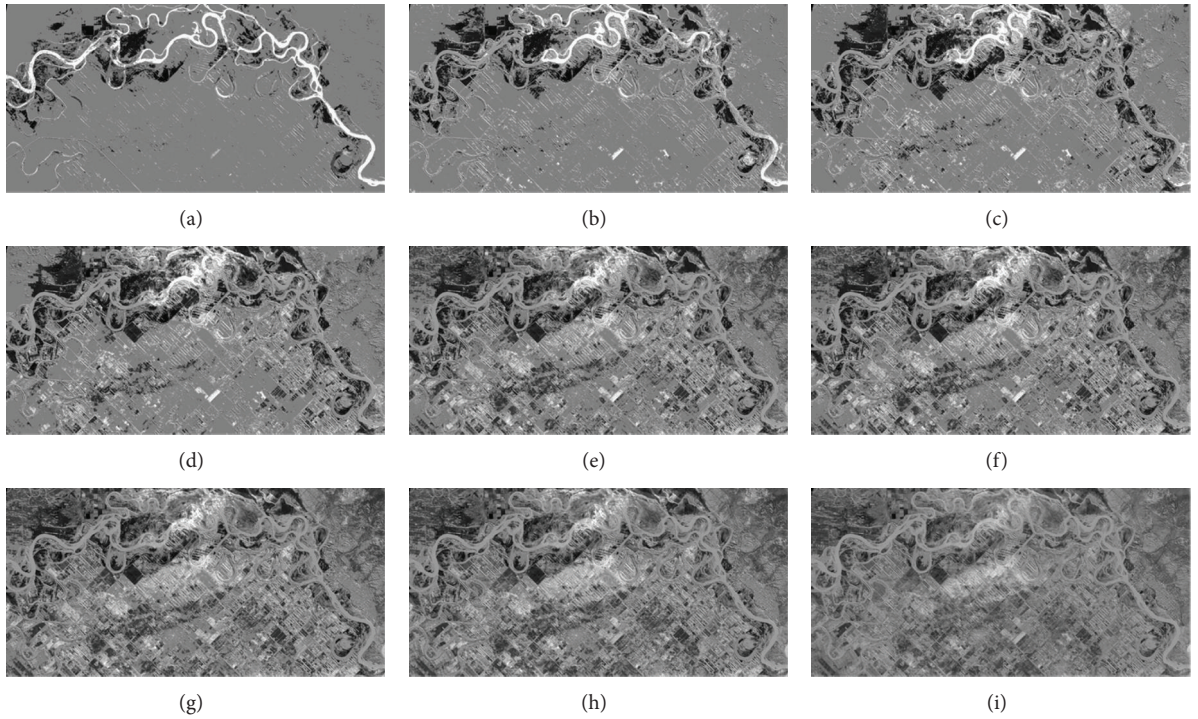


FIGURE 3: Difference images I_Z for different percentiles values: (a) at $q = 0.1$, (b) at $q = 0.2$, (c) at $q = 0.3$, (d) at $q = 0.4$, (e) at $q = 0.5$, (f) at $q = 0.6$, (g) at $q = 0.7$, (h) at $q = 0.8$, and (i) at $q = 0.98$.

in Figure 6(a) and V/V polarization on August 19, 2006, in Figure 6(b).

Figure 3 shows the variation in the difference image with respect to percentile value q. By increasing q, the details in the image increase (and vice versa). In Figure 3(a) we can notice that the ground area around the river is dim, which becomes quite visible in Figure 3(b) but the flood water is not so clear. At percentile value $q = 0.30$ (in Figure 3(c)), the ground area, permanent water, and flood are visible to the required level. As we move to higher percentile values ($q > 0.3$) ground area becomes more prominent gradually which contributes to the change area in final RGB composition. This effect can be observed in Figures 3(d)–3(i).

Figure 4 represents RGB images, generated by respective difference images (given in Figure 3). In Figures 4(a) and 4(b) the flooded area is dim, which fades away around the river. In Figure 4(c) (at $q = 0.30$), the flooded area around the river (at the top center of image) becomes quite visible to the acceptable level. The light ink blue area (at the bottom center of image) is reflecting the flooded pixels at the required level. For higher percentile values ($q > 0.3$) in Figures 4(d)–4(g), RGB images gradually increase the flooded areas at the bottom center of image (in dark ink blue color). The visibility of flooded areas on the top center of image is also not good. The results are quite obvious in Figures 4(h) and 4(i), where flooded areas are more faded around the river, but a lot of flooded areas are seen at the bottom center.

Figure 5 provides comparison of the proposed technique and Dellepiane and Angiati [17] technique. Figure 5(a) is a difference image generated using Dellepiane and Angiati [17] technique and Figure 5(b) is generated using the proposed methodology. There is a clear difference in details in these images. Figure 5(a) shows the ground details more prominently while Figure 5(b) highlights the major required details comparatively. These differences in details contribute a lot to their respective RGB (Figures 5(c) and 5(d)). We can notice the flooded area (in Figure 5(c)) around river (at the top center) is blur (not clear) which degrades visibility. A very high contribution of irrelevant details of difference image in RGB is visible (the blue color at the bottom center and dark blue color at the top right corner of image). Figure 5(d) shows better visibility of flooded area around river (at the top center), low blue color (at the top right corner), and low flooded areas (at the bottom center of image). One can clearly notice the difference in contrast/details of ground area and the contrast of river with flooded areas.

Figure 6(c) is the RGB map generated using Dellepiane and Angiati [17] technique. Although the image (in Figure 6(c)) is enhanced, it highlights the irrelevant details which contribute to flooding (see the blue colored areas at the right center of image). The details at the top right flooded area (in Figure 6(d)) are clear as compared to the flooded areas in Figure 6(c). Figure 6(c) mixes up the details due to overenhancement at the areas around the river while these areas are more clear in Figure 6(d) (in red colors). Figure 6(c) produces unnatural ground details; however more smoothness of image is seen in Figure 6(d) that preserves the natural effect of image to some extent.

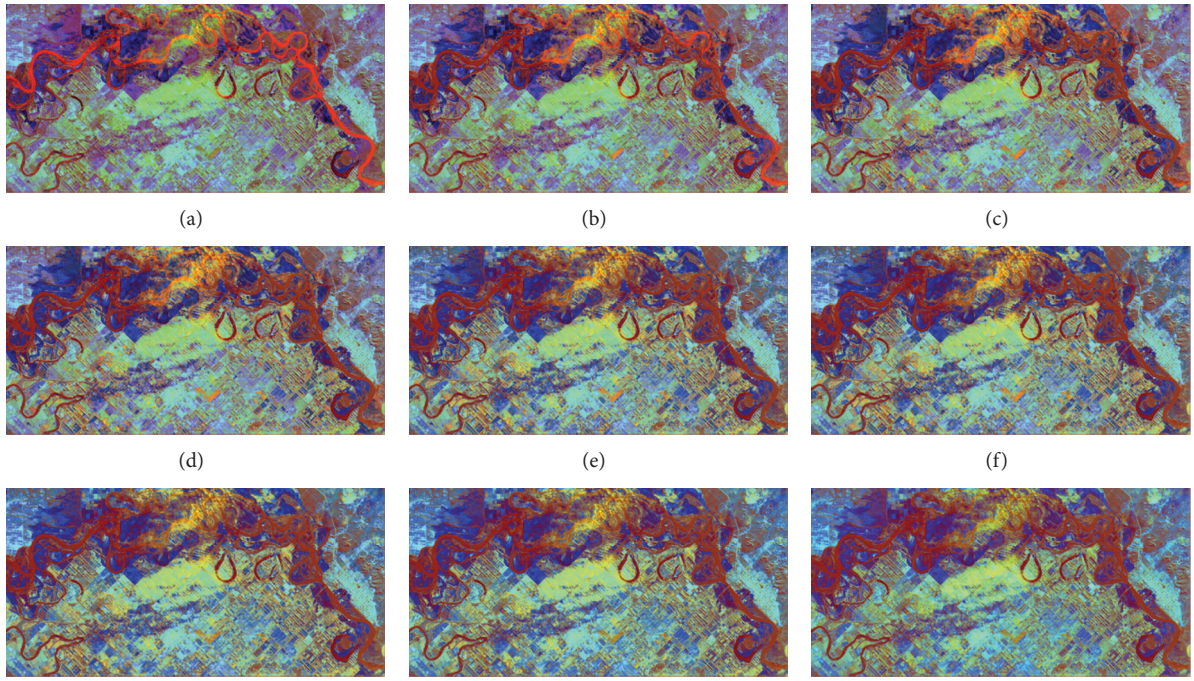


Figure 4: RGB images for different q percentiles values: (a) at $q = 0.1$, (b) at $q = 0.2$, (c) at $q = 0.3$, (d) at $q = 0.4$, (e) at $q = 0.5$, (f) at $q = 0.6$, (g) at $q = 0.7$, (h) at $q = 0.8$, and (i) at $q = 0.98$.

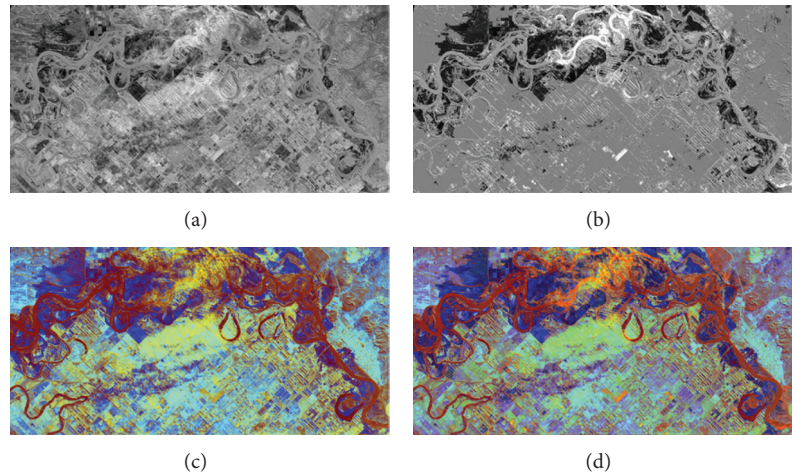
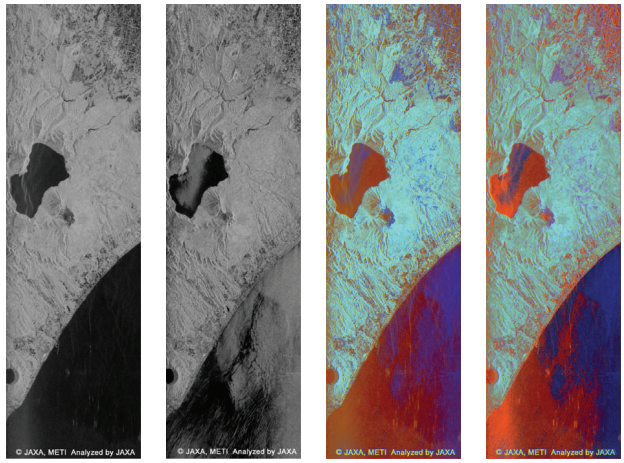


Figure 5: Evaluation of results using images of flood that occurred in Choele Choel City, Argentina. (a) Difference image obtained by Dellepiane and Angiati [17] approach. (b) Difference image obtained by proposed technique. (c) Fast ready map generated using Dellepiane and Angiati [17] approach. (d) Fast ready map generated using proposed technique.

4. Conclusion

A contrast enhancement-based flood mapping approach for SAR images is proposed which is composed of three steps (histogram adaptive clipping, remapping, and adjustable histogram equalization). Pre- and postflooded images are processed using different processing chains and the difference image is produced (by pre- and postimages). A fast ready flood map is then generated, using the combination of processed pre- and postimages (only the third step is applied) with difference images. A specific contrast enhancement technique AHE is used as a third step to remove the overenhancement produced by HE. The proposed technique is an improvement in existing state of the art, which suffers from unwanted details, unnatural look, and overenhancement of the image. The technique produces visually pleasing



(a) (b) (c) (d)

Figure 6: Evaluation of results using images of Tomakomai, Japan. (a) Preimage acquired on 19 August 2006. (b) Postimage acquired on 19 August 2006. (c) Fast ready map generated using Dellepiane and Angiati [17] approach. (d) Fast ready map generated using proposed technique.

results by suppressing the irrelevant details and minimizing overenhancement, thus maintaining quality. Simulation results show the significance of proposed technique.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this article.

References

[1] S. Kuehn, U. Benz, and J. Hurley, "Efficient flood monitoring based on RADARSAT-1 images data and information fusion with object-oriented technology," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS '02)*, vol. 5, pp. 2862–2864, June 2002.

[2] F. Nazir, M. M. Riaz, A. Ghafoor, and F. Arif, "Contrast enhancement based flood monitoring," in *IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Okinawa, Japan, November 2013.

[3] N. Kussul, A. Shelestov, and S. Skakun, "Flood monitoring from SAR data," in *Use of Satellite and In-Situ Data to Improve Sustainability*, NATO Science for Peace and Security Series C: Environmental Security, pp. 19–29, 2011.

[4] H. Rasid and M. A. H. Pramanik, "Areal extent of the 1988 flood in Bangladesh: how much did the satellite imagery show?" *Natural Hazards*, vol. 8, no. 2, pp. 189–200, 1993.

[5] N. M. Robertson and T. Chan, "Aerial image segmentation for flood risk analysis," in *IEEE International Conference on Image Processing (ICIP '09)*, pp. 597–600, November 2009.

[6] R. T. Lowry, E. J. Langham, and N. Murdy, "A preliminary analysis of SAR mapping of Manitoba flood," in *Satellite Hydrology, Fifth Anniversary. William T. Pecora Memorial Symposium on Remote Sensing*, pp. 316–323, 1981.

[7] J. Sanyal and X. X. Lu, "Application of remote sensing in flood management with special reference to monsoon Asia: a review," *Natural Hazards*, vol. 33, no. 2, pp. 283–301, 2004.

[8] Y. Chambenoit, N. Classeau, E. Trouvé, and J.-P. Rudant, "Performance assessment of multitemporal SAR images' visual interpretation," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 6, pp. 3911–3913, Toulouse, France, July 2003.

[9] F. Giordano, M. Goccia, and S. Dellepiane, "Segmentation of coherence maps for flood damage assessment," in *IEEE International Conference on Image Processing (ICIP '05)*, vol. 2, pp. 233–236, September 2005.

[10] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2972–2982, 2006.

[11] M. Zhao, H. Shang, W. Huang, L. Zou, and Y. Zhang, "Flood area extraction from rgb aerophotograph based on chromatic and textural analysis," in *International Conference on Advanced Geographic Information Systems, Appliations and Services Geo-Processing*, pp. 46–52, 2011.

[12] M. Chini, L. Pulvirenti, and N. Pierdicca, "Analysis and interpretation of the COSMO-SkyMed observations of the 2011 Japan tsunami," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 467–471, 2012.

[13] S. Dellepiane, E. Angiati, and G. Vernazza, "Processing and segmentation of COSMO-SkyMed images for flood monitoring," in *Proceedings of the 30th IEEE International Geoscience and Remote Sensing Symposium (IGARSS '10)*, pp. 4807–4810, July 2010.

[14] L. Pulvirenti, N. Pierdicca, M. Chini, and L. Guerriero, "An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic," *Natural Hazards and Earth System Science*, vol. 11, no. 2, pp. 529–540, 2011.

[15] Y.-Q. Jin, "A flooding index and its regional threshold value for monitoring floods in China from SSM/I data," *International Journal of Remote Sensing*, vol. 20, no. 5, pp. 1025–1030, 1999.

[16] G. Schumann, G. Di Baldassarre, and P. D. Bates, "The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2801–2807, 2009.

[17] S. G. Dellepiane and E. Angiati, "A new method for cross-normalization and multitemporal visualization of SAR images for the detection of flooded areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2765–2779, 2012.

[18] T. Arici, S. Dikbas, and A. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1921–1935, 2009.

[19] Q. Wang and R. K. Ward, "Fast image/video contrast enhancement based on weighted thresholded histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 757–764, 2007.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 896328, 14 pages
http://dx.doi.org/10.1155/2013/896328

Research Article

Recognition of the Script in Serbian Documents Using Frequency Occurrence and Co-Occurrence Analysis

Darko Brodić,[1] Zoran N. Milivojević,[2] and Čedomir A. Maluckov[1]

[1] *Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia*
[2] *Technical College Niš, Aleksandra Medvedeva 20, 18000 Niš, Serbia*

Correspondence should be addressed to Darko Brodić; dbrodic@tf.bor.ac.rs

Received 18 July 2013; Accepted 21 August 2013

Academic Editors: S. Bourennane, C. Fossati, and J. Marot

Any document in Serbian language can be written in two different scripts: Latin or Cyrillic. Although characteristics of these scripts are similar, some of their statistical measures are quite different. The paper proposed a method for the extraction of certain script from document according to the occurrence and co-occurrence of the script types. First, each letter is modeled with the certain script type according to characteristics concerning its position in baseline area. Then, the frequency analysis of the script types occurrence is performed. Due to diversity of Latin and Cyrillic script, the occurrence of modeled letters shows substantial statistics dissimilarity. Furthermore, the co-occurrence matrix is computed. The analysis of the co-occurrence matrix draws a strong margin as a criteria to distinguish and recognize the certain script. The proposed method is analyzed on the case of a database which includes different types of printed and web documents. The experiments gave encouraging results.

1. Introduction

Cryptography studies the problems concerning the conversion of information from a readable to some other state. It deals with information which is changing from one to another state. The initial information represents a plain text. When the information becomes encrypted, it is referred as a cipher text [1]. A substitution cipher is a method of encoding. According to it, the units of plain text are replaced with cipher text [2]. They can be single letters, pairs of letters, triplets of letters, mixtures of the above, and so forth. In our application the encryption function is not needed to be injective [3] due to nature of further statistical analysis. It does not matter if it will encrypt two different plain texts into the same cipher text, because decryption of the cipher text is not considered. Hence, the cryptography is used only as a basis for modeling and analyzing documents written in Serbian language. Serbian language represents the European minority language. However, it is distinct due to its capability to be written in Latin and Cyrillic script, interchangeably. According to the baseline characteristics [4], each letter in the text file is replaced with the cipher which is taken from the set of four counterparts only. The basic idea is to distinguish the script (Latin or Cyrillic) according to statistical analysis of the cipher text. It is accomplished with frequency analysis concerning occurrence [5] as well as with the method using statistical measures extracted from gray-level co-occurrence matrix [6]. The letter frequency distribution is a function which assigns each letter a frequency of its occurrence in the text sample [7]. The gray-level co-occurrence matrix (GLCM) have used for the extraction of features needed for texture classification [8]. Nevertheless, it can be exploited for a letter co-occurrence in a text document [9]. At the final stage, the experiment is made on a custom oriented database containing text from printed and Web documents.

The rest of the paper is organized as follows. Section 2 describes the full procedure of the proposed algorithm. Section 3 defines the experiment. Section 4 presents the results from experiment and discusses them. Section 5 makes a conclusion.

2. Proposed Algorithm

The proposed algorithm converts document written in Latin and Cyrillic script which represent the plain text into cipher

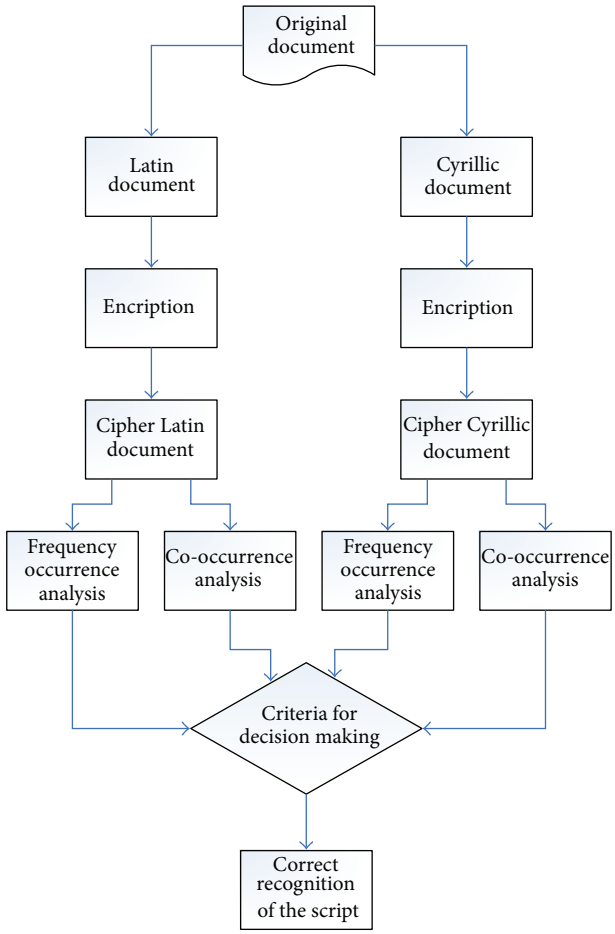


Figure 1: The flow of proposed algorithm.

text according to predefined encryption based on text line structure definition. Then, the equivalent cipher texts are subjected to the frequency and co-occurrence analysis. The results of frequency analysis indicated a substantial difference between cipher texts obtained from Latin and Cyrillic text. Similarly, co-occurrence analysis shows obvious quantitative disparity in some measures. This draws a strong margin as a criterion in order to distinguish and recognize a certain script type (Figure 1).

2.1. Text Line Structure. Text in printed and Web documents is defined as well-formed text type. It is characterized by strong regularity in shape. The distances between the text lines are adequate to be split up. The words are formed regularly with similar distance. Their inter word spacing is decent as well. However, in certain script, the letters or signs have different position according to its baseline. It is shown in Figure 2.

From Figure 2 four virtual lines can be defined [4]:

(i) The top-line,

(ii) The upper-line,

(iii) The base-line, and

(iv) The bottom-line.

Table 1: Definition of script types according to the baseline characteristics.

Script example	Type of script	Designation
a	Short	S
b	Ascender	A
j	Descender	D
lj	Full	F

Accordingly, a text line can be considered as being composed of three vertical zones [4]:

(i) The upper zone,

(ii) The middle zone, and

(iii) The lower zone.

Each text line has at least a middle zone. The upper zone depends on capital letters and letters with ascenders, while the lower zone depends on letters with descenders. Only a few letters occupy the upper and lower zone.

2.2. Encryption. Two different sets are produced. They are A_L and A_C for Latin and Cyrillic alphabet, respectively:

$$A_L = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots, \check{\mathrm{Z}}, \mathrm{a}, \mathrm{b}, \mathrm{c}, \ldots, \check{\mathrm{z}}\},$$
$$A_C = \{\text{А}, \text{Б}, \text{Ц}, \ldots, \text{Ш}, \text{а}, \text{б}, \text{ц}, \ldots, \text{ш}\}. \quad (1)$$

Each of them consists of 60 elements that is, letters, which are valid for Serbian language. Furthermore, both sets A_L and A_C are mapped into set C.

$$f_L : A_L \longmapsto C,$$
$$f_C : A_C \longmapsto C. \quad (2)$$

These mappings are achieved in accordance with the text line area definition. The structure of text line allows definition of following script types [4].

(i) Full letter (F), where letter is present in all three zones.

(ii) Ascender letter (A), where character parts are present in the upper and middle zones.

(iii) Descender letter (D), where character parts are present in the lower and middle zones, and

(iv) Short letter (S), where character parts are present in the middle zone only.

Accordingly, all letters will be replaced with the cipher from the following set:

$$C = \{\mathrm{S}, \mathrm{A}, \mathrm{D}, \mathrm{F}\}. \quad (3)$$

All letters can reach certain position, which belongs to set C with a unique designation according to Table 1.

It should be noted that above mappings are surjective.

Serbian language contains 30 letters. Each letter in Latin has a corresponding equivalent letter in Cyrillic. Table 2 shows Latin and Cyrillic letters as well as theirs designation according to Table 1.

Statistical analysis of the letters and theirs corresponding type for Latin and Cyrillic scripts is shown in Table 3.

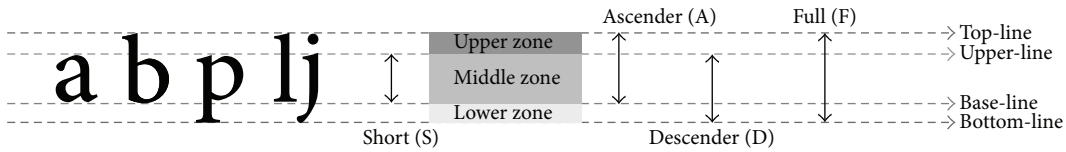


Figure 2: Definitions of the script characteristics.

Table 2: Serbian Latin and equivalent Cyrillic alphabet according to the script types.

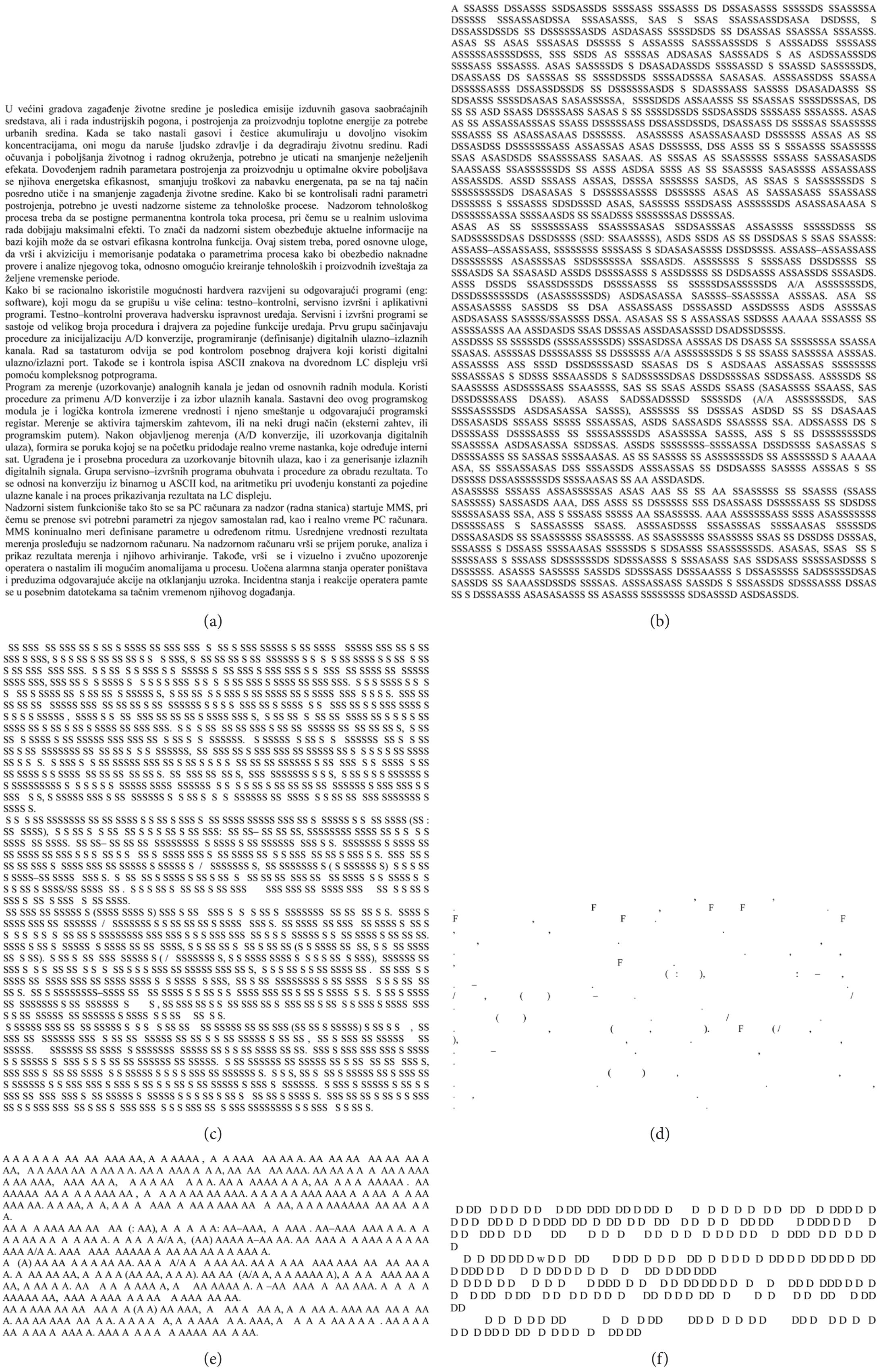
Alphabet	Latin	Script types	Latin	Script types	Cyrillic	Script types	Cyrillic	Script types
1	A	A	a	S	А	A	а	S
2	B	A	b	A	Б	A	б	A
3	C	A	c	S	Ц	F	ц	D
4	Ć	A	ć	A	Ћ	A	ћ	A
5	Č	A	č	A	Ч	A	ч	S
6	D	A	d	A	Д	F	д	S
7	Đ	A	đ	A	Ђ	A	ђ	F
8	Dž	A	dž	A	Џ	F	џ	D
9	E	A	e	S	Е	A	е	S
10	F	A	f	A	Ф	A	ф	F
11	G	A	g	D	Г	A	г	S
12	H	A	h	A	Х	A	х	S
13	I	A	i	S	И	A	и	S
14	J	A	j	D	Ј	A	ј	D
15	K	A	k	A	К	A	к	S
16	L	A	l	A	Л	A	л	S
17	Lj	F	lj	F	Љ	A	љ	S
18	M	A	m	S	М	A	м	S
19	N	A	n	S	Н	A	н	S
20	Nj	F	nj	D	Њ	A	њ	S
21	O	A	o	S	О	A	о	S
22	P	A	p	D	П	A	п	S
23	R	A	r	S	Р	A	р	D
24	S	A	s	S	С	A	с	S
25	Š	A	š	A	Ш	A	ш	S
26	T	A	t	A	Т	A	т	S
27	U	A	u	S	У	A	у	D
28	V	A	v	S	В	A	в	S
29	Z	A	z	S	З	A	з	S
30	Ž	A	ž	A	Ж	A	ж	S

2.3. Frequency Analysis of the Occurrence. In the proposed algorithm, all letters from certain script has been substituted with equivalent members of the set C according to Table 2. These circumstances for Latin document are shown in Figure 3.

Figure 3(b) shows the cipher text which is obtained from Latin documents according to modeling given in Table 2. Figures 3(c)–3(f) shows a subset of cipher text with each element of set C, that is, S, A, D and F, respectively. Statistical analysis of the cipher text shows following: 2217 elements of S, 598 elements of A, 261 elements of D and 8 elements of F types. Accordingly, distribution of C set elements for Latin document is shown in Figure 4.

Currently, the same Latin document is converted into Cyrillic one. Similarly as in Latin document, all letters from Cyrillic document are exchanged with the equivalent members of the set C according to Table 2. These circumstances for Cyrillic document are shown in Figure 5.

Figure 5(b) shows the cipher text which is obtained from Cyrillic documents according to modeling given in Table 2. Figures 5(c)–5(f) shows a subset of cipher text with each element of set C, that is, S, A, D and F, respectively.



U većini gradova zagađenje životne sredine je posledica emisije izduvnih gasova saobraćajnih sredstava, ali i rada industrijskih pogona, i postrojenja za proizvodnju toplotne energije za potrebe urbanih sredina. Kada se tako nastali gasovi i čestice akumuliraju u dovoljno visokim koncentracijama, oni mogu da naruše ljudsko zdravlje i da degradiraju životnu sredinu. Radi očuvanja i poboljšanja životnog i radnog okruženja, potrebno je uticati na smanjenje neželjenih efekata. Dovođenjem radnih parametara postrojenja za proizvodnju u optimalne okvire poboljšava se njihova energetska efikasnost, smanjuju troškovi za nabavku energenata, pa se na taj način posredno utiče i na smanjenje zagađenja životne sredine. Kako bi se kontrolisali radni parametri postrojenja, potrebno je uvesti nadzorne sisteme za tehnološke procese. Nadzorom tehnološkog procesa treba da se postigne permanentna kontrola toka procesa, pri čemu se u realnim uslovima rada dobijaju maksimalni efekti. To znači da nadzorni sistem obezbeđuje aktuelne informacije na bazi kojih može da se ostvari efikasna kontrolna funkcija. Ovaj sistem treba, pored osnovne uloge, da vrši i akviziciju i memorisanje podataka o parametrima procesa kako bi obezbedio naknadne provere i analize njegovog toka, odnosno omogućio kreiranje tehnoloških i proizvodnih izveštaja za željene vremenske periode.

Kako bi se racionalno iskoristile mogućnosti hardvera razvijeni su odgovarajući programi (eng: software), koji mogu da se grupišu u više celina: testno–kontrolni, servisno izvršni i aplikativni programi. Testno–kontrolni proverava hadversku ispravnost uređaja. Servisni i izvršni programi se sastoje od velikog broja procedura i drajvera za pojedine funkcije uređaja. Prvu grupu sačinjavaju procedure za inicijalizaciju A/D konverzije, programiranje (definisanje) digitalnih ulazno–izlaznih kanala. Rad sa tastaturom odvija se pod kontrolom posebnog drajvera koji koristi digitalni ulazno/izlazni port. Takođe se i kontrola ispisa ASCII znakova na dvorednom LC displeju vrši pomoću kompleksnog potprograma.

Program za merenje (uzorkovanje) analognih kanala je jedan od osnovnih radnih modula. Koristi procedure za primenu A/D konverzije i za izbor ulaznih kanala. Sastavni deo ovog programskog modula je i logička kontrola izmerene vrednosti i njeno smeštanje u odgovarajući programski registar. Merenje se aktivira tajmerskim zahtevom, ili na neki drugi način (eksterni zahtev, ili programskim putem). Nakon objavljenog merenja (A/D konverzije, ili uzorkovanja digitalnih ulaza), formira se poruka kojoj se na početku pridodaje realno vreme nastanka, koje određuje interni sat. Ugrađena je i prosebna procedura za uzorkovanje bitovnih ulaza, kao i za generisanje izlaznih digitalnih signala. Grupa servisno–izvršnih programa obuhvata i procedure za obradu rezultata. To se odnosi na konverziju iz binarnog u ASCII kod, na aritmetiku pri uvođenju konstanti za pojedine ulazne kanale i na proces prikazivanja rezultata na LC displeju.

Nadzorni sistem funkcioniše tako što se sa PC računara za nadzor (radna stanica) startuje MMS, pri čemu se prenose svi potrebni parametri za njegov samostalan rad, kao i realno vreme PC računara. MMS koninualno meri definisane parametre u određenom ritmu. Usrednjene vrednosti rezultata merenja prosleđuju se nadzornom računaru. Na nadzornom računaru vrši se prijem poruke, analiza i prikaz rezultata merenja i njihovo arhiviranje. Takođe, vrši se i vizuelno i zvučno upozorenje operatera o nastalim ili mogućim anomalijama u procesu. Uočena alarmna stanja operater poništava i preduzima odgovarajuće akcije na otklanjanju uzroka. Incidentna stanja i reakcije operatera pamte se u posebnim datotekama sa tačnim vremenom njihovog događanja.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 3: Application of the proposed algorithm: (a) Original Latin text, (b) Cipher text according to set C, (c) Only "S" text, (d) Only "F" text, (e) Only "A" text, (f) Only "D" text.

Table 3: Statistical analysis of Latin and Cyrillic script types.

Script	Type of letters	Occurrence of script types				Distribution of script types			
		S	A	D	F	S (%)	A (%)	D (%)	F (%)
Latin	Capital letters	0	28	0	2	0	93.33	0	6.67
Latin	Small letters	12	13	4	1	40.00	43.33	13.33	3.34
Cyrillic	Capital letters	0	27	0	3	0	90.00	0	10.00
Cyrillic	Small letters	21	2	5	2	70.00	6.67	16.66	6.67

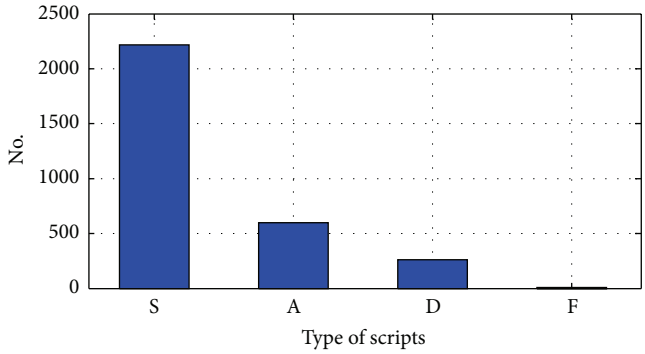


Figure 4: Distribution of C set's elements in cipher text obtained from Latin document.

Statistical analysis of the Cyrillic document image shows following: 2516 elements of S, 53 elements of A, 445 elements of D and 26 elements of F types. It should be noted that the sum of all C set elements in Latin and Cyrillic document is not quite identical. It is valid due to difference in definition of letters in two scripts. In the Cyrillic script, each letter is given one and only one sign. However, in Latin script letters such as dž, lj and nj are represented by two letters. Distribution of C set elements for Cyrillic document is presented in Figure 6.

According to Figures 4 and 6, the comparison chart is drawn. It is shown below in Figure 7.

Quantification of the script type appearance in a document written in Latin and Cyrillic is shown in Table 4.

It is obvious that the Latin document compared to Cyrillic one has slightly smaller number of short (S), descender (D) and full (F) letters. Nonetheless, the crucial margin is seen in ascender (A) letters. Hence, it can be a measure of confidence for detection of the script in a document given in Serbian language.

2.4. Co-Occurrence Analysis. Let **I** be the gray scale image which is under consideration. It has M row and N columns, while T is the total number of gray levels. The spatial relationship of gray levels in the image **I** is expressed by the grayscale co-occurrence matrix (GLCM) **C** [6, 10]. Hence, **C** is a matrix that describes the frequency of one gray level appearing in a specified spatial linear relationship with another gray level within the area of investigation [11]. In order to compute a co-occurrence matrix **C**, we considered a central pixel $I(x, y)$ with a neighborhood defined by the window of interest. This window is defined by two parameters: inter-pixel distance (d) and orientation (θ). Typically, the choice of d is 1 (one pixel), while the value of θ depends on the neighborhood. Because of that, each pixel has 8 neighbors given at following angles $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$. However, the case of neighbors at $\theta = 0^\circ$ or at $\theta = 180^\circ$ is similar to the GLCM definition [12]. So, the choice may fall to 4 neighbors pixels at $\theta = 0^\circ, 45^\circ, 90^\circ$ and 135°, that is, horizontal, right diagonal, vertical and left diagonal [13]. These orientations refer to 4-adjacent pixels at $(x+d, y)$, $(x, y-d)$, $(x-d, y)$ and $(x, y+d)$, where d is 1. For each pixel of the neighborhood, it is counted the number of times a pixel pair appears specified by the distance, and orientation parameters. The (i, j)th entry of **C** represents the number of occasions a pixel with an intensity i is adjacent to a pixel with an intensity j. Hence, for the given image **I**, the co-occurrence matrix **C** is defined as [14]:

$$C(i,j) = \sum_{x=1}^{M} \sum_{y=1}^{N} \begin{cases} 1, & \text{if } I(x,y) = i, \\ & I(x+\Delta x, y+\Delta y) = j \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where i and j are the image intensity values of the image, x and y are the spatial positions in the image **I**. The offset $(\Delta x, \Delta y)$ is specifying the distance between the pixel-of-interest and its neighbor. It depends on the direction θ that is used and the distance d at which the matrix is computed. The square matrix **C** is of the order N. Using a statistical approach like GLCM provides a valuable information about the relative position of the neighboring pixels in an image [12]. In order to normalize matrix **C**, matrix **P** is calculated as [10]:

$$P(i,j) = \frac{C(i,j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} C(i,j)}. \quad (5)$$

The normalized co-occurrence matrix **P** is obtained by dividing each element of **C** by the total number of co-occurrence pairs in **C**.

To illustrate the computing of GLCM, a four gray level image **I** is used. The window parameters are $d = 1$ and $\theta = 0^\circ$ (horizontal). Initial matrix **I** is shown in Figure 8.

The procedure of calculating co-occurrence matrix for grayscale matrix **I** ($d = 1$ and $\theta = 0^\circ$) [12] is given in Figure 9.

In order to GLCM be applied in our case, set C is mapped into set C_N by bijective function as:

$$f_C : C \longmapsto C_N, \quad (6)$$

where $C_N = \{0, 1, 2, 3\}$. Furthermore, the neighborhood is given as 2-connected ($x - d$ and $x + d$ around x, where $d = 1$). According to that, the same document in Latin and Cyrillic script is converted into cipher text. It is shown below in Figure 10.

У већини градова загађење животне средине је последица емисије издувних гасова саобраћајних средстава, али и рада индустријских погона, и постројења за производњу топлотне енергије за потребе урбаних средина. Када се тако настали гасови и честице акумулирају у довољно високим концентрацијама, они могу да наруше људско здравље и да деградирају животну средину. Ради очувања и побољшања животног и радног окружења, потребно је утицати на смањење нежељених ефеката. Довођењем радних параметара постројења за производњу у оптималне оквире побољшава се њихова енергетска ефикасност, смањују трошкови за набавку енергената, па се на тај начин посредно утиче и на смањење загађења животне средине. Како би се контролисали радни параметри постројења, потребно је увести надзорне системе за технолошке процесе. Надзором технолошког процеса треба да се постигне перманентна контрола тока процеса, при чему се у реалним условима рада добијају максимални ефекти. То значи да надзорни систем обезбеђује актуелне информације на бази којих може да се оствари ефикасна контролна функција. Овај систем треба, поред основне улоге, да врши и аквизицију и меморисање података о параметрима процеса како би обезбедио накнадне провере и анализе његовог тока, односно омогућио креирање технолошких и производних извештаја за жељене временске периоде.

Како би се рационално искористиле могућности хардвера развијени су одговарајући програми (енг: софтwаре), који могу да се групишу у више целина: тестно–контролни, сервисно извршни и апликативни програми. Тестно–контролни проверава хадверску исправност уређаја. Сервисни и извршни програми се састоје од великог броја процедура и драјвера за поједине функције уређаја. Прву групу сачињавају процедуре за иницијализацију А/Д конверзије, програмирање (дефинисање) дигиталних улазно–излазних канала. Рад са тастатуром одвија се под контролом посебног драјвера који користи дигитални улазно/излазни порт. Такође се и контрола исписа АСЦИИ знакова на дворедном ЛЦ дисплеју врши помоћу комплексног потпрограма.

Програм за мерење (узорковање) аналогних канала је један од основних радних модула. Користи процедуре за примену А/Д конверзије и за избор улазних канала. Саставни део овог програмског модула је и логичка контрола измерене вредности и њено смештање у одговарајући програмски регистар. Мерење се активира тајмерским захтевом, или на неки други начин (екстерни захтев, или програмским путем). Након објављеног мерења (А/Д конверзије, или узорковања дигиталних улаза), формира се порука којој се на почетку придодаје реално време настанка, које одређује интерни сат. Уграђена је и просебна процедура за узорковање битовних улаза, као и за генерисање излазних дигиталних сигнала. Група сервисно–извршних програма обухвата и процедуре за обраду резултата. То се односи на конверзију из бинарног у АСЦИИ код, на аритметику при увођењу константи за поједине улазне канале и на процес приказивања резултата на ЛЦ дисплеју.

Надзорни систем функционише тако што се са ПЦ рачунара за надзор (радна станица) стартује ММС, при чему се преносе сви потребни параметри за његов самосталан рад, као и реално време ПЦ рачунара. ММС конинуално мери дефинисане параметре у одређеном ритму. Усредњене вредности резултата мерења прослеђују се надзорном рачунару. На надзорном рачунару врши се пријем поруке, анализа и приказ резултата мерења и њихово архивирање. Такође, врши се и визуелно и звучно упозорење оператера о насталим или могућим аномалијама у процесу. Уочена алармна стања оператер поништава и предузима одговарајуће акције на отклањању узрока. Инцидентна стања и реакције оператера памте се у посебним датотекама са тачним временом њиховог догађања.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 5: Application of the proposed algorithm: (a) Original Cyrillic text, (b) Cipher text according to set C, (c) Only "S" text, (d) Only "F" text, (e) Only "A" text, (f) Only "D" text.

Table 4: Percentage of script type occurrence in document.

Type of script (TOS)	Latin	Cyrillic	x times
S	71.88%	82.76%	0.87
A	19.39%	1.74%	11.14
D	8.46%	14.64%	0.57
F	0.27%	0.86%	0.31

It is obvious that the Latin document compared to Cyrillic one has slightly smaller number of short (S), descender (D), and full (F) letters. Nonetheless, the crucial margin is seen in ascender (A) letters. Hence, it can be a measure of confidence for detection of the script in a document given in Serbian language.

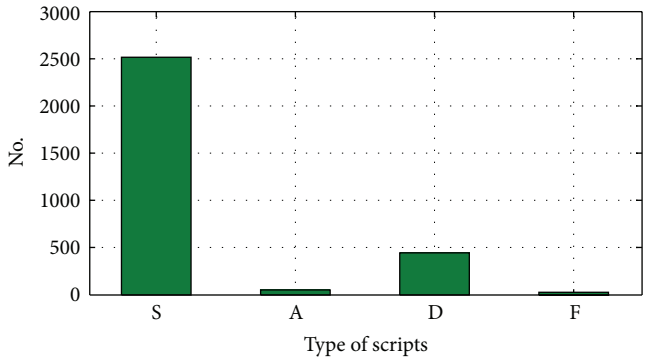


Figure 6: Distribution of C sets elements in cipher text obtained from Cyrillic document.

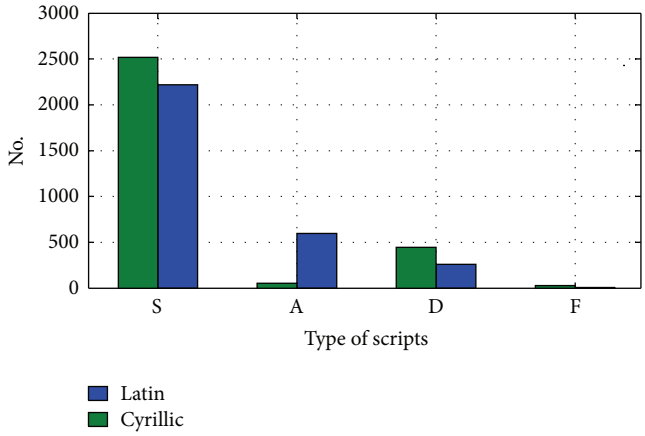


Figure 7: Comparison between distributions of C set elements in Latin and equivalent Cyrillic document.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Figure 8: Initial 4-level grayscale matrix $\mathbf{I}$ (featuring $M = 7$, $N = 4$, and $L = 4$).

Table 5: Normalized cooccurrence matrix.

(a) For cipher text obtained from the Latin text

0.3722	0.2212	0.0623	0.0048
0.2220	0.0343	0.0072	0
0.0623	0.0072	0.0016	0
0.0048	0	0	0

(b) For cipher text obtained from Cyrillic text

0.5863	0.0327	0.1326	0.0064
0.0391	0.0104	0.0144	0
0.1262	0.0200	0.0224	0.0016
0.0072	0	0.0008	0

To evaluate these cipher documents GLCM method is employed. Nevertheless, various statistic measures obtained from the co-occurrence matrix is introduced. The primary goal is to characterize the cipher text. Five descriptors can be used to describe the image [15]:

(i) Uniformity (UNI),

(ii) Entropy (ENT),

(iii) Maximum probability (MAX),

(iv) Dissimilarity (DIS), and

(v) Contrast (CON).

Uniformity (UNI) which is sometimes called angular second moment (ASM) or energy (ENG) measures the image homogeneity. It receives the highest value when GLCM has few entries of large magnitude. In contrast, it is low when all entries are nearly equal. The equation of the uniformity is [15]:

$$\text{UNI} = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)^2. \tag{7}$$

Entropy (ENT) measures the disorder or the complexity of the image. The highest value is found when the values of $P(i, j)$ are allocated quite uniformly throughout the matrix. This happens when the image has no pairs of gray level, with particular preference over others. The equation of the entropy is [15, 16]:

$$\text{ENT} = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) \cdot \log P(i,j). \tag{8}$$

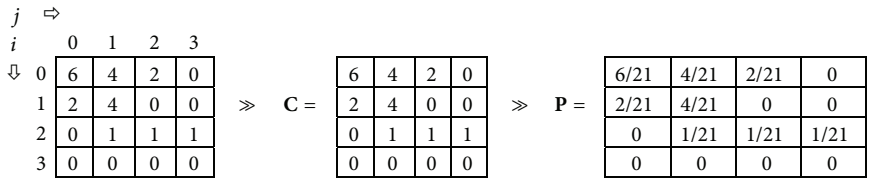


Figure 9: Co-occurrence matrix for grayscale matrix $\mathbf{I}$ ($d = 1$ and $\theta = 0^{\circ}$): (a) The number of occasions a pixel with an intensity I is adjacent to a pixel with intensity j, (b) Co-occurrence matrix $\mathbf{C}$, (c) Normalized matrix $\mathbf{P}$.


Недопустива лакоћа мењања прошлости
Данас незаобилазна „култура сећања" у земљи Србији не наилази увек на плодно тло, поготово када неодвојива баштина те културе треба да буде истина о антифашистичкој борби српског народа.
Мада одавно није у календару државних празника, 7. јули, некада прослављан као Дан устанка у Србији, данас ће ипак бити обележен државном манифестацијом у Белој Цркви код Крупња, код споменика народном хероју Живкин Јовановићу Шпанцу. Одлуку да ова манифестација буде одржана донео је Одбор за неговање традиција ослободилачких ратова Србије, а Влада Србије усвојила је ту одлуку.
Можда би се, помало оптимистички, могло рећи да је то сигнал који указује да су коначно за нама ужарене, страсне расправе о томе када је и како започет устанак у окупираној Србији 1941. године, које су снаге и покрети у Другом светском рату били носиоци антифашистичке борбе и какав су антифашистички капацитети мали партизански, револуционарни покрет и монархистички равногорски покрет.
Прилози које објављујемо уверљиво сведоче о томе да данас незаобилазна „култура сећања" у земљи Србији не наилази увек на плодно тло, поготово не онда када неодвојива баштина те културе треба да буде истина о антифашистичкој борби српског и других југословенских народа током Другог светског рата, а и раније, када се свет средином 1930-их година поларизовао у два блока – онај фашистички и онај њему супротстављени антифашистички, чији ће први борбени суд бити обележен у грађанском рату у Шпанији 1936–1939. године.
Талас расправао томе када је и како започет устанак у окупираној Србији 1941, које су снаге и покрети у Другом светском рату били носиоци антифашистичке борбе, започет је пре више од четврт века, у време када је у јавности отворено и питање „националног помирења", као изворно политичко питање, снажно натопљено идеологијом. Због тога су, прилично разумљиво, домаће снаге супротстављене у Другом светском рату о том питању имале непомирљиве ставове. Дебате о „помирењу" нужно су наметнуле и потребу за превредновањем прошлости, одакле није био дуг пут и до њеног преобликовања, чак радикалне ревизије, што се поклопило и са сличним, снажно присутним тенденцијама у европској и светској историографији.
Слика прошлости је до почетка 1980-их у нашој историографији била увелико обликована по партијској матрици, то је несумњиво, али су у комунистичкој епохи, поготово на њеном залеску, отворене многе бреше, срушене разне догме, што је омогућило стварање једне много сложеније, објективније и нијансиране слике о прохујалом времену, не само када је реч о Другом светском рату.
Српска научна историографија је у то време снажно закорачила у истраживање „белих мрља" прошлости,одважно је и одговорно кренула у рушење различитих табуа. Тиме је потврдила сопствено зрење, показала да је прешла дуг пут од изрицања пуких идеолошких тврдњи у првим годинама после 1945, до критички обликованог знања, које истражује прошлост и о њој гради истинуту слику на основу увида у нове документе, уз снажан ослонац на нове теоријске моделе и нове методолошке поступке.
Истовремено се, међутим, српска историографија суочила и с нарастајућом плимом афирмације неких ставова и погледа обликованих током прних поратних година у заграничној, емигрантској литератури. Док је српска научна историографија у земљи прешла дуг пут трења, за ону заграничну, емигрантску, то се не би могло рећи, напротив – она је остала конзервирана и при аргументима који су важили у првој поратној деценији. Наизглед парадоксално, из великог дела сфере дневне, страначке политике, која се од краја 1980-их наводно суочила с изазовима савременог света, пружена је одлучујућа подршка заграничној историјској литератури, коју су исписивали припадници страна поражених током рата.
Исувише често, та штива су говорила острашћеним језиком пропаганде, писана ненаучно била су преоптерећена идеолошким и политичким садржајима и искључивостима, а онда су послужила и као основица за ванаучну ревизију историографске слике епохе.
У таквој, обрнутој „историографској перспективи", сасвим логично, дошло је и до потискивања истине о допринос српског народа антифашистичкој борби и кључним носиоцима те борбе током Другог светског рата.
У обимној студији „Рат и историографија", из које објављујемо један део, ове проблеме помно анализира наш познати историчар др Љубодраг Димић, дописни члан САНУ и редовни професор Филозофског факултета у Београду.
„Слика прошлости с недопустивом лакоћом доживљава промене под утицајем ревизије историје које намеће политика, док настојања да се кратно историјским тумачења 'умири' у тачки поделнако удаљеној од 'левих' и 'десних' крајности, захтевају време, труд, посвећеност струци, научно звање", констатује и упозорава професор Димић у једној другој студији, посвећеној 27. марту 1941. године.
(a)
(b)
Nedopustiva lakoća menjanja prošlosti
Danas nezaobilazna „kultura sećanja" u zemlji Srbiji ne nailazi uvek na plodno tlo, pogotovo kada neodvojiva baština te kulture treba da bude istina o antifašističkoj borbi srpskog naroda.
Mada odavno nije u kalendaru državnih praznika, 7. juli, nekada proslavljan kao Dan ustanka u Srbiji, danas će ipak biti obeležen državnom manifestacijom u Beloj Crkvi kod Krupnja, kod spomenika narodnom heroju Žikici Jovanoviću Špancu. Odluku da ova manifestacija bude održana doneo je Odbor za negovanje tradicija oslobodilačkih ratova Srbije, a Vlada Srbije usvojila je tu odluku.
Možda bi se, pomalo optimistički, moglo reći da je to signal koji ukazuje da su konačno za nama užarene, strasne rasprave o tome kada je i kako započet ustanak u okupiranoj Srbiji 1941. godine, koje su snage i pokreti u Drugom svetskom ratu bili nosioci antifašističke borbe i kakav su antifašistički kapaciteti mali partizanski, revolucionarni pokret i monarhistički ravnogorski pokret.
Prilozi koje objavljujemo uverljivo svedoče o tome da danas nezaobilazna „kultura sećanja" u zemlji Srbiji ne nailazi uvek na plodno tlo, pogotovo ne onda kada neodvojiva baština te kulture treba da bude istina o antifašističkoj borbi srpskog i drugih jugoslovenskih naroda tokom Drugog svetskog rata, a i ranije, kada se svet sredinom 1930-ih godina polarizovao u dva bloka – onaj fašistički i onaj njemu suprotstavljeni antifašistički, čiji će prvi borbeni sudar biti obeležen u građanskom ratu u Španiji 1936–1939. godine.
Talas raspravao tome kada je i kako započet ustanak u okupiranoj Srbiji 1941, koje su snage i pokreti u Drugom svetskom ratu bili nosioci antifašističke borbe, započet je pre više od četvrt veka, u vreme kada je u javnosti otvoreno i pitanje „nacionalnog pomirenja", kao izvorno političko pitanje, snažno natopljeno ideologijom. Zbog toga su, prilično razumljivo, domaće snage suprotstavljene u Drugom svetskom ratu o tom pitanju imale nepomirljive stavove. Debate o „pomirenju" nužno su nametnule i potrebu za prevrednovanjem prošlosti, odakle nije bio dug put i do njenog preoblikovanja, čak radikalne revizije, što se poklopilo i sa sličnim, snažno prisutnim tendencijama u evropskoj i svetskoj istoriografiji.
Slika prošlosti je do početka 1980-ih u našoj istoriografiji bila uveliko oblikovana po partijskoj matrici, to je nesumnjivo, ali su u komunističkoj epohi, pogotovo na njenom zalasku, otvorene mnoge breše, srušene razne dogme, što je omogućilo stvaranje jedne mnogo složenije, objektivnije i nijansirane slike o prohujalom vremenu, ne samo kada je reč o Drugom svetskom ratu.
Srpska naučna istoriografija je u to vreme snažno zakoračila u istraživanje „belih mrlja" prošlosti,odvažno je i odgovorno krenula u rušenje različitih tabua. Time je potvrdila sopstveno zrenje, pokazala da je prešla dug put od izricanja pukih ideoloških tvrdnji u prvim godinama posle 1945, do kritički oblikovanog znanja, koje istražuje prošlost i o njoj gradi istinitu sliku na osnovu uvida u nove dokumente, uz snažan oslonac na nove teorijske modele i nove metodološke postupke.
Istovremeno se, međutim, srpska istoriografija suočila i s narastajućom plimom afirmacije nekih stavova i pogleda oblikovanih tokom prvih poratnih godina u zagranični, emigrantskoj literaturi. Dok je srpska naučna istoriografija u zemlji prešla dug put trenja, za onu zagraničnu, emigrantsku, to se ne bi moglo reći, naprotiv – ona je ostala konzervirana i pri argumentima koji su važili u prvoj poratnoj deceniji. Naizgled paradoksalno, iz velikog dela sfere dnevne, stranačke politike, koja se od kraja 1980-ih navodno suočila s izazovima savremenog sveta, pružena je odlučujuća podrška zagraničnoj istorijskoj literaturi, koju su ispisivali pripadnici strana poraženih tokom rata.
Isuviše često, ta štiva su govorila ostrašćenim jezikom propagande, pisana nenaučno bila su preopterećena ideološkim i političkim sadržajima i isključivostima, a onda su poslužila i kao osnovica za vannaučnu reviziju istoriografske slike epohe.
U takvoj, obrnutoj „istoriografskoj perspektivi", sasvim logično, došlo je i do potiskivanja istine o doprinosu srpskog naroda antifašističkoj borbi i ključnim nosiocima te borbe tokom Drugog svetskog rata.
U obimnoj studiji „Rat i istoriografija", iz koje objavljujemo jedan deo, ove probleme pomno analizira naš poznati istoričar dr Ljubodrag Dimić, dopisni član SANU i redovni profesor Filozofskog fakulteta u Beogradu.
„Slika prošlosti s nedopustivom lakoćom doživljava promene pod uticajem revizije istorije koje nameće politika, dok nastojanja da se klatno istorijskih tumačenja 'umiri' u tački podjednako udaljenoj od 'levih' i 'desnih' krajnosti, zahtevaju vreme, trud, posvećenost struci, naučno zvanje", konstatuje i upozorava profesor Dimić u jednoj drugoj studiji, posvećenoj 27. martu 1941. godine.
(c)
(d)

Figure 10: Document conversion: (a) Original Cyrillic text, (b) Cipher text obtained from Cyrillic text according to set C_N, (c) Original Latin text (equivalent to Cyrillic one), (d) Cipher text obtained from Latin text according to set C_N.

Maximum probability (MAX) extracts the most probable difference between gray scale value in pixels. It is defined as [15]:

$$\mathrm{MAX} = \max\left\{P\left(i, j\right)\right\}. \tag{9}$$

Dissimilarity (DIS) is a measure of the variation in gray level pairs of the image. It depends on distance from the diagonal weighted by its probability. The equation of the dissimilarity is [15]:

$$\mathrm{DIS} = \sum_{i=1}^{N}\sum_{j=1}^{N} P\left(i, j\right) \cdot \left|i - j\right|. \tag{10}$$

PANORAMA

U doba moga detinjstva, dakle vrlo davno, u naš grad je došla stalna panorama. U jednoj od glavnih ulica, njen "direktor", mršav Austrijanac, iznajmio je prazan dućan na uglednom mestu, i u njemu otvorio tu radnju. Iznad vrata je bio solidan natpis, zlatnim slovima na crnom staklu, "Panorama svijeta".
Velika prostorija je lepo urenena i diskretno osvetljena. Ništa tu nije ličilo na jevtine šatre u kojima su dosad boravile prolazne panorame. Sve je odavalo trajnost i stalnost. Unaokolo je išao uzak crven ćilim, u uglovima su bile veštačke palme, ali najveći deo prostora, celu sredinu, zauzimala je – panorama. To je bila drvena granevina u obliku višestrane prizme, od finog drveta koje je bilo tako obojeno da ostavlja utisak mahagonija. Ta poligonska zgradica ličila je na drveno turbe, a krov joj je dopirao do ispod dućanskog plafona. Na njoj su uokrug, u pravilnim razmacima, bili otvori za gledanje: po dva velika durbinska oka s naročitim staklima, uokvirena crnim kaučukom. Pored svakog takvog otvora bila je mala stolica bez naslona, presvučena crvenim plišem. Svega petnaest otvora – na svakoj strani poligona po jedan – i isto toliko sedišta.
U nevidljivoj unutrašnjosti te zgrade bile su kružno rasporenene slike u boji, pomoću naročito postavljenih ogledala i jakog električnog osvetljenja one su pred gledaočevim očima postajale uveličane, jarko osvetljene i veoma žive i plastične. Naročita, i opet nevidljiva, mašinerija pokretala je ceo taj pojas slika uokrug. Svaka slika zadržavala se pred svakim gledaocem dva do tri minuta, a onda se kretala dalje udesno, i na njeno mesto dolazila je nova slika, sleva. Tako, dok svaki gledalac ne vidi sve slike.
Program se sastojao od petnaest slika jedne odrenene zemlje ili jednog velikog grada, i menjao se svake nedelje. Jednom je to bio Rio de Žaneiro, drugi put Lisabon, pa Cejlon, i sve tako redom, iz sedmice u sedmicu, sve daleke zemlje i nepoznati gradovi. Cena ulaznice nije bila visoka, pa ipak je za siromaha naka trećeg razreda realne gimnazije predstavljala znatan novac do koga nije bilo uvek lako doći. Ali taj novac se morao naći. Panorama je za mene postala neka vrsta neophodne droge. Dovijao sam se na sve moguće načine kako da donem do niklenog novca koji otvara vrata "Panorame svijeta", a čim bih uspeo da ga dobijem, trčao sam pravo u glavnu ulicu u predavao ga debeloj ženi za kasom na ulazu u panoramu. A kad bih jednom seo na crveni tabure, za mene je počinjao pravi, veliki i obasjani život. Sve što je dotle značilo moj stvarni život, tonulo je u nepostojanje. A sve što sam čitao u romanima ili želeo i gradio u mašti, sve se moglo povezati sa ovim slikama. Moje vidno polje, a sa njim i cela svest, bilo je potpuno ispunjeno zemljama i gradovima koji su preda mnom klizili i u kojima sam se gubio. Tišina je bila svečana i potpuna, samo je s vremena na vreme fini dečački sluh mogao da uhvati slab šum, jedva čujnu škripu mehanizma.

(a)

ПАНОРАМА

У доба мога детињства, дакле врло давно, у наш град је дошла стална панорама. У једној од главних улица, њен "директор", мршав Аустријанац, изнајмио је празан дућан на угледном месту, и у њему отворио ту радњу. Изнад врата је био солидан натпис, златним словима на црном стаклу, "Панорама свијета".
Велика просторија је лепо уренена и дискретно осветљена. Ништа ту није личило на јевтине шатре у којима су досад боравиле пролазне панораме. Све је одавало трајност и сталност. Унаоколо је ишао узак црвен ћилим, у угловима су биле вештачке палме, али највећи део простора, целу средину, заузимала је – панорама. То је била дрвена граневина у облику вишестране призме, од финог дрвета које је било тако обојено да оставља утисак махагонија. Та полигонска зградица личила је на дрвено турбе, а кров јој је допирао до испод дућанског плафона. На њој су уокруг, у правилним размацима, били отвори за гледање: по два велика дурбинска ока с нарочитим стаклима, уоквирена црним каучуком. Поред сваког таквог отвора била је мала столица без наслона, пресвучена црвеним плишем. Свега петнаест отвора – на свакој страни полигона по један – и исто толико седишта.
У невидљивој унутрашњости те зграде биле су кружно распоренене слике у боји, помоћу нарочито постављених огледала и јаког електричног осветљења оне су пред гледаочевим очима постајале увеличане, јарко осветљене и веома живе и пластичне. Нарочита, и опет невидљива, машинерија покретала је цео тај појас слика уокруг. Свака слика задржавала се пред сваким гледаоцем два до три минута, а онда се кретала даље удесно, и на њено место долазила је нова слика, слева. Тако, док сваки гледалац не види све слике.
Програм се састојао од петнаест слика једне одренене земље или једног великог града, и мењао се сваке недеље. Једном је то био Рио де Жанеиро, други пут Лисабон, па Цејлон, и све тако редом, из седмице у седмицу, све далеке земље и непознати градови. Цена улазнице није била висока, па ипак је за сиромаха нака трећег разреда реалне гимназије представљала знатан новац до кога није било увек лако доћи. Али тај новац се морао наћи. Панорама је за мене постала нека врста неопходне дроге. Довијао сам се на све могуће начине како да донем до никленог новца који отвара врата "Панораме свијета", а чим бих успео да га добијем, трчао сам право у главну улицу у предавао га дебелој жени за касом на улазу у панораму. А кад бих једном сео на црвени табуре, за мене је почињао прави, велики и обасјани живот. Све што је дотле значило мој стварни живот, тонуло је у непостојање. А све што сам читао у романима или желео и градио у машти, све се могло повезати са овим сликама. Моје видно поље, а са њим и цела свест, било је потпуно испуњено земљама и градовима који су преда мном клизили и у којима сам се губио. Тишина је била свечана и потпуна, само је с времена на време фини дечачки слух могао да ухвати слаб шум, једва чујну шкрипу механизма.

(b)

Beograd dobija Pupinov spomenik

Ispred tehničkih fakulteta u Beogradu sledećeg septembra biće postavljen spomenik Mihajlu Pupinu, u okviru programa kojim će se obeležiti 160 godina od rođenja tog naučnika. Značajnu sumu za izgradnju obezbediće Fond "Mladen Selak".

U okviru programa kojim će se obeležiti 160 godina od rođenja Mihajla Pupina, sledećeg septembra će naspram spomenika Nikoli Tesli, ispred zgrade tehničkih fakulteta u Beogradu, biti postavljen spomenik tom naučniku. Spomenik će biti delo profesera Miroljuba Stamenkovića.

Značajnu sumu za izgradnju obezbediće Fond "Mladen Selak". To je Fond jednog od najbogatijih i najuticajnijih Srba u Americi.

"Ja duboko, duboko, verujem u nauku, duboko verujem u moć čoveka, da čovek je tu da stvara, da pomaže, da širi dobro oko sebe. Za mene su Pupin i Tesla ljudi koji zaslužuju naše najveće poštovanje", kaže Mladen Selak.

Selak je i finansirao izložbu o Pupinovim delima u Srpskoj akademiji nauke i umetnosti, a sada objavljuje knjigu koju će podeliti našim školama. Za autobiografiju "Sa pašnjaka do naučenjaka", Pupin je dobio Pulicerovu nagradu.

Selak kaže da je literatura skupa, ali i da je ta knjiga dobro urađena. "Knjiga nosi istorijat čoveka koji je ovaj svet zadužio", kaže Selak.

Prema njegovim rečima, telefoni, koje mi danas upotrebljavamo, su Pupinovo delo.

Fondacija je već obnovila Pupinovu rodnu kuću u Idvoru i izdala prigodnu poštansku marku. Sve u cilju da se delo velikog naučnika otrgne od zaborava, kako bi prave vrednosti bile putokaz mladima.

(c)

Београд добија Пупинов споменик

Испред техничких факултета у Београду следећег септембра биће постављен споменик Михајлу Пупину, у оквиру програма којим ће се обележити 160 година од рођења тог научника. Значајну суму за изградњу обезбедиће Фонд "Младен Селак".

У оквиру програма којим ће се обележити 160 година од рођења Михајла Пупина, следећег септембра ће наспрам споменика Николи Тесли, испред зграде техничких факултета у Београду, бити постављен споменик том научнику. Споменик ће бити дело професера Миролуба Стаменковића.

Значајну суму за изградњу обезбедиће Фонд "Младен Селак". То је Фонд једног од најбогатијих и најутицајнијих Срба у Америци.

"Ја дубоко, дубоко, верујем у науку, дубоко верујем у моћ човека, да човек је ту да ствара, да помаже, да шири добро око себе. За мене су Пупин и Тесла људи који заслужују наше највеће поштовање", каже Младен Селак.

Селак је и финансирао изложбу о Пупиновим делима у Српској академији науке и уметности, а сада објављује књигу коју ће поделити нашим школама. За аутобиографију "Са пашњака до научењака", Пупин је добио Пулицерову награду.

Селак каже да је литература скупа, али и да је та књига добро урађена. "Књига носи историјат човека који је овај свет задужио", каже Селак.

Према његовим речима, телефони, које ми данас употребљавамо, су Пупиново дело.

Фондација је већ обновила Пупинову родну кућу у Идвору и издала пригодну поштанску марку. Све у циљу да се дело великог научника отргне од заборава, како би праве вредности биле путоказ младима.

(d)

Figure 11: Custom-oriented database: (a) Printed document in Latin, (b) Printed document in Cyrillic, (c) Web document in Latin, (d) Web document in Cyrillic.

Contrast (CON) or inertia is a measure of the intensity contrast between a pixel and its neighbor over the entire image. Hence, it shows the amount of local variations present in the image. If the image is constant, then the contrast will be equal 0. The highest value of contrast is obtained when the image has random intensity and the pixel intensity and neighbor intensity are very different. The equation of the contrast is [15, 16]:

$$\mathrm{CON} = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)\cdot(i-j)^2. \tag{11}$$

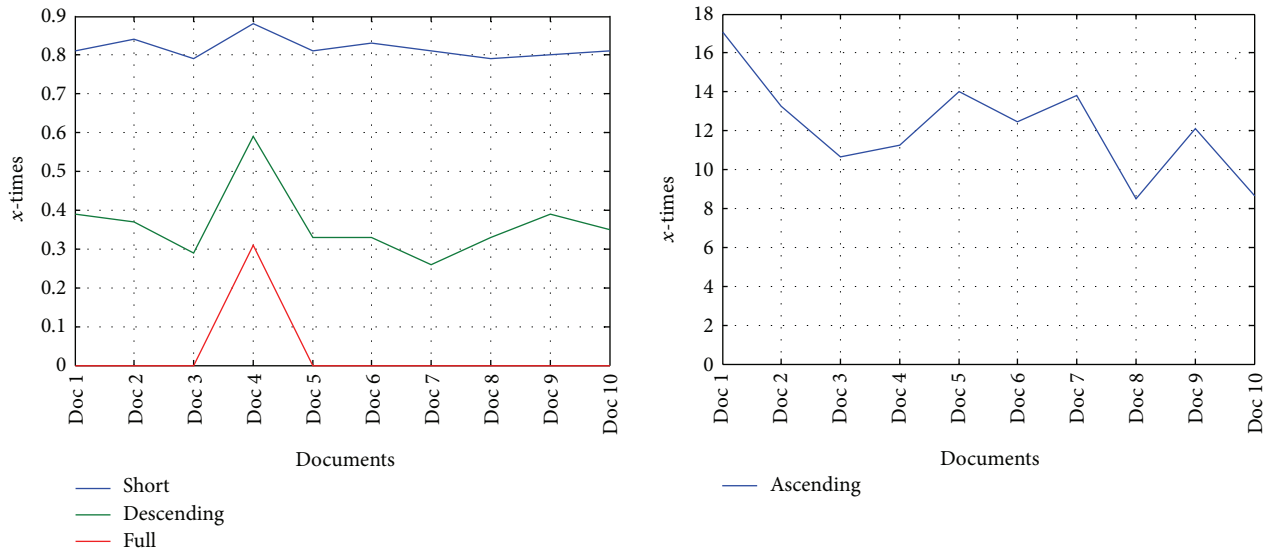


Figure 12: The ratio of the script type occurrence: (a) short, descending and full, (b) ascending.

Table 6: Cooccurrence descriptors for Latin and Cyrillic cipher text.

Serbian language	Latin	Cyrillic	Characterization
Uniformity (energy)	0.2459	0.3811	Latin < Cyrillic
Entropy	−1.6298	−1.4363	Latin > Cyrillic
Maximum probability	0.3722	0.5863	Latin < Cyrillic
Dissimilarity	0.7356	0.6669	Latin > Cyrillic
Contrast	1.0423	1.2660	Latin < Cyrillic

From the above results, it is clear that co-occurrence descriptors can fully characterize the difference between Latin and Cyrillic script. This means that frequency analysis of the occurrence can be supplemented with additional attributes in order to define a strong margin as a criterion to distinguish a certain script.

A brief look at the normalized co-occurrence matrix **P** for the same document written in Latin and Cyrillic scripts (text representing the excerpt of the first four paragraphs from a document given in Figure 10) shows quite a different characterization. The test results are given in Table 5.

Furthermore, the calculation of five co-occurrence descriptors shows the values given in Table 6.

3. Experiments

For the sake of the experiment, a custom-oriented database is created. It consists of 10 documents. These documents represent excerpts from printed and web documents written in Serbian language. The documents are created in both scripts: Latin and Cyrillic. Printed documents are created from PDF documents, while web documents are extracted from web news. The total length of documents given in the database is approx. 75000 letter characters per script (approx. 40 pages). The length of printed documents is from 2273 to 15840 letter characters. Web documents are smaller compared to printed documents. Their length is from 1231 to 2502 letter characters. It should be noted that all documents have more than 1000 letter characters. The example of the printed and web document from the database is shown in Figure 11.

4. Results and Discussion

According to the proposed algorithm, all documents from the database are converted into equivalent cipher texts and subjected to the frequency and co-occurrence analysis. First, the frequency analysis of the script type occurrence in Latin as well as in Cyrillic documents is examined (Table 7). The obtained results for each document are given in Table 8.

The final processing of the results is based on cumulative measures like sum, average, max and min of script type occurrence in the database. According to that the criteria are established. All these are shown in Table 9.

From cumulative results given in Table 10 some criteria can be established. It can be noted that the biggest margin between results are seen in the ratio of ascending letters. This ratio has the value of at least 8. Hence, it is the strongest point of qualitative characterization and recognition of the certain script. Furthermore, the smaller number of short and descending scripts are common in Latin compared to Cyrillic documents. At the and, full letters are quite rare in a Latin document. However, its characterization in criteria form is

Table 7: Frequency analysis of the script type occurrence in documents from database.

Printed documents										
Type of script	Doc 1		Doc 2		Doc 3		Doc 4		Doc 5	
	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic
S	2243	2764	11396	13593	1510	1914	2217	2516	2069	2542
A	906	53	4060	306	693	65	598	53	897	64
D	183	468	724	1933	82	286	261	445	151	461
F	0	7	0	8	0	8	8	26	0	12
Web documents										
Type of script	Doc 6		Doc 7		Doc 8		Doc 9		Doc 10	
	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic
S	1486	1799	1358	1682	783	996	1657	2078	1328	1637
A	598	48	636	46	408	48	750	62	588	68
D	99	304	75	292	58	174	134	344	99	284
F	0	7	0	9	0	13	0	18	0	9

The above results are further processed in order to calculate the ratio of script type occurrence between Latin and Cyrillic document. Complete results are given in Table 8.

Table 8: The ratio of script type occurrence between Latin and Cyrillic documents.

Type of script	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6	Doc 7	Doc 8	Doc 9	Doc 10
S	0.81	0.84	0.79	0.88	0.81	0.83	0.81	0.79	0.80	0.81
A	17.09	13.27	10.66	11.28	14.02	12.46	13.83	8.50	12.10	8.65
D	0.39	0.37	0.29	0.59	0.33	0.33	0.26	0.33	0.39	0.35
F	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00

The results are presented in Figure 12.

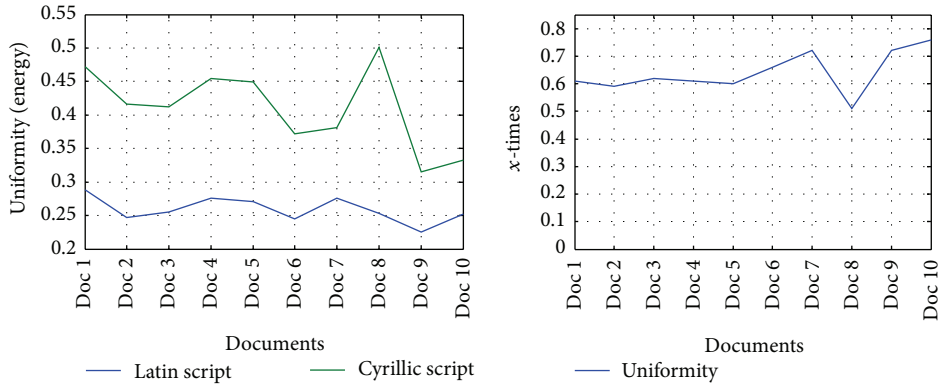
Table 9: The ratio of script type occurrence measures.

Type of script	$\sum$		Ratio			Criteria
	Latin	Cyrillic	Average	Max.	Min.	
S	26047	31521	0.82	0.88	0.79	>0.75
A	10134	813	12.21	17.09	8.50	>8
D	1866	4991	0.36	0.59	0.26	<0.6
F	8	117	0.03	0.31	0.00	?

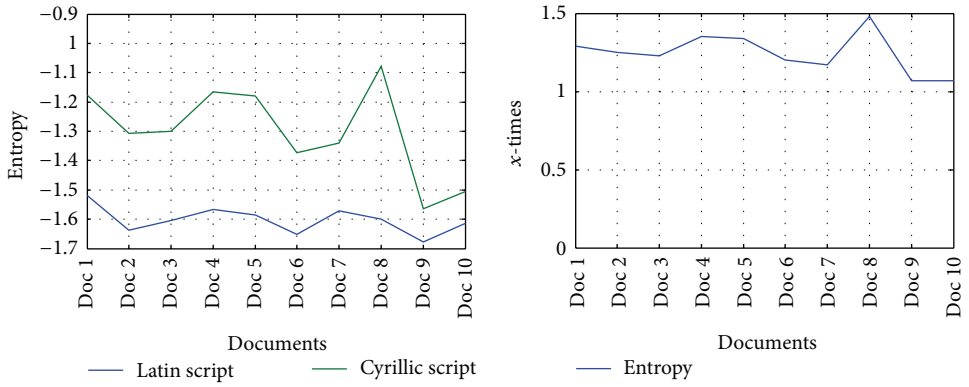
Table 10: GLCM five descriptors of the script type co-occurrence in documents from database.

Printed documents										
	Doc 1		Doc 2		Doc 3		Doc 4		Doc 5	
	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic
Uniformity	0.2885	0.4725	0.2473	0.4167	0.2557	0.4120	0.2759	0.4545	0.2707	0.4498
Entropy	−1.5191	−1.1774	−1.6379	−1.3079	−1.6047	−1.2999	−1.5675	−1.1650	−1.5847	−1.1799
Max. probability	0.4655	0.6636	0.3952	0.6139	0.4120	0.6098	0.4439	0.6457	0.4349	0.6405
Dissimilarity	0.6847	0.5933	0.7469	0.6592	0.7502	0.6427	0.7064	0.6041	0.7117	0.6217
Contrast	1.0324	1.1790	1.1106	1.2859	1.1258	1.2261	1.0577	1.1449	1.0630	1.1949
Web documents										
	Doc 6		Doc 7		Doc 8		Doc 9		Doc 10	
	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic
Uniformity	0.2447	0.3714	0.2754	0.3817	0.2533	0.5005	0.2252	0.3147	0.2522	0.3325
Entropy	−1.6524	−1.3738	−1.5725	−1.3412	−1.5990	−1.0779	−1.6778	−1.5650	−1.6144	−1.5059
Max. probability	0.3964	0.5650	0.4409	0.5753	0.3972	0.6844	0.3195	0.5154	0.4016	0.5318
Dissimilarity	0.7723	0.7320	0.6912	0.7209	0.7294	0.5686	0.8317	0.7667	0.7256	0.7416
Contrast	1.1862	1.3869	1.0287	1.3681	1.0459	1.1158	1.2122	1.4220	1.0641	1.3716

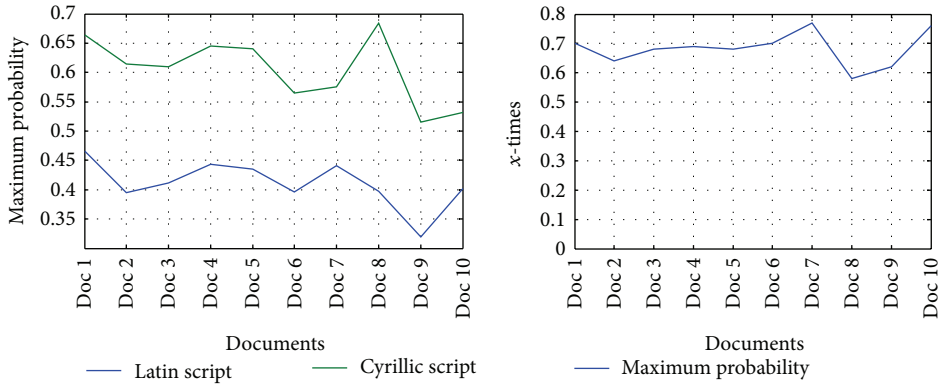
The above results are further processed in order to calculate the ratio of script type co-occurrence in between Latin and Cyrillic document. These results are shown in Table 11.



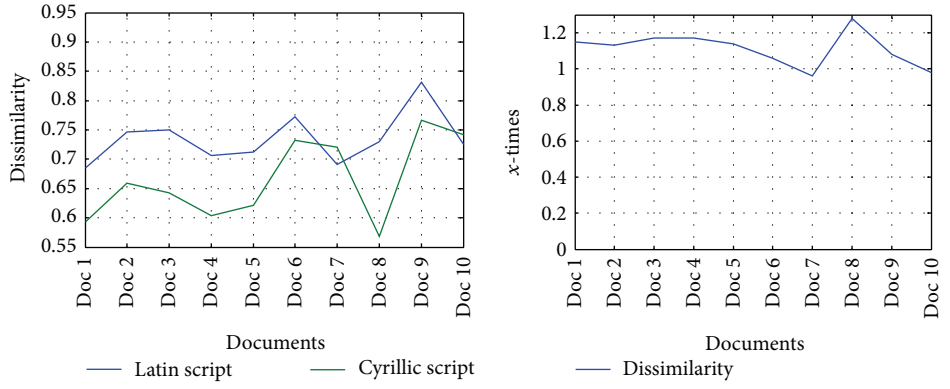
(a)



(b)



(c)



(d)

Figure 13: Continued.

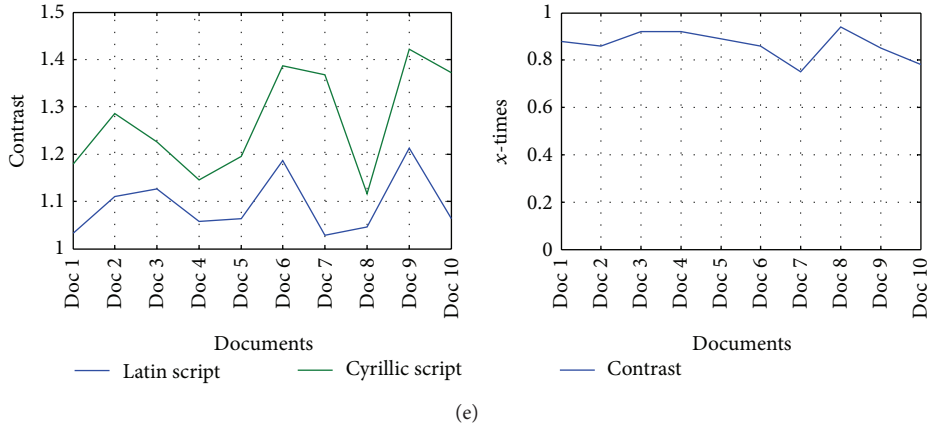


Figure 13: Illustrations of co-occurrence descriptors in Latin and Cyrillic text (left) and its ratio (right): (a) Uniformity, (b) Entropy, (c) Maximum probability, (d) Dissimilarity, (e) Contrast.

Table 11: The ratio of the co-occurrence descriptors between Latin and Cyrillic documents.

	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6	Doc 7	Doc 8	Doc 9	Doc 10
Uniformity	0.61	0.59	0.62	0.61	0.60	0.66	0.72	0.51	0.72	0.76
Entropy	1.29	1.25	1.23	1.35	1.34	1.20	1.17	1.48	1.07	1.07
Max. probability	0.70	0.64	0.68	0.69	0.68	0.70	0.77	0.58	0.62	0.76
Dissimilarity	1.15	1.13	1.17	1.17	1.14	1.06	0.96	1.28	1.08	0.98
Contrast	0.88	0.86	0.92	0.92	0.89	0.86	0.75	0.94	0.85	0.78

The final processing of the above results is based on cumulative measures like average, max. and min. of script type co-occurrence in the database. According to that certain criteria are established. All these are shown in Table 12.

Table 12: The ratio of script type co-occurrence descriptors.

	Latin			Cyrillic			Ratio			Criteria
	Min.	Max.	Average	Min.	Max.	Average	Max.	Min.	Average	
Uniformity	0.23	0.29	0.26	0.31	0.50	0.41	0.76	0.51	0.64	0.3
Entropy	−1.68	−1.52	−1.60	−1.57	−1.08	−1.30	1.48	1.07	1.25	?
Max. probability	0.32	0.47	0.42	0.52	0.68	0.60	0.77	0.58	0.68	0.5
Dissimilarity	0.68	0.83	0.74	0.57	0.77	0.67	1.28	0.96	1.11	?
Contrast	1.03	1.21	1.12	1.12	1.42	1.27	0.94	0.75	0.86	?

quite problematic due to their absence in Latin documents from time to time.

Furthermore, the analysis of the script type co-occurrence in Latin as well as in Cyrillic documents is examined according to GLCM method. The obtained results for each document are given in Table 10.

The co-occurrence descriptor for Latin and Cyrillic text and its ratio is presented in Figure 13.

From the above results, some criteria can be established. It is clear that uniformity and maximum probability receive the most distinct values in Latin and Cyrillic text. Hence, these descriptors are suitable for qualitative characterization of Latin and Cyrillic text as well as for creating criteria to distinguish a certain script type. From the above results, the margin criteria should be uniformity of 0.3 and maximum probability of 0.5. These values of both descriptors represent the strong margin in qualifying the script in certain Serbian text. If we accompany them with the criteria obtained from frequency analysis of the script type occurrence, then the full criteria of decision making can be established. This will lead to correct recognition of the script in Serbian text.

5. Conclusion

The paper proposed the algorithm for recognition of exact script in Serbian document. Documents in Serbian language can be written in two different scripts: Latin or Cyrillic. The proposed algorithm converts document written in Latin and Cyrillic script into cipher text. This way, all alphabetic characters are exchanged with only four different encrypted signs according to predefined encryption based on text line structure definition. Such ciphers texts are then subjected to the frequency and co-occurrence analysis. According to the obtained results a criteria for recognition of the certain script is proposed. The proposed method is applied to the custom-oriented database which includes different types of printed

and web documents. The experiment shows encouraging results. Possible applications can be seen in the area of web page recognition.

Future work will be toward the recognition of related languages as well as different languages written in the same script.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this article ("Recognition of the Script in Serbian Documents using Frequency Occurrence and Co-occurrence Analysis" by Darko Brodić, Zoran N. Milivojević, Čedomir A. Maluckov).

References

[1] C. Smith, "Basic cryptanalysis techniques," Tech. Rep., SANS Institute, 2001.

[2] J. Hoffstein, J. Pipher, and J. H. Silverman, *An Introduction to Mathematical Cryptography*, Springer, New York, NY, USA, 2008.

[3] K. Ruohonen, "Mathematical cryptology," Lecture Notes, 2010.

[4] A. Zramdini and R. Ingold, "Optical font recognition using typographical features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 877–882, 1998.

[5] P. Quaresma, "Frequency analysis of the Portuguese language," TR, 2008/003, University of Coimbra, Portugal, 2008.

[6] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[7] G. Jaskiewicz, "Analysis of letter frequency distribution in the voynich manuscript," in *Proceedings of the International Workshop (CS&P '11)*, pp. 250–261, Putusk, Poland, 2011.

[8] S. Cagnoni, *Evolutionary Image Analysis and Signal Processing*, vol. 213 of *Studies in Computational Intelligence*, Springer, Berlin, Germany, 2009.

[9] R. Nosaka, C. H. Suryanto, and K. Fukui, "Rotation invariant co-occurrence among adjacent LBPs," in *Proceedings of the 11th International Conference on Computer Vision (ACCV '11)*, vol. 7088 of *Lecture Notes in Computer Science*, pp. 82–91, Daejeon, Republic of Korea, 2011.

[10] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[11] A. Baraldi and F. Parmiggiani, "Investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 293–304, 1995.

[12] F. R. Renzetti and L. Zortea, "Use of a gray level co-occurrence matrix to characterize duplex stainless steel phases microstructure," *Frattura ed Integrità Strutturale*, vol. 16, no. 5, pp. 43–51, 2011.

[13] R. Susomboon, D. Raicu, J. Furst, and T. B. Johnson, "A Co-occurrence texture semi-invariance to direction, distance and patient size," in *Medical Imaging 2008: Image Processing*, vol. 6914 of *Procedings of SPIE*, San Diego, Calif, USA, February 2008.

[14] A. Eleyan and H. Demirel, "Co-occurrence matrix and its statistical features as a new approach for face recognition," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 19, no. 1, pp. 97–107, 2011.

[15] D. A. Clausi and Y. Zhao, "Rapid extraction of image texture by co-occurrence using a hybrid data structure," *Computers and Geosciences*, vol. 28, no. 6, pp. 763–774, 2002.

[16] R. Jobanputra and D. A. Clausi, "Preserving boundaries for image texture segmentation using grey level co-occurring probabilities," *Pattern Recognition*, vol. 39, no. 2, pp. 234–245, 2006.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 369103, 13 pages
http://dx.doi.org/10.1155/2013/369103

Research Article

Rao and Wald Tests for Adaptive Detection in Partially Homogeneous Environment with a Diversely Polarized Antenna

Chaozhu Zhang, Jing Zhang, and Chengyuan Liu

College of Information and Communications Engineering, Harbin Engineering University, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Jing Zhang; zhangjing_heu@hrbeu.edu.cn

Received 27 April 2013; Accepted 14 July 2013

Academic Editors: S. Bourennane, C. Fossati, J. Marot, and K. Spinnler

This study considers Rao test and Wald test for adaptive detection based on a diversely polarized antenna (DPA) in partially homogeneous environment. The theoretical expressions for the probability of false alarm and detection are derived, and constant false alarm rate (CFAR) behaviour is remarked on. Furthermore, the monotonicities of detection probability of the two detectors are proved, and a polarization optimization detection algorithm to enhance the detection performance is proposed. The numerical simulations are conducted to attest to the validity of the above theoretical analysis and illustrate the improvement in the detection performance of the proposed optimization algorithm.

1. Introduction

Detecting a signal of interest in the presence of noise is often encountered in radar/sonar signal processing. In an ideal situation, the noise in the training data is assumed to share the same covariance matrix as that in the test data. This situation is often referred to as a homogeneous environment. In this case, many classic algorithms, such as the generalized likelihood ratio test (GLRT) detector [1] and adaptive matched filter (AMF) detector [2], are widely used. A prominent feature of the two detectors is constant false alarm rate (CFAR). The GLRT detector is obtained through replacing all the unknown parameters with their maximum likelihood (ML) estimates under each hypothesis within one step [1]. We refer to this procedure as a one-step design procedure. On the opposite, the AMF detector is derived with an ad hoc two-step design procedure [2]. In other words, this design procedure is the first to assume that the noise covariance matrix is known, and to obtain the GLRT test by maximizing over other unknown parameters. The ML estimate of the noise covariance matrix based on the training data alone is then substituted into this test. In the following, the GLRT and AMF detectors are referred to as the one-step and two-step GLRT-based detectors, respectively.

However, nonhomogeneous environments may be encountered in many practical applications [3]. In airborne radars, for instance, the ground clutter in the test and training data is generated by reflections from different portions of the ground. The power level fluctuations of the ground clutter might arise due to variations in terrain [4]. Several models have been proposed for the nonhomogeneous environments [5]. One of these models is the partially homogeneous model, where the noise covariance matrix has the same structure in the test and training data samples but may differ by a scaling factor. This partially homogeneous model is also used in a wireless communication system with fades over multiple sources of interference [6]. The detection problems in the partially homogeneous environments have recently attracted much attention [6–17]. Many detection algorithms, such as matched subspace detector (MSD) [12] and adaptive subspace detector (ASD) [13], are proposed to deal with detection problems in the partially homogeneous environment. Notice that the ASDs contain the common one-step and two-step GLR-based detectors for the case in which the scaling factor of the test data may deviate from that of the training data [3]. Additionally, the CFAR ASD has been proved to be a uniformly most powerful invariant test [6, 17]. Interestingly, the CFAR ASD exhibits great robustness to the scaling of the test data, whereas Kelly's GLR detector [1] and Robey's AMF [2] are sensitive to the scaling factor and even cease to be CFAR [6]. Reference [16] showed that the CFAR ASD is

GLRT when the test measurement is not constrained to have the same noise level as the training data.

Moreover, the GLRT has been the most commonly employed one in signal processing. Nevertheless, since there is no particular a priori reason to exploit the GLRT rather than the others, during the last three decades, Rao test [18] and Wald test [19] applied to practical signal processing detection problems have started to appear in open literature. For instance, in [20], an adaptive detector based on Rao criterion is devised to discriminate the presence of a deterministic signal, with unknown amplitude, in Gaussian noise with unknown but structured (AR parameterized) spectra. In [21, 22], Rao and Wald tests are, respectively, derived to detect a signal with unknown amplitude in homogeneous Gaussian disturbance with unknown covariance matrix. In [23], Rao test has been applied to the problem of radar space-time adaptive processing (STAP), while in [24, 25], Rao and Wald tests are devised with reference to adaptive detection of distributed targets in non-Gaussian clutter. In [26–28], Rao and Wald tests have been applied to multiple-input multiple-output radar detection in compound-Gaussian clutter. Finally, in [29, 30], the coincidence and statistical equivalence of the GLRT, Rao test, and Wald test are proved.

As is well known, a diversely polarized antenna (DPA) has some inherent advantages over a scalar sensor, since it can handle signals based on their polarization characteristics. The tripole antenna in this paper is a common diversely polarized antenna, and it consists of three mutually perpendicular short dipoles—all centered at the same location, and it has as its output the three components of the impinging electric filed [31].

Generally, the performance of a radar system is associated with many characteristics of the transmitted signals, such as polarization [32]. Therefore, the system performance (for example, detection capability) can be improved by optimally selecting the transmitted signals. In [33], the authors have developed a polarimetric detector based only on several primary data vectors and shown that this test statistic has the standard F-distribution. Hence, the detection performance of the polarimetric detector can be improved by optimally choosing the polarization of the transmitted pulses to maximize the noncentrality parameter. Based on a diversely polarized antenna, [34, 35] addressed the problems of adaptive detection and performance enhancement by optimally selecting the polarization of the transmitted pulses of the DPA in homogeneous and partially homogeneous environments. Hence, the detection performance of the polarimetric detector can be improved by optimally choosing the polarization of the transmitted pulses to maximize the noncentrality parameter.

In this paper, we study Rao test and Wald test for adaptive detection based on a DPA in partially homogeneous environment. Our main contribution is to derive the expressions for the probabilities of false alarm and detection of Rao test and Wald test with unknown noise covariance matrix structure based on a DPA, in the partially homogeneous environment. It is found from these expressions derived that Rao test and Wald test have the CFAR property. The other contribution of this paper is that the monotonicities of detection probability of the two detectors are proved and a polarization optimization detection algorithm to enhance the detection performance is proposed. The improvement in the detection performance of the DPA is achieved by using the proposed algorithm to optimally select the polarization of the transmitted pulses.

The rest of this paper is organised as follows. The statement of the problem and the description of the signal and noise models are given in Section 2. The theoretical expressions for the probability of false alarm and detection of Rao and Wald detectors are derived in Section 3. The monotonicities of detection probability of the two detectors are proved in Section 4. Polarization optimization detection algorithm is proposed in Section 5. The simulation results of the two detectors as well as the GLRT are displayed in Section 6. Lastly, conclusions are given in Section 7.

2. Signal Model

In this section, we consider a detection problem in partially homogeneous environments. The received Q-dimensional complex vector $\mathbf{x}$, commonly called primary data or test data, is constrained to be of the form

$$\mathbf{x} = \Sigma\mathbf{s} + \mathbf{n}, \tag{1}$$

where Σ is a known $Q \times q$ dimension signal subspace matrix representing the system response associated with the characteristics of the transmitted signals (e.g., polarization), and suppose that $Q > q$ and $\operatorname{rank}(\Sigma) = q$; $\mathbf{s}$ is a q-dimensional deterministic but unknown complex vector accounting for the target reflectivity and the channel propagation effects; $\mathbf{n}$ is a noise data vector and is assumed to have a complex circular Gaussian distribution with zero mean and covariance matrix $\mu\mathbf{R}$, that is, $\mathbf{n} \sim \mathrm{CN}(0, \mu\mathbf{R})$, where $\mathbf{R}$ is an unknown, positive definite noise covariance matrix structure and μ is an unknown scaling of the noise in the test data. Notice that the scaling factor accounts for the noise power mismatch between the primary and secondary data. The arbitrary scaling between the primary and secondary data is important in some realistic scenarios [3].

Suppose that K ($K > Q$) secondary data samples free of the target signal, that is $\{\mathbf{y}_k, k = 1, \ldots, K \mid \mathbf{y}_k \sim \mathrm{CN}(0, \mathbf{R})\}$, are available. The problem of detection is to decide whether the target signal is present or not in the range cell under test. This problem can be posed in terms of a binary hypotheses test. We let the null hypothesis (H_0) be that no target signal is present and let the alternative hypothesis (H_1) be that the data contains target signal. Hence, the detection problem is to decide between the null hypothesis and the alternative one and can be stated as a parameter test:

$$\begin{aligned} H_0 &: \begin{cases} \mathbf{x} \sim \mathrm{CN}(0, \mu\mathbf{R}), \\ \mathbf{y}_k \sim \mathrm{CN}(0, \mathbf{R}), \quad k = 1, \ldots, K, \end{cases} \\ H_1 &: \begin{cases} \mathbf{x} \sim \mathrm{CN}(\Sigma\mathbf{s}, \mu\mathbf{R}), \\ \mathbf{y}_k \sim \mathrm{CN}(0, \mathbf{R}), \quad k = 1, \ldots, K. \end{cases} \end{aligned} \tag{2}$$

3. Adaptive detectors

Now, a brief introduction about ASD is explained in order to study Rao test and Wald test performances.

In practice, a prior knowledge on the covariance matrix structure is usually unknown. According to [35], the detector used to handle the detection problem with unknown $\mathbf{R}$, which is referred to as ASD, is

$$\Psi = \frac{\mathbf{x}^H \widehat{\mathbf{R}}^{-1} \Sigma \left(\Sigma^H \widehat{\mathbf{R}}^{-1} \Sigma\right)^{-1} \Sigma^H \widehat{\mathbf{R}}^{-1} \mathbf{x}}{\mathbf{x}^H \widehat{\mathbf{R}}^{-1} \mathbf{x}} \underset{H_1}{\overset{H_0}{\lessgtr}} g_{\mathrm{ASD}}, \tag{3}$$

where $g_{\mathrm{ASD}} \in (0,1)$ is the detection threshold and $\widehat{\mathbf{R}} = \sum_{k=1}^{K} \mathbf{y}_k \mathbf{y}_k^H$. The superscript "$H$" denotes complex conjugation transpose.

The false alarm probability function of ASD can be written as

$$F\left(g_{\mathrm{ASD}}\right) = P_{\mathrm{FA}}^{\mathrm{ASD}}\left\{\Psi > g_{\mathrm{ASD}}|_{H_0}\right\} = \int_0^1 P_{\mathrm{FA}|\rho}^{\mathrm{ASD}} f_\rho\left(\rho\right) d\rho, \tag{4}$$

where ρ denotes a loss factor whose distribution is

$$f_\rho\left(\rho\right) = \frac{K! \rho^{K-Q+q} (1-\rho)^{Q-q-1}}{(Q-q-1)!\,(K-Q+q)!}, \quad 0 < \rho < 1,$$

$$P_{\mathrm{FA}|\rho}^{\mathrm{ASD}} = \left(\frac{1-g_{\mathrm{ASD}}}{1-g_{\mathrm{ASD}}\rho}\right)^{K-Q+1} \sum_{j=1}^{q} C_{K-Q+q-j}^{q-j} \left[\frac{g_{\mathrm{ASD}}(1-\rho)}{1-g_{\mathrm{ASD}}\rho}\right]^{q-j}. \tag{5}$$

Furthermore, the probability detection function of ASD is [3]

$$Z\left(g_{\mathrm{ASD}}\right) = P_D^{\mathrm{ASD}}\left\{\Phi > g_{\mathrm{ASD}}|_{H_1}\right\} = \int_0^1 P_{D|\rho}^{\mathrm{ASD}} f_\rho\left(\rho\right) d\rho, \tag{6}$$

where the probability of detection conditioned on ρ is

$$\begin{aligned} P_{D|\rho}^{\mathrm{ASD}} &= 1 - \left[\frac{g_{\mathrm{ASD}}(1-\rho)}{1-g_{\mathrm{ASD}}\rho}\right]^{q-1} \left(\frac{1-g_{\mathrm{ASD}}}{1-g_{\mathrm{ASD}}\rho}\right)^{K-Q+1} \\ &\quad \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \left[\frac{g_{\mathrm{ASD}}(1-\rho)}{1-g_{\mathrm{ASD}}}\right]^{j} \\ &\quad \times \exp\left[\frac{-\Gamma\rho\left(1-g_{\mathrm{ASD}}\right)}{1-g_{\mathrm{ASD}}\rho}\right] \sum_{m=0}^{j-1} \frac{1}{m!} \left[\frac{\Gamma\rho\left(1-g_{\mathrm{ASD}}\right)}{1-g_{\mathrm{ASD}}\rho}\right]^{m}, \\ \Gamma &= \mathbf{s}^H \left[\Sigma^H \left(\mu \mathbf{R}\right)^{-1} \Sigma\right] \mathbf{s}. \end{aligned} \tag{7}$$

Reference [29] proved that Rao test, Wald test, and GLRT [16] coincide in the presence of partially homogeneous complex circular Gaussian disturbance with unknown covariance matrix and Rao test which can be written as

$$\Phi = \frac{Q\left(K+1-Q\right)\Psi}{\left(K+1\right)\left(K+1+Q\Psi\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} g_{R,1} \tag{8}$$

$$\Longleftrightarrow \Psi \underset{H_1}{\overset{H_0}{\lessgtr}} g_{R,2} = \frac{g_{R,1}(K+1)^2}{Q\left[K+1-Q-(K+1)\,g_{R,1}\right]}, \tag{9}$$

where $g_{R,1} \in (0, Q(K+1-Q)/(K+1)(K+1+Q))$ denotes the appropriate modification of thethreshold in [29]. Let $\alpha = (Q(K+1-Q)-(K+1)(Q+K+1)g_{R,1})/(Q(K+1-Q)-(K+1)(Q+\rho K+\rho)g_{R,1})$ $(0<\alpha<1)$. The false alarm probability of Rao test can be written as

$$\begin{aligned} &P_{\mathrm{FA}}^{\mathrm{Rao}}\left\{\Phi > g_{R,1}|_{H_0}\right\} \\ &\quad = P_{\mathrm{FA}}^{\mathrm{ASD}}\left\{\Psi > g_{R,2}|_{H_0}\right\} \\ &\quad = P_{\mathrm{FA}}^{\mathrm{ASD}}\left\{\Psi > \frac{g_{R,1}(K+1)^2}{Q\left[K+1-Q-(K+1)\,g_{R,1}\right]}\bigg|_{H_0}\right\} \\ &\quad = F\left(\frac{g_{R,1}(K+1)^2}{Q\left[K+1-Q-(K+1)\,g_{R,1}\right]}\right) \\ &\quad = \int_0^1 P_{\mathrm{FA}|\rho}^{\mathrm{Rao}} f_\rho\left(\rho\right) d\rho, \end{aligned} \tag{10}$$

where

$$P_{\mathrm{FA}|\rho}^{\mathrm{Rao}} = \alpha^{K-Q+1} \sum_{j=1}^{q} C_{K-Q+q-j}^{q-j} (1-\alpha)^{q-j}. \tag{11}$$

The detection probability of Rao test can be written as

$$\begin{aligned} &P_D^{\mathrm{Rao}}\left\{\Phi > g_{R,1}|_{H_1}\right\} \\ &\quad = P_D^{\mathrm{ASD}}\left\{\Psi > g_{R,2}|_{H_1}\right\} \\ &\quad = P_D^{\mathrm{ASD}}\left\{\Psi > \frac{g_{R,1}(K+1)^2}{Q\left[K+1-Q-(K+1)\,g_{R,1}\right]}\bigg|_{H_1}\right\} \\ &\quad = Z\left(\frac{g_{R,1}(K+1)^2}{Q\left[K+1-Q-(K+1)\,g_{R,1}\right]}\right) \\ &\quad = \int_0^1 P_{D|\rho}^{\mathrm{Rao}} f_\rho\left(\rho\right) d\rho, \end{aligned} \tag{12}$$

where

$$\begin{aligned} P_{D|\rho}^{\mathrm{Rao}} &= 1 - (1-\alpha)^{q-1} \alpha^{K-Q+1} \\ &\quad \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \left(\frac{1-\alpha}{\alpha}\right)^{j} \exp\left(-\Gamma\rho\alpha\right) \\ &\quad \times \sum_{m=0}^{j-1} \frac{1}{m!} (\Gamma\rho\alpha)^{m}. \end{aligned} \tag{13}$$

From [29] we can get Wald test

$$\Omega = \frac{Q\Psi}{\left(K+1-Q\right)\left(1-\Psi\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} g_{W,1} \tag{14}$$

$$\Longleftrightarrow \Psi \underset{H_1}{\overset{H_0}{\lessgtr}} g_{W,2} = \frac{g_{W,1}\left(K+1-Q\right)}{Q + \left(K+1-Q\right) g_{W,1}}, \tag{15}$$

where $g_{W,1} \in (0, \infty)$ denotes the appropriate modification of the threshold in [29, p. 387, formula (12)]. Let $\beta = Q/(Q+(1-\rho)(K+1-Q)g_{W,1})$, $0 < \beta < 1$, and the false alarm probability of Wald test can be written as

$$\begin{aligned} P_{\mathrm{FA}}^{\mathrm{Wald}} &\left\{\Omega > g_{W,1}|_{H_0}\right\} \\ &= P_{\mathrm{FA}}^{\mathrm{ASD}} \left\{\Psi > g_{W,2}|_{H_0}\right\} \\ &= P_{\mathrm{FA}}^{\mathrm{ASD}} \left\{\Psi > \frac{g_{W,1}(K+1-Q)}{Q+(K+1-Q)g_{W,1}}\bigg|_{H_0}\right\} \\ &= F\left(\frac{g_{W,1}(K+1-Q)}{Q+(K+1-Q)\,g_{W,1}}\right) \\ &= \int_0^1 P_{\mathrm{FA}|\rho}^{\mathrm{Wald}} f_\rho(\rho)\, d\rho, \end{aligned} \tag{16}$$

where

$$P_{\mathrm{FA}|\rho}^{\mathrm{Wald}} = \beta^{K-Q+1} \sum_{j=1}^{q} C_{K-Q+q-j}^{q-j} (1-\beta)^{q-j}. \tag{17}$$

The detection probability of Wald test can be written as

$$\begin{aligned} P_D^{\mathrm{Wald}} &\left\{\Omega > g_{W,1}\big|_{H_1}\right\} \\ &= P_D^{\mathrm{ASD}} \left\{\Psi > g_{W,2}\big|_{H_1}\right\} \\ &= P_D^{\mathrm{ASD}} \left\{\Psi > \frac{g_{W,1}(K+1-Q)}{Q+(K+1-Q)g_{W,1}}\bigg|_{H_1}\right\} \\ &= Z\left(\frac{g_{W,1}(K+1-Q)}{Q+(K+1-Q)\,g_{W,1}}\right) \\ &= \int_0^1 P_{D|\rho}^{\mathrm{Wald}} f_\rho(\rho)\, d\rho, \end{aligned} \tag{18}$$

where

$$\begin{aligned} P_{D|\rho}^{\mathrm{Wald}} &= 1 - (1-\beta)^{q-1}\beta^{K-Q+1} \\ &\quad \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \left(\frac{1-\beta}{\beta}\right)^j \exp(-\Gamma\rho\beta) \\ &\quad \times \sum_{m=0}^{j-1} \frac{1}{m!}(\Gamma\rho\beta)^m. \end{aligned} \tag{19}$$

From (10) and (16) we can see that Rao test and Wald test have the desirable constant false alarm rate (CFAR) property with respect to both the shared noise covariance matrix structure $\mathbf{R}$ and the scaling μ of the noise in the test data.

4. The Monotonicities of Detection Probability

In this section, the monotonicities of detection probability are proved. Firstly, a proposition is introduced, and then a polarization optimization detection algorithm is proposed to enhance the detection performance of the two detectors.

Proposition 1. *Both P_D^{Rao} and P_D^{Wald} are monotonically increasing functions for $\Gamma > 0$. (Since $\mathbf{R}^{-1}$ is positive definite, we have $\Gamma > 0$.)*

Proof. We transform (14) into an equivalent form as (20) in [35], and then we have $0 < \alpha < 1$. Using (17), we can obtain

$$\begin{aligned} \frac{dP_D^{\mathrm{Rao}}}{d\Gamma} &= \int_0^1 \frac{dP_{D|\rho}^{\mathrm{Rao}}}{d\Gamma} f_\rho(\rho)\, d\rho \\ &= \int_0^1 (1-\alpha)^{q-1}\alpha^{K-Q+1} \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \\ &\quad \times \left(\frac{1-\alpha}{\alpha}\right)^j (\rho\alpha)\exp(-\Gamma\rho\alpha)\, f_\rho(\rho) \\ &\quad \times \sum_{m=0}^{j-1} \left\{\frac{1}{m!}(\Gamma\rho\alpha)^m - \frac{m}{m!}(\Gamma\rho\alpha)^{m-1}\right\} d\rho \\ &= \int_0^1 (1-\alpha)^{q-1}\alpha^{K-Q+1} \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \\ &\quad \times \left(\frac{1-\alpha}{\alpha}\right)^j (\rho\alpha)\exp(-\Gamma\rho\alpha)\, f_\rho(\rho) \\ &\quad \times \frac{1}{(j-1)!}(\Gamma\rho\alpha)^{j-1} d\rho. \end{aligned} \tag{20}$$

It follows from $0 < \alpha < 1$ and $\Gamma > 0$ that the function of ρ in the integral in the right-hand side of (20) is positive for $0 < \rho < 1$ and is zero for $\rho = 0$ or 1. Thus, the integral of this function over $[0, 1]$ is greater than zero; namely, the derivative in (20) is positive. □

Proof. We transform (8) into an equivalent form as (20) in [35], and then we have $g_{W,1} \in (0, \infty)$. Using (12), we can obtain

$$\begin{aligned} \frac{dP_D^{\mathrm{Wald}}}{d\Gamma} &= \int_0^1 \frac{dP_{D|\rho}^{\mathrm{Wald}}}{d\Gamma} f_\rho(\rho)\, d\rho \\ &= \int_0^1 (1-\beta)^{q-1}\beta^{K-Q+1} \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \\ &\quad \times \left(\frac{1-\beta}{\beta}\right)^j (\rho\beta)\exp(-\Gamma\rho\beta)\, f_\rho(\rho) \\ &\quad \times \sum_{m=0}^{j-1} \left\{\frac{1}{m!}(\Gamma\rho\beta)^m - \frac{m}{m!}(\Gamma\rho\beta)^{m-1}\right\} d\rho \\ &= \int_0^1 (1-\beta)^{q-1}\beta^{K-Q+1} \times \sum_{j=1}^{K-Q+1} C_{K-Q+q}^{q+j-1} \end{aligned}$$

$$\times \left(\frac{1-\beta}{\beta}\right)^{j} (\rho\beta) \exp(-\Gamma\rho\beta) f_{\rho}(\rho) \times \frac{1}{(j-1)!}(\Gamma\rho\beta)^{j-1} d\rho. \tag{21}$$

It can be proved in the same way that the derivative in (21) is positive too. The proof is completed. □

From the proposition above we can see that the greater the value of Γ ($\Gamma > 0$), the better detection performance. The detection performance of Rao test and Wald test can be enhanced by designing the system response Σ to maximize the parameter Γ. The system response matrix can be parameterized as $\Sigma = \Sigma(\varepsilon)$. The problem of performance enhancement of Rao test and Wald test can be formulated as

$$\widehat{\varepsilon} = \arg \max_{\varepsilon} \left\{ \mathbf{s}^{H} \left[\Sigma^{H}(\varepsilon) (\mu \mathbf{R})^{-1} \Sigma(\varepsilon) \right] \mathbf{s} \right\}. \tag{22}$$

5. Polarization Optimization Detection Algorithm

The matrix $\mathbf{V}$ is the response of the diversely polarized sensor array [33]. If the array is a tripole antenna, it can be written as

$$\mathbf{V} = \begin{bmatrix} -\sin\varphi & -\cos\varphi\sin\psi \\ \cos\varphi & -\sin\varphi\sin\psi \\ 0 & \cos\psi \end{bmatrix}, \tag{23}$$

where φ and ψ denote the elevation and azimuth angles of the target return with $\phi \in [0, \pi]$ and $\psi \in [-\pi, \pi]$.

The vector $\mathbf{z}_p(t)$ is the pth pulse of the narrowband transmitted signal which can be represented by

$$\mathbf{z}_p(t) = \begin{bmatrix} z_{1p} \\ z_{2p} \end{bmatrix} a_p(t) = \begin{bmatrix} \cos\alpha_p & \sin\alpha_p \\ -\sin\alpha_p & \cos\alpha_p \end{bmatrix} \begin{bmatrix} \cos\beta_p \\ j\sin\beta_p \end{bmatrix} a_p(t), \tag{24}$$

where z_{1p} and z_{2p} are the signal components on the polarization basis of transmitter, α_p and β_p are the orientation and ellipticity angles of polarization ellipse with $\alpha_p \in [-\pi/2, \pi/2]$ and $\beta_p \in [-\pi/4, \pi/4]$, and $a_p(t)$ ($p = 1, \ldots, P$) is the complex envelope of the pth transmitted signal pulse and each element of $\mathbf{a}_p = [a_p(t_{1p}), \ldots, a_p(t_{Mp})]^T$ ($p = 1, \ldots, P$) with t_{mp} ($m = 1, \ldots, M$) denoting the mth sampling instant within the pth pulse.

The polarization matrix of each diversely polarized pulse ($p = 1, \ldots, P$) is given by

$$\mathbf{E}_p = \begin{bmatrix} z_{1p} & 0 & z_{2p} \\ 0 & z_{2p} & z_{1p} \end{bmatrix}. \tag{25}$$

So the system response matrix can be written as

$$\Sigma = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{V}\mathbf{E}_1 \\ \vdots \\ \mathbf{a}_P \otimes \mathbf{V}\mathbf{E}_P \end{bmatrix}, \tag{26}$$

and matrix Σ has dimension $3MP \times 3$, where P is the number of the transmitted pulses.

The noise covariance matrix is supposed to be $\mathbf{R} = \sigma_n^2(\mathbf{I}_P \otimes \mathbf{C}_{3M\times 3M})$, where σ_n^2 is the noise power of each sample, $\mathbf{I}_P$ denotes the P-dimensional identity matrix, and $\mathbf{C}_{m\times n}$ is Gaussian shaped with one-lag correlation coefficient $\rho_c = 0.9$ [35]. That is to say,

$$\mathbf{C}_{m\times n} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}, \tag{27}$$

where $c_{ij} = 0.9^{(i-j)^2}$, $i = 1, \ldots, m$; $j = 1, \ldots, n$.

Suppose $\mathbf{R}' = \mathbf{I}_P \otimes \mathbf{C}_{3M\times 3M}$ and $\Gamma = (1/\mu\sigma_n^2)\mathbf{s}^H[\Sigma^H(\mathbf{R}')^{-1}\Sigma]\mathbf{s}$. Note that $\mathbf{R}'$ is real symmetric matrix and $(\mathbf{R}')^{-1}$ is also real symmetric matrix. So $(\mathbf{R}')^{-1}$ can be decomposed into $(\mathbf{R}')^{-1} = \mathbf{G}^T\mathbf{G}$ uniquely, where $\mathbf{G}$ is $3MP \times 3MP$ dimension real upper triangular matrix. Therefore, the fitness can be written as

$$\Gamma = \frac{1}{\mu\sigma_n^2}\mathbf{s}^H\Sigma^H\mathbf{G}^T\mathbf{G}\Sigma\mathbf{s} = \frac{1}{\mu\sigma_n^2}(\mathbf{G}\Sigma\mathbf{s})^H\mathbf{G}\Sigma\mathbf{s}. \tag{28}$$

Suppose that $\mathbf{H} = \mathbf{G}\Sigma\mathbf{s}$ and $\mathbf{H}$ is $3MP \times 1$ dimension complex vector. Then finding the maximum value of Γ is equivalent to finding the maximum modulus value of $\mathbf{H}$.

Note that $\mathbf{C}_{3M\times 3M}$ is real symmetric matrix and $\mathbf{C}_{3M\times 3M}^{-1}$ is also real symmetric matrix. So it can be decomposed into $\mathbf{C}_{3M\times 3M}^{-1} = \mathbf{g}^T\mathbf{g}$ uniquely, where $\mathbf{g}$ is $3M \times 3M$ dimension real upper triangular matrix. Now we have an important discovery: $\mathbf{G} = \mathbf{I}_P \otimes \mathbf{g}$.

Proof. From the above analysis, we can get

$$\begin{aligned} &\mathbf{C}_{3M\times 3M}^{-1} = \mathbf{g}^T\mathbf{g} \\ &\Longleftrightarrow \mathbf{I}_P \otimes \mathbf{C}_{3M\times 3M}^{-1} = \mathbf{I}_P \otimes \mathbf{g}^T\mathbf{g} \\ &\Longleftrightarrow \mathbf{I}_P^{-1} \otimes \mathbf{C}_{3M\times 3M}^{-1} = \left(\mathbf{I}_P \otimes \mathbf{g}^T\right)\left(\mathbf{I}_P \otimes \mathbf{g}\right) \\ &\Longleftrightarrow \left(\mathbf{I}_P \otimes \mathbf{C}_{3M\times 3M}\right)^{-1} = \left(\mathbf{I}_P \otimes \mathbf{g}\right)^T\left(\mathbf{I}_P \otimes \mathbf{g}\right). \end{aligned} \tag{29}$$

Due to $(\mathbf{I}_P \otimes \mathbf{C}_{3M\times 3M})^{-1} = \mathbf{G}^T\mathbf{G}$, we can get

$$\mathbf{G} = \mathbf{I}_P \otimes \mathbf{g}. \tag{30}$$

Then $\mathbf{H}$ can be written as

$$\begin{aligned} \mathbf{H} = \mathbf{G}\Sigma\mathbf{s} &= \left(\mathbf{I}_P \otimes \mathbf{g}\right)\Sigma\mathbf{s} \\ &= \operatorname{diag}[\mathbf{g}, \ldots, \mathbf{g}] \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{V}\mathbf{E}_1 \\ \vdots \\ \mathbf{a}_P \otimes \mathbf{V}\mathbf{E}_P \end{bmatrix} \mathbf{s} \\ &= \begin{bmatrix} \mathbf{g}\mathbf{a}_1 \otimes \mathbf{V}\mathbf{E}_1\mathbf{s} \\ \vdots \\ \mathbf{g}\mathbf{a}_P \otimes \mathbf{V}\mathbf{E}_P\mathbf{s} \end{bmatrix} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_P]^T, \end{aligned} \tag{31}$$

where $\mathbf{h}_p = \mathbf{g}\mathbf{a}_p \otimes \mathbf{V}\mathbf{E}_p\mathbf{s}, p = 1, \ldots, P$ is a P-dimensional complex vector group, and each one of them is a $3M \times 1$ dimension complex vector.

It is considered in our system that the polarization parameters of different transmitted signal pulses are independent of each other; that is, when $i \neq j$ $(i = 1, \ldots, P; j = 1, \ldots, P)$ (α_i, β_i) and (α_j, β_j) are independent of each other. Thus, we can get a conclusion that finding the maximum modulus value of $\mathbf{H}$ is equivalently decomposed into finding the maximum modulus value of every vector in the complex vector group: $\mathbf{h}_p, p = 1, \ldots, P$. □

Now we analyze the complex vector group: $\mathbf{h}_p = \mathbf{g}\mathbf{a}_p \otimes \mathbf{V}\mathbf{E}_p\mathbf{s}, p = 1, \ldots, P$, where real upper triangular matrix $\mathbf{g}$ is fixed; when transmitted signal pulses and the sampling form are fixed, the complex envelope of the pth transmitted signal pulse $\mathbf{a}_p$ is fixed; when the target is deterministic, the target reflectivity vector $\mathbf{s}$ is fixed; in the same pulse interval, we assume that the elevation and azimuth angles of the target fixed, that is, $\mathbf{V}$, are fixed. Thus, there are two variable parameters (α_p, β_p) to be optimized in each vector $\mathbf{h}_p, p = 1, \ldots, P$. Therefore, the proposed algorithm is to optimally choose the parameters (α_p, β_p) to meet the maximum modulus value of every vector in the complex vector group: $\mathbf{h}_p, p = 1, \ldots, P$.

The optimization detection algorithm is to find the maximum fitness function value: $\Gamma(\varepsilon) = \mathbf{s}^H[\Sigma^H(\varepsilon)(\mu\mathbf{R})^{-1}\Sigma(\varepsilon)]\mathbf{s}$, and there are $N_1 = 9M^2P^2 + 36MP + 3$ multiplications in the fitness. The proposed algorithm is the equivalently decomposed of previous method [35]. There are p fitness functions: $\mathbf{h}_p = \mathbf{g}\mathbf{a}_p \otimes \mathbf{V}\mathbf{E}_p\mathbf{s}, p = 1, \ldots, P$, and they totally have $N_2 = 9M^2P + 30MP$ multiplications. The multiplication number of proposed method is a linear increasing as the parameters increase, while it is a square increasing in the previous method. From Figure 1 we can see that the proposed method is much more efficient than the previous method.

In a special circumstance, $\lambda_1\mathbf{a}_1 = \lambda_2\mathbf{a}_2 = \cdots = \lambda_P\mathbf{a}_P = \mathbf{a}$ $(\lambda_k \in R, k = 1, \ldots, P)$, that is, $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_P$, are linear correlation, for example, rectangular pulses [34]. Thus, we get a conclusion that finding the maximum modulus value of every vector in the complex vector group: $\mathbf{h}_p, p = 1, \ldots, P$ is degraded equivalent to finding the maximum modulus value of any vector.

6. Experiment Results and Discussions

The experiment results are done by MATLAB program in a PC computer with CPU: inter I3-2100, 3.1 GHz dual-core processor, and 2 GB memory.

6.1. Simulation Results of the Detection Performances. The received data model with a coherent radar in [3] is adopted. We set $Q = 8$, $K = 48$. In order to decrease the computational burden, the probability of false alarm is set to be 0.01.

The performances of the ASD, Rao test, and Wald test operating with three polarimetric channels are compared with that of the single and dual channel detectors in Figure 2.

Table 1: The maximum fitness function values got by two methods ($\times 10^6$).

Methods	P	1	2	3	4
Proposed method	Fitness 1	1.5560	1.5560	1.5560	1.5560
	Fitness 2		1.9668	1.9668	1.9668
	Fitness 3			1.2143	1.2143
	Fitness 4				1.5471
	Sum	1.5560	3.5228	4.7371	6.2842
Previous method	Fitness	1.5560	3.5228	4.7371	6.2842

It is shown in Figure 2 that the more the used polarimetric channels, the better the detection performance. In particular, the three-channel detector significantly outperforms the dual-channel detector due to the exploitation of the HV channelwith higher SNR.

In Figure 3, the probability of detection of the dual-channel detector as a function of SNR for different values of noise level is plotted. As expected, the increase in the value of μ results in a performance loss due to greater noise power received by the radar system.

From Figures 2 and 3 we can see that the detection performance curves of ASD and Rao and Wald tests coincide exactly. For this reason, Figures 2 and 3 are conducted to attest to the coincidence of Rao test, Wald test, and GLRT and illustrate the validity of the expressions for the probabilities of false alarm and detection of Rao test and Wald test with unknown noise covariance matrix structure based on a DPA, in the partially homogeneous environment.

6.2. Simulation Results of Detection Performance Optimization Algorithm. In this section, we validated the analytical performance of the algorithms by computer simulations. In the following simulations, we select $P = 1, 2, 3, 4$; $M = 2$; $\sigma_n^2 = 1/3$ and $\mu = 3$; $\mathbf{s} = [2i, -1i, 0.5]^T$; $\mathbf{a}_1 = [7 + 8i, 8 - 2i]^T$; $\mathbf{a}_2 = [5 + 3i, 6 - 9i]^T$; $\mathbf{a}_3 = [3 + 7i, 4 - 4i]^T$; $\mathbf{a}_4 = [1 + 5i, 2 - 8i]^T$ in normal circumstance.

The analytical solution can be solved by the proposed method theoretically, but the solution procedure is very complex. Therefore, we use Taguchi optimization algorithm to solve this problem.

As shown in Figure 4 and Table 1, the sum of the maximum fitness function values in the proposed method is the same as the maximum fitness function value in the previous method. And from Table 2 we can see that the two methods get the same optimal polarization parameters. There is no doubt that the numerical simulations are conducted to attest to the validity of the above theoretical equivalence relation.

From Table 3 we can see that the proposed method costs less time than the previous method. The numerical simulations confirm the truth that the multiplication number of proposed method is a linear increasing as the parameters increase, while it is a square increasing in the previous method.

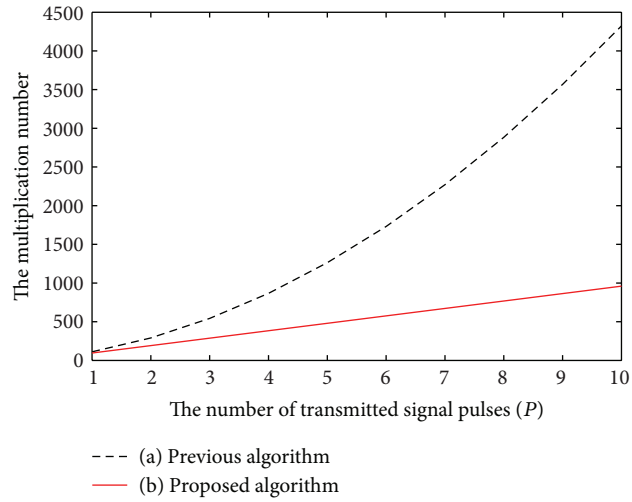


Figure 1: The multiplication numbers of two methods.

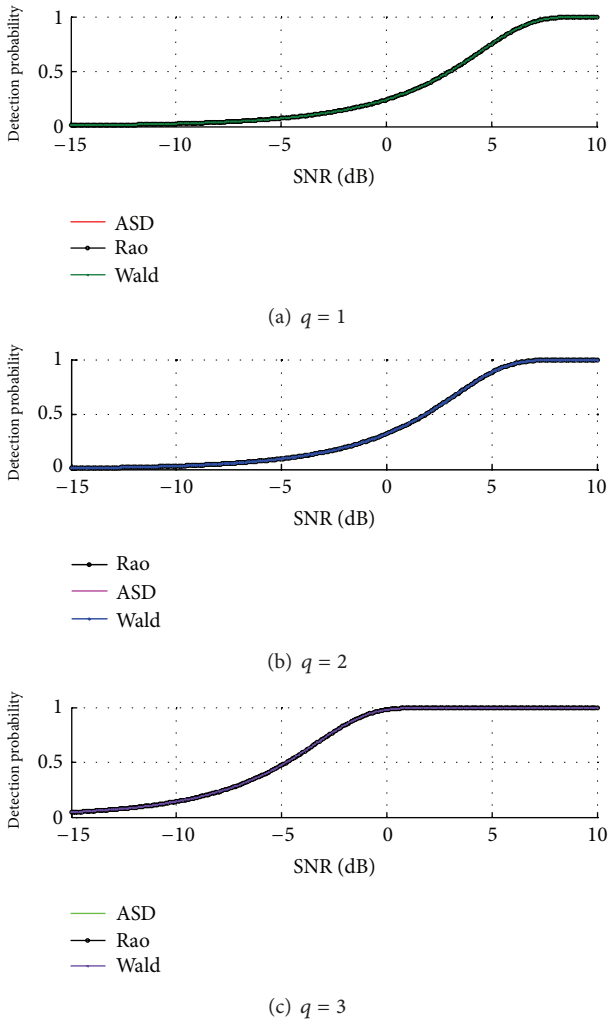


Figure 2: Detection performances of ASD, Rao test and Wald test with different polarimetric channels.

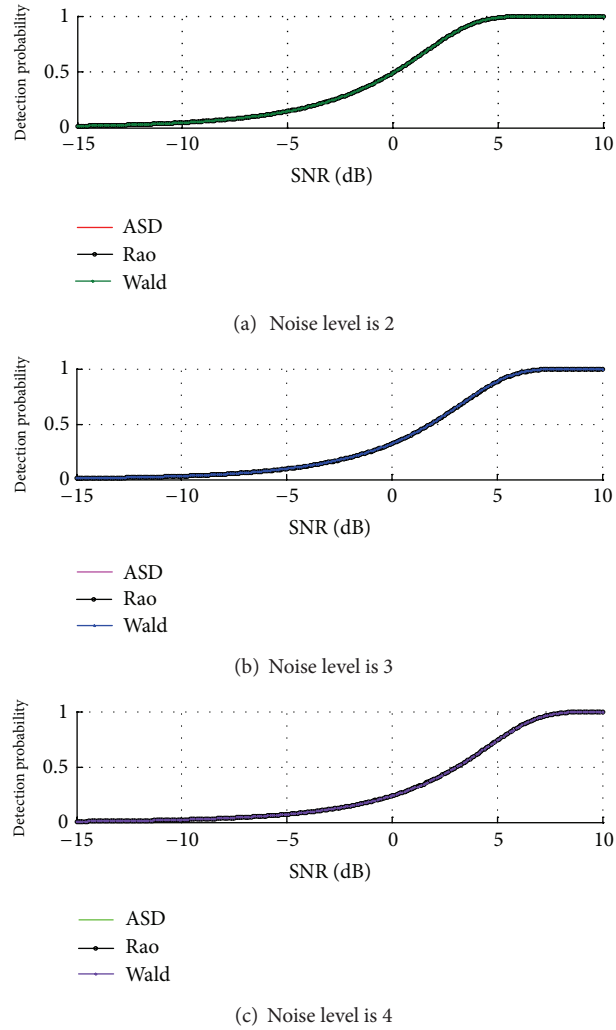
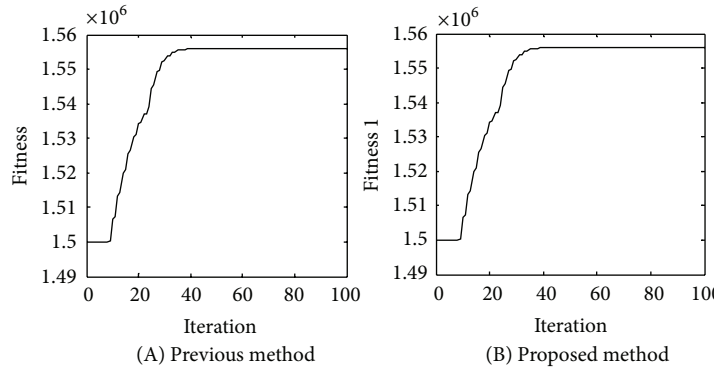
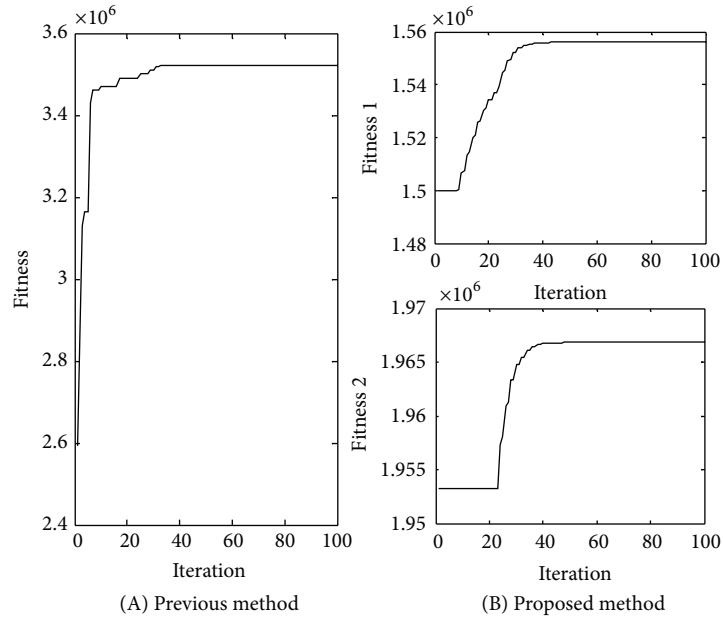


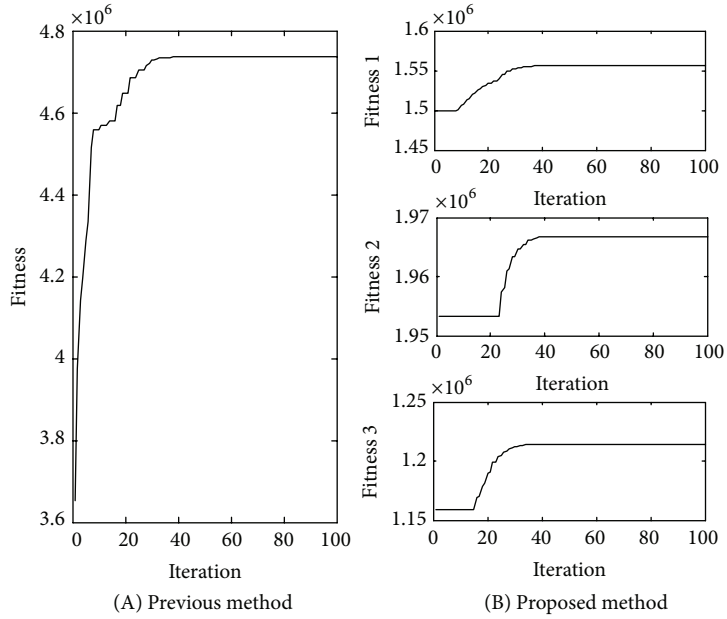
Figure 3: Detection performances of ASD, Rao test and Wald test with $q = 2$ for different noise level $\mu = 2, 3, 4$.



(a) $P = 1$

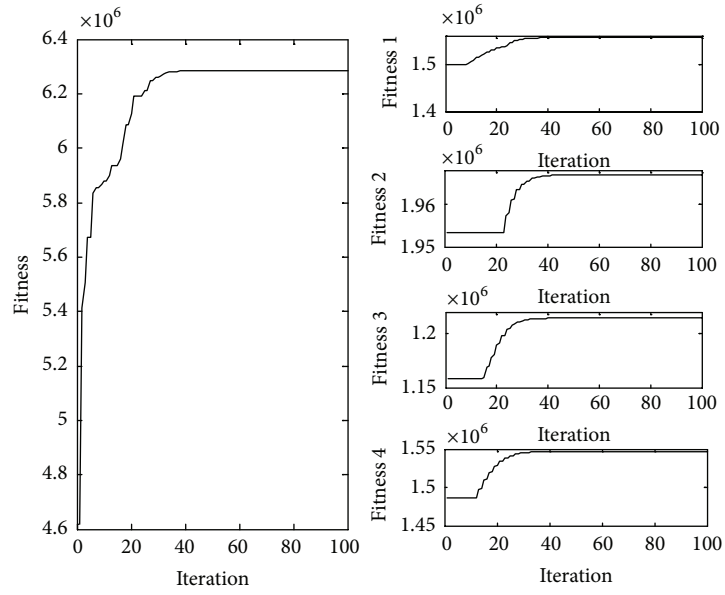


(b) $P = 2$



(c) $P = 3$

Figure 4: Continued.



(d) $P = 4$

Figure 4: The fitness curves of two methods.

Table 2: The optimal polarization parameters got by two methods (rad).

(α_p, β_p)	Previous method	Proposed method
α_1	0.5648	0.5648
β_1	1.7805	1.7805
α_2	0.4070	0.4070
β_2	1.7996	1.7996
α_3	0.5013	0.5013
β_3	2.0538	2.0538
α_4	0.3776	0.3776
β_4	2.1467	2.1467

Table 3: The time cost by two methods (ms).

P	Previous method	Proposed method
1	47.139	43.702
2	164.46	87.175
3	354.16	130.52
4	457.63	178.48

From the above theoretical analysis and simulation experiments we can get a conclusion that the proposed method can get the same detection performance as previous method, but it is more efficient than previous method.

Figure 5 depicts a three-dimensional distribution of modulus value of h_p when the orientation angle α_p and the ellipticity angle β_p of polarization ellipse are valued within the range: $\alpha_p \in [-\pi/2, \pi/2]$ and $\beta_p \in [-\pi/4, \pi/4]$. From Figure 3 we can see the proposed method. Comparing Tables 2 and 3 with Figure 5, we can prove that the proposed algorithm is reliable.

Table 4: The maximum fitness function values got by two methods in a special circumstance ($\times 10^7$).

Methods	P	4
Proposed method	Fitness 1	0.1556
	Fitness 2	0.6224
	Fitness 3	1.4004
	Fitness 4	2.4896
	Sum	4.6880
Previous method	Fitness	4.6880

Here we select $P = 4$; $M = 2$; $\mu = 3$; $\mathbf{s} = [2i, -1i, 0.5]^T$; $\mathbf{a}_1 = [7 + 8i, 8 - 2i]^T$; $\mathbf{a}_2 = 2\mathbf{a}_1$; $\mathbf{a}_3 = 3\mathbf{a}_1$; $\mathbf{a}_4 = 4\mathbf{a}_1$ in special circumstance. From Figure 6, Table 4 and Table 5 we can see that the following experiment results get the same conclusion as Section 5 in detection performance analysis. However, from Table 6 we can see that the efficiency of the proposed method is 9 times more than the efficiency of the previous method. The numerical simulations are conducted to attest to the validity of the above theoretical analysis.

We consider two different cases ($q = 3$, $P = 4$, $\mu = 3$) to illustrate the advantage of our optimization algorithm. In Case 1, the polarization state is fixed. In Case 2, the polarization state can be arbitrarily selected, and we use the proposed optimization algorithm in Case 2. When depicting the detection probability curves of Rao test and Wald test, the probability of false alarm is set to be 10^{-4}, and we choose the statistical data model in [35].

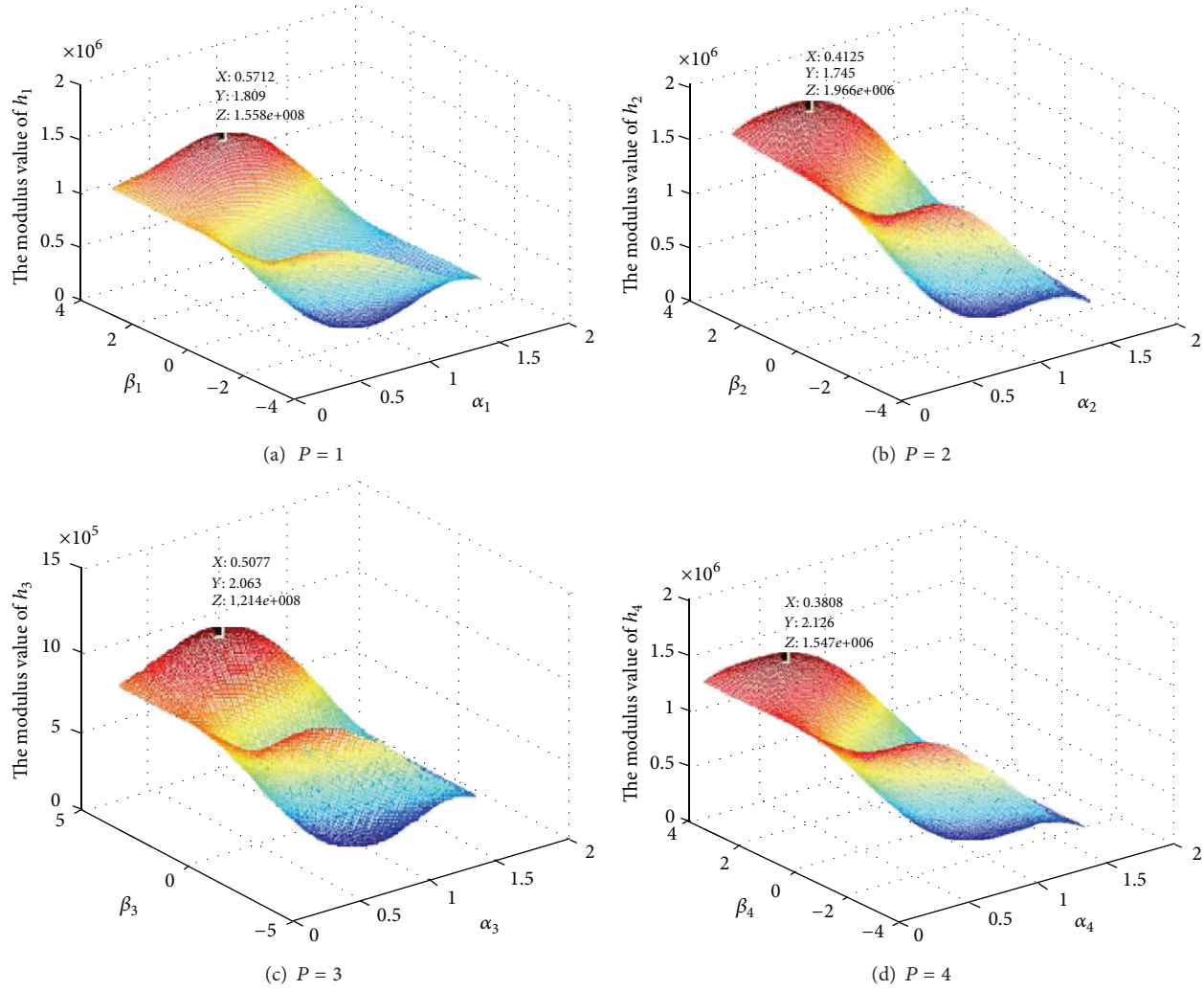


Figure 5: The distribution of modulus value of **h**.

Table 5: The optimal polarization parameters got by two methods in a special circumstance (rad).

(α_p, β_p)	Previous method	Proposed method
α_1	0.5648	0.5648
β_1	1.7805	1.7805
α_2	0.5648	0.5648
β_2	1.7805	1.7805
α_3	0.5648	0.5648
β_3	1.7805	1.7805
α_4	0.5648	0.5648
β_4	1.7805	1.7805

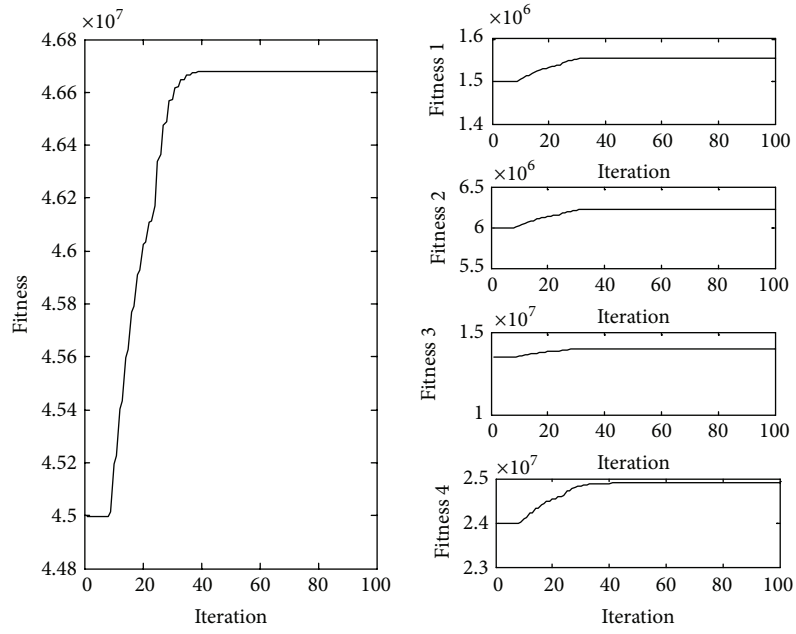
We can see that the detection performance of both the detectors is indeed enhanced by utilizing the proposed algorithm. The gains with respect to the conventional case are approximately 1 dB in Case 2, respectively, when the detection probability is 0.9. Comparing Figures 7(a) and 7(b), we can find that the performances of Rao test and Wald test are exactly the same and it is attested to coincidence of Rao test and Wald test again.

Table 6: The time cost by two methods in a special circumstance (ms).

Previous method	Proposed method
443.72	43.542

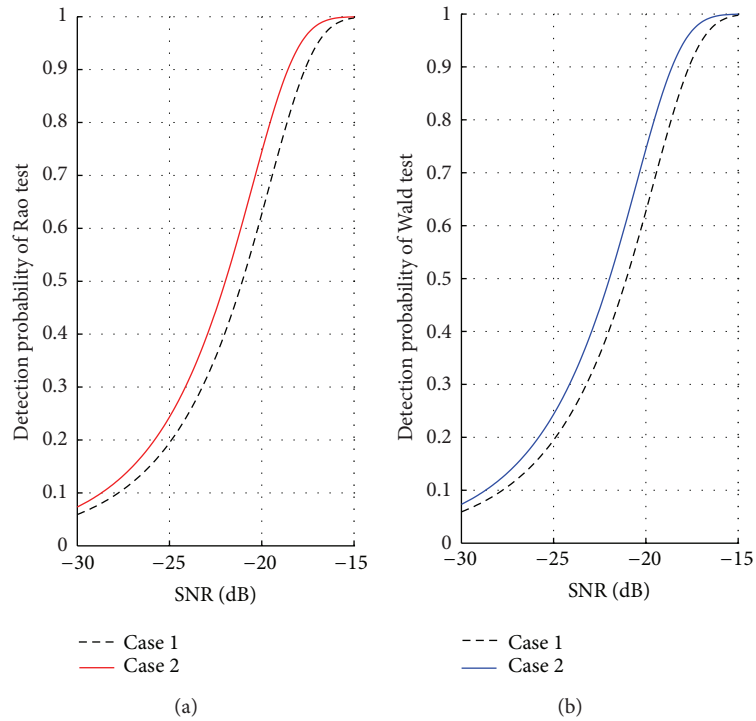
7. Conclusions

In this paper, we study Rao test and Wald test for adaptive detection based on a DPA in partially homogeneous environment. Firstly, we derive the expressions for the probabilities of false alarm and detection of Rao test and Wald test with unknown noise covariance matrix structure based on a DPA, in the partially homogeneous environment. It is found from these derived expressions that Rao test and Wald test have the CFAR property. Secondly, the monotonicities of



(a) Previous method (b) Proposed method

Figure 6: The fitness curves of two methods in special circumstance.



(a) (b)

Figure 7: Performance comparisons of Rao test and Wald test between different cases.

detection probability of the two detectors are proved, and a polarization optimization detection algorithm to enhance the detection performance is proposed. The improvement in the detection performance of the DPA is achieved by using the proposed algorithm to optimally select the polarization of the transmitted pulses. The theoretical analyses and the numerical simulations are conducted to attest to detection of the performance advantage of the proposed optimization algorithm. What is more, the proposed method was much more efficient than the previous method.

Acknowledgments

This work is supported by a Grant from the National Natural Science Fund of China (no. 61172159) and the Fundamental Research Funds for the Central Universities (HEUCFT1101).

References

[1] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 2, pp. 115–127, 1986.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, 1992.

[3] J. Liu, Z. Zhang, Y. Yang, and H. Liu, "A CFAR adaptive subspace detector for first-order or second-order gaussian signals based on a single observation," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5126–5140, 2011.

[4] K. F. McDonald, "Exact performance of STAP algorithms with mismatched steering and clutter statistics," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2750–2763, 2000.

[5] P. Wang, Z. Sahinoglu, M. Pun, H. Li, and B. Himed, "Knowledge-aided adaptive coherence estimator in stochastic partially homogeneous environments," *IEEE Signal Processing Letters*, vol. 18, no. 3, pp. 193–196, 2011.

[6] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Transactions on Signal Processing*, vol. 53, no. 2 I, pp. 427–438, 2005.

[7] L. Cai, Y. She, H. Wang, and T. Luo, "Bayesian detection in partially homogeneous environment with orthogonal rejection," in *Proceedings of the 4th International Conference on Intelligent Control and Information Processing (ICICIP '13)*, pp. 443–446, Beijing, China, June 2013.

[8] O. Besson, L. L. Scharf, and S. Kraut, "Adaptive detection of a signal known only to lie on a line in a known subspace, when primary and secondary data are partially homogeneous," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4698–4705, 2006.

[9] E. Conte, A. De Maio, and G. Ricci, "GLRT-based adaptive detection algorithms for range-spread targets," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1336–1348, 2001.

[10] O. Besson, "Detection of a signal in linear subspace with bounded mismatch," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 3, pp. 1131–1139, 2006.

[11] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Modified GLRT and AMF framework for adaptive detectors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 1017–1051, 2007.

[12] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2157, 1994.

[13] S. Kraut and L. L. Scharf, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, 2001.

[14] M. N. Desai and R. S. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3115–3127, 2003.

[15] S. Bidon, O. Besson, and J. Tourneret, "The adaptive coherence estimator is the generalized likelihood ratio test for a class of heterogeneous environments," *IEEE Signal Processing Letters*, vol. 15, pp. 281–284, 2008.

[16] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2538–2541, 1999.

[17] S. Kraut and L. L. Scharf, "UMP invariance of the multi-rank adaptive coherence estimator," in *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1863–1867, November 2003.

[18] C. R. Rao, "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 44, pp. 50–57, 1948.

[19] A. Wald, "Tests of statistical hypothesis concerning several parameters when the number of observations is large," *Transactions of American Mathematical Society*, vol. 54, pp. 426–482, 1943.

[20] P. M. Baggenstoss and S. M. Kay, "An adaptive detector for deterministic signals in noise of unknown spectra using the Rao test," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1460–1468, 1992.

[21] A. De Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3577–3584, 2007.

[22] A. De Maio, "A new derivation of the adaptive matched filter," *IEEE Signal Processing Letters*, vol. 11, no. 10, pp. 792–793, 2004.

[23] K. J. Sohn, H. Li, and B. Himed, "Parametric GLRT for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5351–5360, 2007.

[24] E. Conte and A. De Maio, "Distributed target detection in compound-Gaussian noise with Rao and Wald tests," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 568–582, 2003.

[25] G. Cui, L. Kong, X. Yang, and J. Yang, "The rao and wald tests designed for distributed targets with polarization MIMO radar in compound-gaussian clutter," *Circuits, Systems, and Signal Processing*, vol. 31, no. 1, pp. 237–254, 2012.

[26] N. Li, G. Cui, L. Kong, and X. Yang, "Rao and Wald tests design of multiple-input multiple-output radar in compound-Gaussian clutter," *IET Radar, Sonar and Navigation*, vol. 6, no. 8, pp. 729–738, 2012.

[27] L. Kong, G. Cui, X. Yang, and J. Yang, "Rao and Wald tests design of polarimetric multiple-input multiple-output radar in compound-Gaussian clutter," *IET Signal Processing*, vol. 5, no. 1, pp. 85–96, 2011.

[28] G. Cui, L. Kong, and X. Yang, "Multiple-input multiple-output radar detectors design in non-Gaussian clutter," *IET Radar, Sonar and Navigation*, vol. 4, no. 5, pp. 724–732, 2010.

[29] A. De Maio and S. Iommelli, "Coincidence of the rao test, wald test, and GLRT in partially homogeneous environment," *IEEE Signal Processing Letters*, vol. 15, pp. 385–388, 2008.

[30] A. De Maio, S. M. Kay, and A. Farina, "On the invariance, coincidence, and statistical equivalence of the GLRT, Rao test, and wald test," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 1967–1979, 2010.

[31] R. T. Compton, "The tripole antenna: an adaptive array with full polarization flexibility," *IEEE Transactions on Antennas and Propagation*, vol. 29, no. 6, pp. 944–952, 1981.

[32] M. Hurtado, J. Xiao, and A. Nehorai, "Target estimation, detection, and tracking: a look at adaptive polarimetric design," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 42–52, 2009.

[33] M. Hurtado and A. Nehorai, "Polarimetric detection of targets in heavy inhomogeneous clutter," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1349–1361, 2008.

[34] J. Liu, Z. Zhang, and Y. Yang, "Optimal waveform design for generalized likelihood ratio and adaptive matched filter detectors using a diversely polarized antenna," *Signal Processing*, vol. 92, no. 4, pp. 1126–1131, 2012.

[35] J. Liu, Z. Zhang, and Y. Yang, "Performance enhancement of subspace detection with a diversely polarized antenna," *IEEE Signal Processing Letters*, vol. 19, no. 1, pp. 4–7, 2012.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 130134, 9 pages
http://dx.doi.org/10.1155/2013/130134

Research Article

An MR Brain Images Classifier System via Particle Swarm Optimization and Kernel Support Vector Machine

Yudong Zhang,[1,2] Shuihua Wang,[1,3] Genlin Ji,[1] and Zhengchao Dong[2]

[1] *School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China*
[2] *Brain Imaging Lab and MRI Unit, New York State Psychiatry Institute and Columbia University, New York, NY 10032, USA*
[3] *School of Electronic Science and Engineering, Nanjing University, Nanjing, Jiangsu 210046, China*

Correspondence should be addressed to Yudong Zhang; zhangyudongnuaa@gmail.com

Received 22 July 2013; Accepted 13 August 2013

Academic Editors: S. Bourennane, C. Fossati, and J. Marot

Automated abnormal brain detection is extremely of importance for clinical diagnosis. Over last decades numerous methods had been presented. In this paper, we proposed a novel hybrid system to classify a given MR brain image as either normal or abnormal. The proposed method first employed digital wavelet transform to extract features then used principal component analysis (PCA) to reduce the feature space. Afterwards, we constructed a kernel support vector machine (KSVM) with RBF kernel, using particle swarm optimization (PSO) to optimize the parameters C and σ. Fivefold cross-validation was utilized to avoid overfitting. In the experimental procedure, we created a 90 images dataset *brain* downloaded from Harvard Medical School website. The abnormal brain MR images consist of the following diseases: glioma, metastatic adenocarcinoma, metastatic bronchogenic carcinoma, meningioma, sarcoma, Alzheimer, Huntington, motor neuron disease, cerebral calcinosis, Pick's disease, Alzheimer plus visual agnosia, multiple sclerosis, AIDS dementia, Lyme encephalopathy, herpes encephalitis, Creutzfeld-Jakob disease, and cerebral toxoplasmosis. The 5-folded cross-validation classification results showed that our method achieved 97.78% classification accuracy, higher than 86.22% by BP-NN and 91.33% by RBF-NN. For the parameter selection, we compared PSO with those of random selection method. The results showed that the PSO is more effective to build optimal KSVM.

1. Introduction

Magnetic resonance imaging (MRI) is an imaging technique that produces high quality images of the anatomical structures of the human body, especially in the brain, and provides rich information for clinical diagnosis and biomedical research. The diagnostic values of MRI are greatly magnified by the automated and accurate classification of the MRI images.

Wavelet transform is an effective tool for feature extraction from MR brain images, because they allow analysis of images at various levels of resolution due to its multiresolution analytic property. However, this technique requires large storage and is computationally expensive [1]. In order to reduce the feature vector dimensions and increase the discriminative power, the principal component analysis (PCA) has been used. PCA is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data [2]. Then, the problem of how to classify on the input data comes.

In recent years, researchers have proposed a lot of approaches for this goal, which fall into two categories. One category is supervised classification, including support vector machine (SVM) [3] and k-nearest neighbors (k-NN) [4]. The other category is unsupervised classification, including self-organization feature map (SOFM) [3] and fuzzy c-means [5]. While all these methods achieved good results, yet the supervised classifier performs better than unsupervised classifier in terms of classification accuracy (success classification rate) [6].

Among supervised classification methods, the SVMs are state-of-the-art classification methods based on machine learning theory [7]. Compared with other methods such as artificial neural network, decision tree, and Bayesian

network, SVMs have significant advantages of high accuracy, elegant mathematical tractability, and direct geometric interpretation. Besides, it does not need a large number of training samples to avoid overfitting [8].

Original SVMs are linear classifiers. In this paper, we introduced in the kernel SVMs (KSVMs), which extends original linear SVMs to nonlinear SVM classifiers by applying the kernel function to replace the dot product form in the original SVMs [9]. The KSVMs is allowed to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear, and the transformed space may be high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space [10].

The structure of the rest of this paper was organized as follows. Section 2 gave the detailed procedures of preprocessing, including the discrete wavelet transform (DWT) and principal component analysis (PCA). Section 3 first introduced the motivation and principles of linear SVM and then extended it to soft margin, dual from. Section 4 introduced the method of PSO-KSVM. It first gave the principles of KSVM and then used the particle swarm optimization algorithm to optimize the values of parameters C and σ; finally it used K-fold cross-validation to protect the classifier from overfitting. The pseudocodes and flowchart were listed. Experiments in Section 5 created a dataset *brain* of 90 brain MR images and showed the results of each step. We compared our proposed PSO-KSVM method with traditional BP-NN and RBF-NN methods. Final Section 6 was devoted to conclusions and discussions.

2. Preprocessing

2.1. Feature Extraction. The most conventional tool of signal analysis is Fourier transform (FT), which breaks down a time domain signal into constituent sinusoids of different frequencies, thus transforming the signal from time domain to frequency domain. However, FT has a serious drawback as discarding the time information of the signal. For example, analyst cannot tell when a particular event took place from a Fourier spectrum. Thus, the classification will decrease as the time information is lost.

Gabor adapted the FT to analyze only a small section of the signal at a time. The technique is called windowing or short-time Fourier transform (STFT) [11]. It adds a window of particular shape to the signal. STFT can be regarded as a compromise between the time information and frequency information. It provides some information about both time and frequency domain. However, the precision of the information is limited by the size of the window.

Wavelet transform (WT) represents the next logical step: a windowing technique with variable size. Thus, it preserves both time and frequency information of the signal. The development of signal analysis is shown in Figure 1.

Another advantage of WT is that it adopts "scale" instead of traditional "frequency," namely, it does not produce a time-frequency view but a time-scale view of the signal. The time-scale view is a different way to view data, but it is a more natural and powerful way.

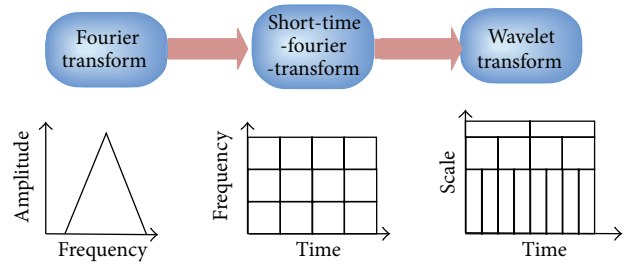


Figure 1: The development of signal analysis.

2.2. Discrete Wavelet Transform. The discrete wavelet transform (DWT) is a powerful implementation of the WT using the dyadic scales and positions. The basic fundamental of DWT is introduced as follows. Suppose that $x(t)$ is a square-integrable function, then the continuous WT of $x(t)$ relative to a given wavelet $\psi(t)$ is defined as

$$W_{\psi}(a,b) = \int_{-\infty}^{\infty} x(t)\,\psi_{a,b}(t)\,dt, \tag{1}$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right). \tag{2}$$

Here, the wavelet $\psi_{a,b}(t)$ is calculated from the mother wavelet $\psi(t)$ by translation and dilation: a is the dilation factor, and b is the translation parameter (both real positive numbers). There are several different kinds of wavelets which have gained popularity throughout the development of wavelet analysis. The most important wavelet is the Harr wavelet, which is the simplest one and often the preferred wavelet in a lot of applications.

Equation (1) can be discretized by restraining a and b to a discrete lattice ($a = 2^b$ & $a > 0$) to give the DWT, which can be expressed as follows:

$$\begin{aligned} \mathrm{ca}_{j,k}(n) &= \mathrm{DS}\left[\sum_{n} x(n)\, g_j^*\left(n - 2^j k\right)\right], \\ \mathrm{cd}_{j,k}(n) &= \mathrm{DS}\left[\sum_{n} x(n)\, h_j^*\left(n - 2^j k\right)\right]. \end{aligned} \tag{3}$$

Here $\mathrm{ca}_{j,k}$ and $\mathrm{cd}_{j,k}$ refer to the coefficients of the approximation components and the detail components, respectively. $g(n)$ and $h(n)$ denote the low-pass filter and high-pass filter, respectively. j and k represent the wavelet scale and translation factors, respectively. DS operator means the downsampling.

The above decomposition process can be iterated with successive approximations being decomposed in turn, so that one signal is broken down into various levels of resolution. The whole process is called wavelet decomposition tree, shown in Figure 2.

2.3. 2D DWT. In case of 2D images, the DWT is applied to each dimension separately. Figure 3 illustrates the schematic

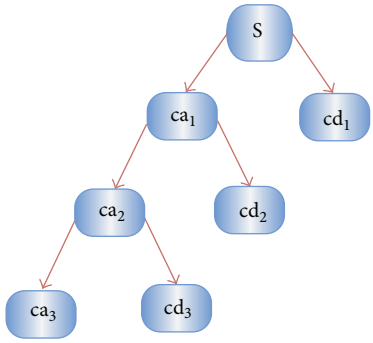


Figure 2: A 3-level wavelet decomposition tree.

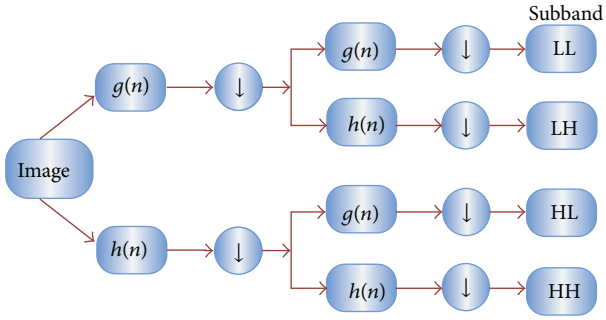


Figure 3: Schematic diagram of 2D DWT.

diagram of 2D DWT. As a result, there are 4 subband (LL, LH, HH, and HL) images at each scale. The sub-band LL is used for the next 2D DWT.

The LL subband can be regarded as the approximation component of the image, while the LH, HL, and HH subbands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level 3 decomposition via Harr wavelet was utilized to extract features.

2.4. Feature Reduction. Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features [12].

PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors, so that it uncorrelated with each other, it orders the resulting orthogonal components, so that those with the largest variation come first, and it eliminates those components contributing the least to the variation in the data set.

It should be noted that the input vectors should be normalized to have zero mean and unity variance before performing PCA. The normalization is a standard procedure. Details about PCA could be seen in [13].

3. SVM Classifier

The introduction of support vector machine (SVM) is a landmark of the field of machine learning [14]. The advantages of SVMs include high accuracy, elegant mathematical tractability and direct geometric interpretation [15]. Recently, multiple improved SVMs have grown rapidly, among which the kernel SVMs are the most popular and effective. Kernel SVMs have the following advantages [16]: (1) work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics, and computer vision; (2) have few tunable parameters; and (3) training often employs convex quadratic optimization [17]. Hence, solutions are global and usually unique, thus avoiding the convergence to local minima exhibited by other statistical learning systems, such as neural networks.

3.1. Principles of Linear SVMs. Given a p-dimensional training dataset of size Nin the form

$$\{(x_n, y_n) \mid x_n \in R^p, y_n \in \{-1, +1\}\}, \quad n = 1, \ldots, N, \tag{4}$$

where y_n is either −1 or 1 corresponding to the class 1 or 2. Each x_n is a p-dimensional vector. The maximum-margin hyperplane which divides class 1 from class 2 is the support vector machine we want. Considering that any hyperplane can be written in the form of

$$\mathbf{w}\mathbf{x} - b = 0, \tag{5}$$

where $\cdot$ denotes the dot product and $\mathbf{w}$ denotes the normal vector to the hyperplane. We want to choose the $\mathbf{w}$ and b to maximize the margin between the two parallel (as shown in Figure 4) hyperplanes as large as possible while still separating the data. So we define the two parallel hyperplanes by the equations as

$$\mathbf{w}\mathbf{x} - b = \pm 1. \tag{6}$$

Therefore, the task can be transformed to an optimization problem. That is, we want to maximize the distance between the two parallel hyperplanes, subject to prevent data falling into the margin. Using simple mathematical knowledge, the problem can be finalized as

$$\begin{aligned} &\min_{\mathbf{w},b} \|\mathbf{w}\| \\ &\text{s.t. } y_n(\mathbf{w}x_n - b) \geq 1, \quad n = 1, \ldots, N. \end{aligned} \tag{7}$$

In practical situations the $\|\mathbf{w}\|$ is usually replaced by

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{s.t. } y_n(\mathbf{w}x_n - b) \geq 1, \quad n = 1, \ldots, N. \end{aligned} \tag{8}$$

The reason leans upon the fact that $\|\mathbf{w}\|$ is involved in a square root calculation. After it is superseded with formula

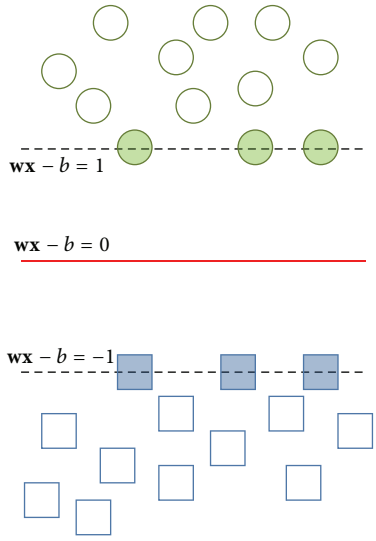


Figure 4: The concept of parallel hyperplanes.

(8), the solution will not change, but the problem is altered into a quadratic programming optimization that is easy to solve by using Lagrange multipliers and standard quadratic programming techniques and programs.

3.2. Soft Margin. However, in practical applications, there may exist no hyperplane that can split the samples perfectly. In such case, the "*soft margin*" method will choose a hyperplane that splits the given samples as clean as possible, while still maximizing the distance to the nearest cleanly split samples.

Positive slack variables ξ_n are introduced to measure the misclassification degree of sample x_n (the distance between the margin and the vectors x_n that lying on the wrong side of the margin). Then, the optimal hyperplane separating the data can be obtained by the following optimization problem:

$$\min_{\mathbf{w},\xi,b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{N}\xi_n$$
$$\text{s.t.} \begin{cases} y_n(\mathbf{w}x_n - b) \geq 1 - \xi_n \\ \xi_n \geq 0, \end{cases} \quad n = 1, \ldots, N, \tag{9}$$

where C is the error penalty. Therefore, the optimization becomes a tradeoff between a large margin and a small error penalty. The constraint optimization problem can be solved using "*Lagrange multiplier*" as

$$\min_{\mathbf{w},\xi,b} \max_{\alpha,\beta} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\alpha_n\left[y_n(\mathbf{w}x_n - b) - 1 + \xi_n\right] - \sum_{n=1}^{N}\beta_n\xi_n \right\}. \tag{10}$$

The min-max problem is not easy to solve, so Cortes and Vapnik proposed a *dual form* technique to solve it.

3.3. Dual Form. The dual form of formula (9) can be designed as

$$\max_{\alpha} \sum_{n=1}^{N}\alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_m\alpha_n y_m y_n k(x_m, x_n),$$
$$\text{s.t.} \begin{cases} 0 \leq \alpha_n \leq C, \\ \sum_{n=1}^{N}\alpha_n y_n = 0, \end{cases} \quad n = 1, \ldots, N. \tag{11}$$

The key advantage of the dual form function is that the slack variables ξ_n vanish from the dual problem, with the constant C appearing only as an additional constraint on the Lagrange multipliers. Now, the optimization problem (11) becomes a *quadratic programming* (QP) problem, which is defined as the optimization of a quadratic function of several variables subject to linear constraints on these variables. Therefore, numerous methods can solve formula (9) within milliseconds, like interior point method, active set method, augmented Lagrangian method, conjugate gradient method, simplex algorithm, and so forth.

4. PSO-KSVM

4.1. Kernel SVMs. Linear SVMs have the downside to linear hyperplane, which cannot separate complicated distributed practical data. In order to generalize it to nonlinear hyperplane, the kernel trick is applied to SVMs [18]. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. In another point of view, the KSVMs allow to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear, and the transformed space may be higher dimensional; thus though the classifier is a hyperplane in the higher-dimensional feature space, it may be nonlinear in the original input space. For each kernel, there should be at least one adjusting parameter so as to make the kernel flexible and tailor itself to practical data. In this paper, RBF kernel is chosen due to its excellent performance. The kernel is written as

$$k(x_m, x_n) = \exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right). \tag{12}$$

Put formula (12) into formula (11), and we got the final SVM training function as

$$\max_{\alpha} \sum_{n=1}^{N}\alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_m\alpha_n y_m y_n \exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right),$$
$$\text{s.t.} \begin{cases} 0 \leq \alpha_n \leq C, \\ \sum_{n=1}^{N}\alpha_n y_n = 0, \end{cases} \quad n = 1, \ldots, N. \tag{13}$$

It is still a quadratic programming problem, and we chose interior point method to solve the problem. However, there is still an outstanding issue, that is, the value of parameters C and σ in (13).

4.2. PSO. To determine the best parameter of C and σ, traditional method uses trial-and-error methods. It will cause heavy computation burden and cannot guarantee to find the optimal or even near-optimal solutions. Fei, W. [19] and Chenglin et al. [20] proposed to use PSO to optimize the parameters, respectively, and independently. The PSO is a populated global optimization method, deriving from the research of the movement of bird flocking or fish schooling. It is easy and fast to implement. Besides, we introduced in the cross-validation to construct the fitness function used for PSO.

PSO performs search via a swarm of particles which is updated from iteration to iteration. To seek for the optimal solution, each particle moves in the direction of its previously best position (p_{best}) and the best global position in the swarm (g_{best}) as follows:

$$\begin{aligned} &p_{\text{best}_i} = p_i\left(k^*\right) \\ &\text{s.t. fitness}\left(p_i\left(k^*\right)\right) = \min_{k=1,\ldots,t}\left[\text{fitness}\left(p_i\left(k\right)\right)\right], \\ &g_{\text{best}} = p_{i^*}\left(k^*\right) \\ &\text{s.t. fitness}\left(p_{i^*}\left(k^*\right)\right) = \min_{\substack{i=1,\ldots,P \\ k=1,\ldots,t}}\left[\text{fitness}\left(p_i\left(k\right)\right)\right], \end{aligned} \tag{14}$$

where i denotes the particle index, P denotes the total number of particles, k denotes the iteration index, and t denotes the current iteration number, and p denotes the position. The velocity and position of particles can be updated by the following equations:

$$\begin{aligned} v_i\left(t+1\right) &= wv_i\left(t\right) + c_1 r_1\left(p_{\text{best}_i}\left(t\right) - p_i\left(t\right)\right) \\ &\quad + c_2 r_2\left(g_{\text{best}}\left(t\right) - p_i\left(t\right)\right), \\ p_i\left(t+1\right) &= p_i\left(t\right) + v_i\left(t+1\right), \end{aligned} \tag{15}$$

where v denotes the velocity. The inertia weight w is used to balance the global exploration and local exploitation. The r_1 and r_2 are uniformly distributed random variables within range $(0, 1)$. The c_1 and c_2 are positive constant parameters called "acceleration coefficients." Here, the particle encoding is composed of the parameters C and σ in (13).

4.3. Cross-Validation. In this paper we choose 5-fold considering the best compromise between computational cost and reliable estimates. The dataset is randomly divided into 5 mutually exclusively subsets of approximately equal size, in which 4 subsets are used as training set, and the last subset is used as validation set. The abovementioned procedure repeated 5 times, so each subset is used once for validation. The fitness function of PSO chose the classification accuracy of the 5-fold cross-validation:

$$\text{fitness} = \frac{1}{5}\sum_{i=1}^{5}\left|\frac{y_s}{y_s + y_m}\right|. \tag{16}$$

Here y_s and y_m denote the number of successful classification and misclassification, respectively. PSO is performed to maximize the fitness function (classification accuracy).

4.4. Pseudocodes of Our Method. In total, our method can be described as the following three stages, and the flowchart is depicted in Figure 5.

Step 1: Collecting MR brain images dataset.

Step 2: Preprocessing (including feature extraction and feature reduction).

Step 3: Fivefolded cross-validation.

Step 4: Determining the best parameter.

- Step 4.1: Initializing PSO. The particles correspond to C and σ.
- Step 4.2: For each particle i, computer the fitness values.
  - Step 4.2.1: Decoding the particle to parameters C and σ.
  - Step 4.2.2: Using interior method to train KSVM according to (13).
  - Step 4.2.3: Calculating classification error according to (16) as the fitness values.
- Step 4.3: Updating the g_{best} and p_{best} according to (14).
- Step 4.4: Updating the velocity and position of each particle according to (15).
- Step 4.5: If stopping criteria is met, then jump to Step 4.6; otherwise return to Step 4.2.
- Step 4.6: Decoding the optimal particle to corresponding parameter C^* and σ^*.

Step 5: Constructing KSVM via the optimal C^* and σ^* according to (13).

Step 6: Submitting new MRI brains to the trained KSVM and outputting the prediction.

5. Experiments and Discussions

The experiments were carried out on the platform of P4 IBM with 3.3 GHz processor and 2 GB RAM, running under Windows XP operating system. The algorithm was in-house developed via the wavelet toolbox, the biostatistical toolbox of 32 bit MATLAB 2012a (the MathWorks). The programs can be run or tested on any computer platforms where MATLAB is available.

5.1. Database. The datasets *brain* consists of 90 T2-weighted MR brain images in axial plane and 256 × 256 in-plane resolution, which were downloaded from the website of Harvard Medical School (URL: http://www.med.harvard.edu/aanlib/home.html). The abnormal brain MR images of the dataset consist of the following diseases: glioma, metastatic adenocarcinoma, metastatic bronchogenic carcinoma, meningioma, sarcoma, Alzheimer, Huntington, motor neuron disease, cerebral calcinosis, Pick's disease, Alzheimer plus visual agnosia, multiple sclerosis, AIDS dementia, Lyme encephalopathy, herpes encephalitis, Creutzfeld-Jakob disease, and cerebral toxoplasmosis. The samples of each disease are illustrated in Figure 6.

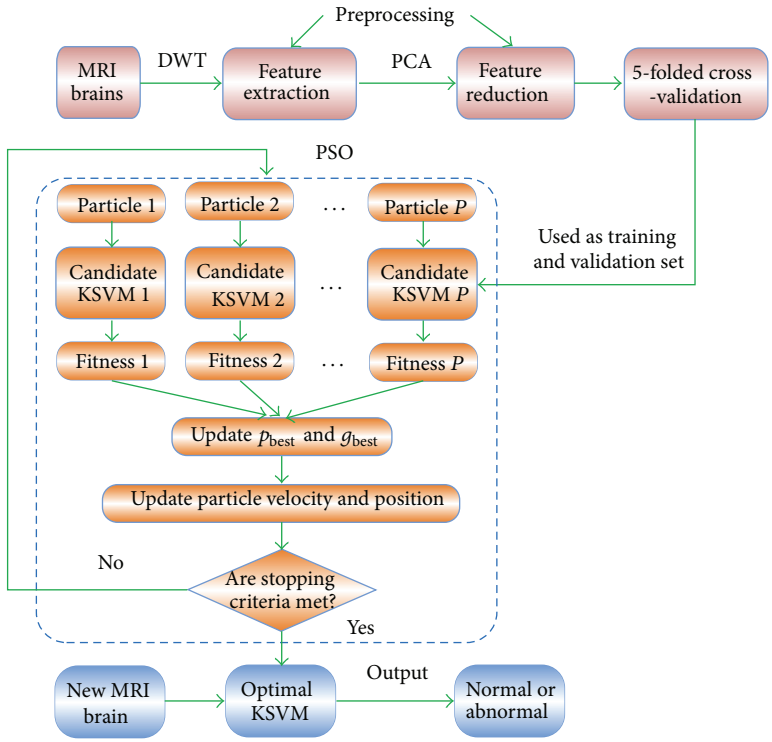


Figure 5: Methodology of our proposed PSO-KSVM algorithm.

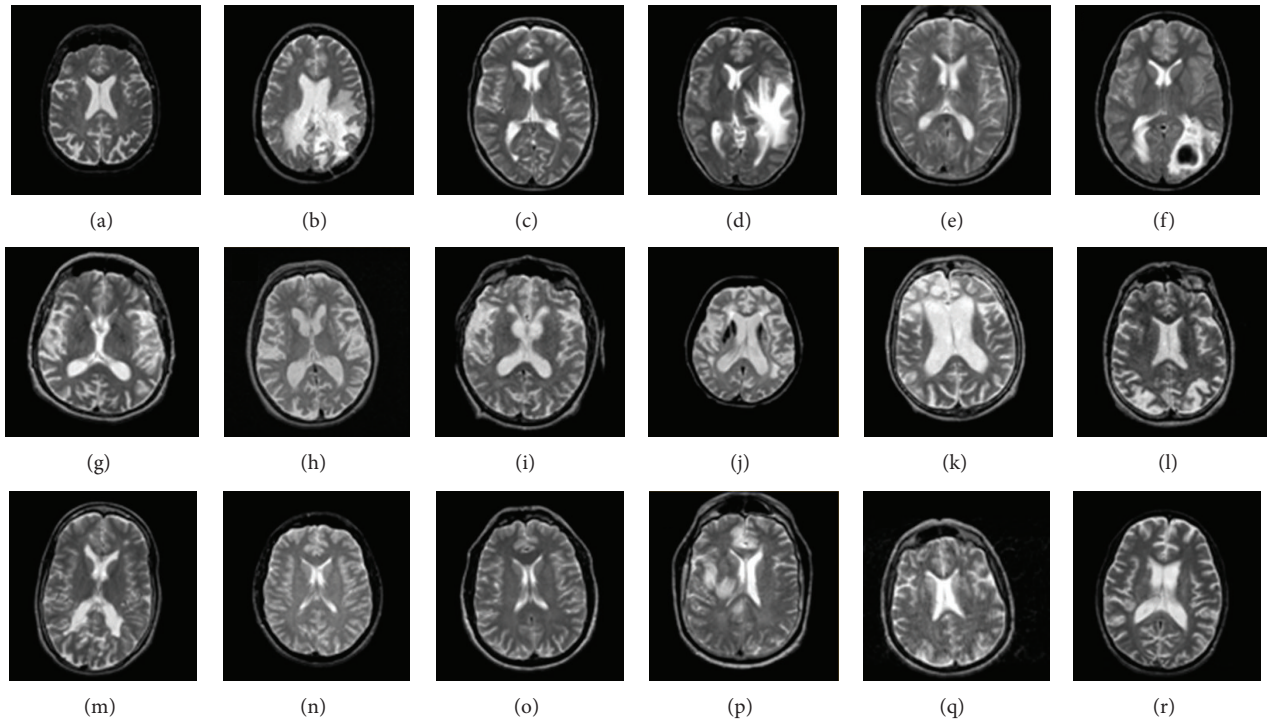


Figure 6: Sample of brain MRIs: (a) normal brain; (b) glioma, (c) metastatic adenocarcinoma; (d) metastatic bronchogenic carcinoma; (e) meningioma; (f) sarcoma; (g) Alzheimer; (h) Huntington; (i) motor neuron disease; (j) cerebral calcinosis; (k) Pick's disease; (l) Alzheimer plus visual agnosia; (m) multiple sclerosis; (n) AIDS dementia; (o) Lyme encephalopathy; (p) herpes encephalitis; (q) Creutzfeld-Jakob disease; and (r) cerebral toxoplasmosis.

Table 1: Detailed data of PCA.

Number of prin. comp.	1	2	3	4	5	6	7	8	9	10
variance (%)	32.81	44.53	52.35	57.71	61.97	65.02	67.78	70.18	72.46	74.56
Number of prin. comp.	11	12	13	14	15	16	17	18	19	20
variance (%)	76.41	78.2	79.7	81.12	82.27	83.38	84.35	85.29	86.05	86.8
Number of prin. comp.	21	22	23	24	25	26	27	28	29	30
variance (%)	87.53	88.2	88.8	89.35	89.86	90.35	90.84	91.3	91.73	92.15
Number of prin. comp.	31	32	33	34	35	36	37	38	**39**	40
variance (%)	92.54	92.9	93.24	93.58	93.9	94.21	94.5	94.76	**95.02**	95.27

Table 2: Methods of comparison between BP-NN, RBF-NN, and PSO-KSVM.

Method	Confusion matrix	Success Cases	Sensitivity	Specificity	Classification accuracy
BP-NN	374 11 51 14	388	88.0%	56%	86.22%
RBF-NN	393 7 32 18	411	92.47%	72%	91.33%
PSO-KSVM	417 2 8 23	440	98.12%	92%	97.78%

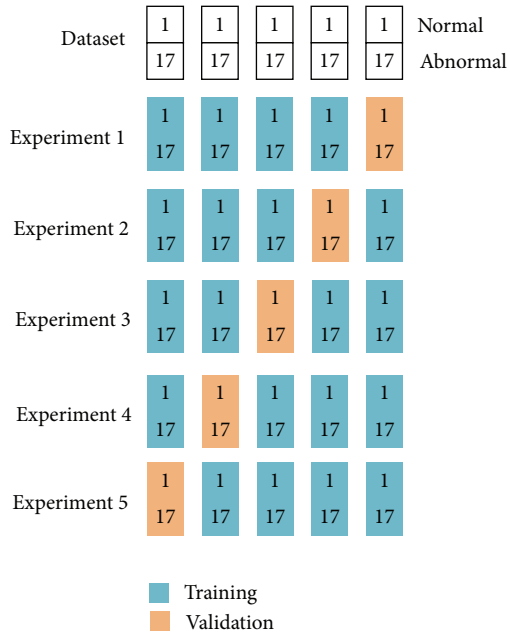


Figure 7: Illustration of 5-fold cross-validation of *brain* dataset (we divided the dataset into 5 groups, and for each experiment, 4 groups were used for training, and the rest one group was used for validation. Each group was used once for validation).

We randomly selected 5 images for each type of brain. Since there are 1 type of normal brain and 17 types of abnormal brain in the dataset, $5^*(1 + 17) = 90$ images were selected to construct the *brain* dataset, consisting of 5 normal and 85 abnormal brain images in total.

The setting of the training images and validation images was shown in Figure 7. We divided the dataset into 5 equally distributed groups; each groups contain one normal brain and 17 abnormal brains. Since 5-fold cross-validation was used, we would perform 5 experiments. In each experiment, 4 groups were used for training, and the left 1 group was used for validation. Each group was used once for validation. In total, in this cross validation way, 360 images were for training, and 90 images were for validation.

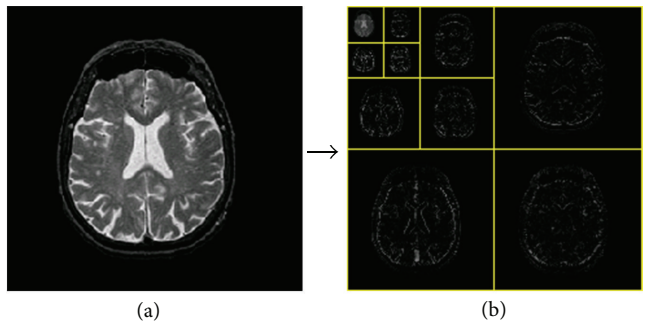


Figure 8: The procedures of 3-level 2D DWT: (a) normal brain MRI; (b) level 3 wavelet coefficients.

5.2. Feature Extraction. The three levels of wavelet decomposition greatly reduce the input image size as shown in Figure 8. The top left corner of the wavelet coefficients image denotes for the approximation coefficients at level 3, of which the size is only $32 \times 32 = 1024$. The border distortion should be avoided. In our algorithm, symmetric padding method [21] was utilized to calculate the boundary value.

5.3. Feature Reduction. As stated above, the extracted features were reduced from 65536 to 1024 by the DWT procedure. However, 1024 was still too large for calculation. Thus, PCA was used to further reduce the dimensions of features. The

Table 3: Parameters of comparison by random selection method (the final row corresponds to our proposed method).

	σ	C	Success case	Classification accuracy
Random 1	0.625	124.71	410	91.11%
Random 2	1.439	185.13	412	91.56%
Random 3	1.491	136.20	423	94.00%
Random 4	1.595	176.78	409	90.89%
Random 5	1.836	160.80	401	89.11%
Random 6	1.973	137.90	401	89.11%
Random 7	1.654	87.01	396	88.00%
Random 8	1.372	149.96	427	94.89%
Optimized	1.132	143.3	440	97.78%

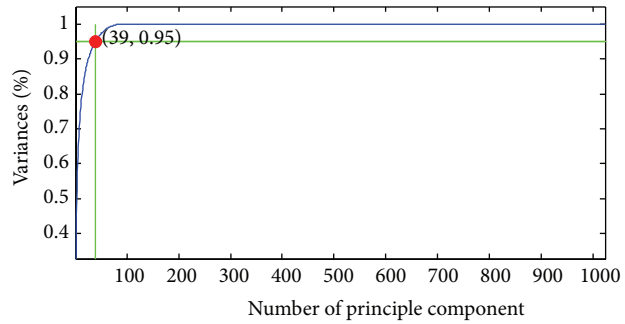


Figure 9: The curve of variances against number of principle components (here we found that 39 features can achieve 95.02% variances).

curve of cumulative sum of variance versus the number of principle components was shown in Figure 9.

The variances versus the number of principle components from 1 to 40 were listed in Table 1. It showed that only 39 principle components (bold font in table), which were only 39/1024 = 3.81% of the original features, could preserve 95.02% of total variance.

5.4. Classification Accuracy. The KSVM used the RBF as the kernel function. We compared our PSO-KVSM method with one hidden-layer Back Propagation-Neural Network (BP-NN) and RBF-Neural Network (RBF-NN). The results were shown in Table 2. It showed that BP-NN correctly matched 388 cases with 86.22% classification accuracy. RBF-NN correctly matched 411 cases with 91.33% classification accuracy. Our PSO-KSVM correctly matched 440 brain images with 97.78% classification accuracy. Therefore, our method had the most excellent classification performance.

5.5. Parameter Selection. The final parameters obtained by PSO were $C = 143.3$ and $\sigma = 1.132$. We compared this case with random selection method, which randomly generated the values of C in the range of $(50, 200)$ and σ in the range of $[0.5, 2]$, and then we compared them with the optimized values by PSO ($C = 143.3$ and $\sigma = 1.132$). The results achieved by random selection method were shown in Table 3. We saw that the classification accuracy varied with the change of parameters σ and C, so it was important to determine the optimal values before constructing the classifier. The random selection method was difficult to come across the best values, so PSO was an effective method for this problem compared to random selection method.

6. Conclusions and Discussions

In this study we had developed a novel DWT + PCA + PSO-KSVM hybrid classification system to distinguish between normal and abnormal MRIs of the brain. We picked up RBF as the kernel function of SVM. The experiments demonstrated that the PSO-KSVM method obtained 97.78% classification accuracy on the 5-folded 90-image dataset, higher than 86.22% of BP-NN and 91.33% of RBF-NN.

Future work should focus on the following four aspects. First, the proposed SVM based method could be employed for MR images with other contrast mechanisms such as T1-weighted, proton density weighted, and diffusion weighted images. Second, the computation time could be accelerated by using advanced wavelet transforms such as the lift-up wavelet. Third, Multiclassification, which focuses on brain MRIs of specific disorders, can also be explored. Forth, novel kernels will be tested to increase the classification accuracy and accelerate the algorithm.

The DWT can efficiently extract the information from original MR images with litter loss. The advantage of DWT over Fourier transforms is the spatial resolution; namely, DWT captures both frequency and location information. In this study we choose the Harr wavelet, although there are other outstanding wavelets such as Daubechies series. We will compare the performance of different families of wavelet in future work. Another research direction lies in the stationary wavelet transform and the wavelet packet transform.

The importance of PCA was demonstrated in the Discussion. If we omitted the PCA procedures, we meet a huge feature space (1024 dimensions) which will cause heavy computation burden and lowered the classification accuracy. There are some other excellent feature reduction methods such as ICA, manifold learning. In the future, we will focus on investigating the performance of those algorithms.

The reason we choose RBF kernel is that RBF takes the form of exponential function, which enlarge the sample distances to the uttermost extent. In the future, we will try to test other kernel functions.

The importance of introducing PSO is to determine the optimal values of parameters C and σ. From random selection method, we found it is hard to get the optimal values at the parameter space. Therefore, the PSO is an effective way to find the optimal values. Integrating PSO to KSVM enhance the classification capability of KSVM.

The most important contribution of this paper is the propose of a hybrid system, integrating DWT, PCA, PSO, KSVM, and CV, used for identifying normal MR brains from abnormal MR brains. It would be useful to help clinicians to diagnose the patients.

References

[1] M. Emin Tagluk, M. Akin, and N. Sezgin, "Classification of sleep apnea by using wavelet transform and artificial neural networks," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1600–1607, 2010.

[2] J. Camacho, J. Picó, and A. Ferrer, "'The best approaches in the on-line monitoring of batch processes based on PCA: does the modelling structure matter?' [Anal. Chim. Acta Volume 642 (2009) 59–68]," *Analytica Chimica Acta*, vol. 658, no. 1, p. 106, 2010.

[3] S. Chaplot, L. M. Patnaik, and N. R. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86–92, 2006.

[4] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," *Medical Image Analysis*, vol. 7, no. 4, pp. 513–527, 2003.

[5] J.-Y. Yeh and J. C. Fu, "A hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1285–1295, 2008.

[6] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors*, vol. 12, no. 9, pp. 12489–12505, 2012.

[7] N. S. Patil, P. S. Shelokar, V. K. Jayaraman, and B. D. Kulkarni, "Regression models using pattern search assisted least square support vector machines," *Chemical Engineering Research and Design A*, vol. 83, no. 8, pp. 1030–1037, 2005.

[8] D. Li, W. Yang, and S. Wang, "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine," *Computers and Electronics in Agriculture*, vol. 74, no. 2, pp. 274–279, 2010.

[9] T. A. F. Gomes, R. B. C. Prudêncio, C. Soares, A. L. D. Rossi, and A. Carvalho, "Combining meta-learning and search techniques to select parameters for support vector machines," *Neurocomputing*, vol. 75, pp. 3–13, 2012.

[10] R. Hable, "Asymptotic normality of support vector machine variants and other regularized kernel methods," *Journal of Multivariate Analysis*, vol. 106, pp. 92–117, 2012.

[11] L. Durak, "Shift-invariance of short-time Fourier transform in fractional Fourier domains," *Journal of the Franklin Institute*, vol. 346, no. 2, pp. 136–146, 2009.

[12] Y. Zhang, S. Wang, Y. Huo, L. Wu, and A. Liu, "Feature extraction of brain MRI by stationary wavelet transform and its applications," *Journal of Biological Systems*, vol. 18, no. 1, pp. 115–132, 2010.

[13] Y. Zhang, L. Wu, and G. Wei, "A new classifier for polarimetric SAR images," *Progress in Electromagnetics Research*, vol. 94, pp. 83–104, 2009.

[14] "Support vector machine," 2012, http://en.wikipedia.org/wiki/Support_vector_machine.

[15] P. Martiskainen, M. Järvinen, J. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, "Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines," *Applied Animal Behaviour Science*, vol. 119, no. 1-2, pp. 32–38, 2009.

[16] S. Bermejo, B. Monegal, and J. Cabestany, "Fish age categorization from otolith images using multi-class support vector machines," *Fisheries Research*, vol. 84, no. 2, pp. 247–253, 2007.

[17] A. M. S. Muniz, H. Liu, K. E. Lyons et al., "Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in Parkinson disease on ground reaction force during gait," *Journal of Biomechanics*, vol. 43, no. 4, pp. 720–726, 2010.

[18] J. Acevedo-Rodríguez, S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Siegmann, and F. López-Ferreras, "Computational load reduction in decision functions using support vector machines," *Signal Processing*, vol. 89, no. 10, pp. 2066–2071, 2009.

[19] S.-W. Fei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Systems with Applications*, vol. 37, no. 10, pp. 6748–6752, 2010.

[20] Z. Chenglin, S. Xuebin, S. Songlin, and J. Ting, "Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9908–9912, 2011.

[21] A. Messina, "Refinements of damage detection methods based on wavelet analysis of dynamical shapes," *International Journal of Solids and Structures*, vol. 45, no. 14-15, pp. 4068–4097, 2008.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 917928, 7 pages
http://dx.doi.org/10.1155/2013/917928

Research Article

Aircraft Detection from VHR Images Based on Circle-Frequency Filter and Multilevel Features

Feng Gao,[1,2] Qizhi Xu,[1,2] and Bo Li[1,2]

[1] *Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China*
[2] *State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Bo Li; boli@buaa.edu.cn

Received 16 July 2013; Accepted 7 August 2013

Academic Editors: S. Bourennane, C. Fossati, and J. Marot

Aircraft automatic detection from very high-resolution (VHR) images plays an important role in a wide variety of applications. This paper proposes a novel detector for aircraft detection from very high-resolution (VHR) remote sensing images. To accurately distinguish aircrafts from background, a circle-frequency filter (CF-filter) is used to extract the candidate locations of aircrafts from a large size image. A multi-level feature model is then employed to represent both local appearance and spatial layout of aircrafts by means of Robust Hue Descriptor and Histogram of Oriented Gradients. The experimental results demonstrate the superior performance of the proposed method.

1. Introduction

With the advent of the very high-resolution (VHR) earth observation satellite programs, the spatial resolution of remote sensing images dramatically increased from tens of meters to tens of decimeters, such as QuickBird, GeoEye-1, and WorldView-2. High-resolution images of a small number of locations are publicly available via Google Earth at an astonishing ground sampling distance (GSD) of 0.15 meter [1]. The high-resolution satellite sensors can acquire high definition images of ground objects with abundant spatial details and contextual information, which make it feasible to detect and recognize artificial targets, such as aircraft [2, 3], vehicle [1, 4, 5], ship [6], and building [7, 8]. In this paper, we concentrate on the problem of detecting aircrafts from such high-resolution aerial and satellite imagery. Aircraft automatic detection from VHR images plays an important role in a wide variety of applications. Detecting and tracking aircrafts in aerial videos is an important component in visual surveillance systems. Images of military airport, along with the distribution of aircrafts, can provide valuable information for military monitoring. However, while image resolution upgrades to decimeter level, background of ground targets becomes complex and disturbing. It is therefore a challenging task to detect ground targets in a cluttered background. The presence of complex structures, such as buildings and airport terminals, can cause many false alarms.

Various methods have been developed for object detection from remote sensing images. For example, Lei et al. [9] proposed a color-enhanced rotation-invariant Hough forest method for detecting aircrafts and buildings. Liu et al. [2] presented a coarse-to-fine method by integrating shape prior; the pose of an aircraft is roughly estimated by template matching, and a shape prior is used to segment the target. Li et al. [10] proposed detecting aircrafts using a contour-based spatial model. Inglada [11] proposed a supervised learning method which used geometric image features to characterize classes of objects. Akçay and Aksoy [12] presented a geospatial object detection method by using hierarchical segmentation which combined spectral and structural information. Recently, there are also some methods concentrating on detecting objects in synthetic aperture radar (SAR) image [13, 14] and hyperspectral image [15, 16].

Another group of methods utilizs a sliding window approach. In this approach, a classifier is first trained to recognize an object; a detection window is scanned over

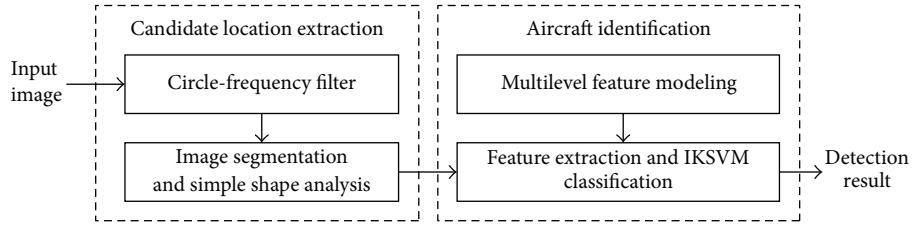


Figure 1: Flow diagram of the proposed method.

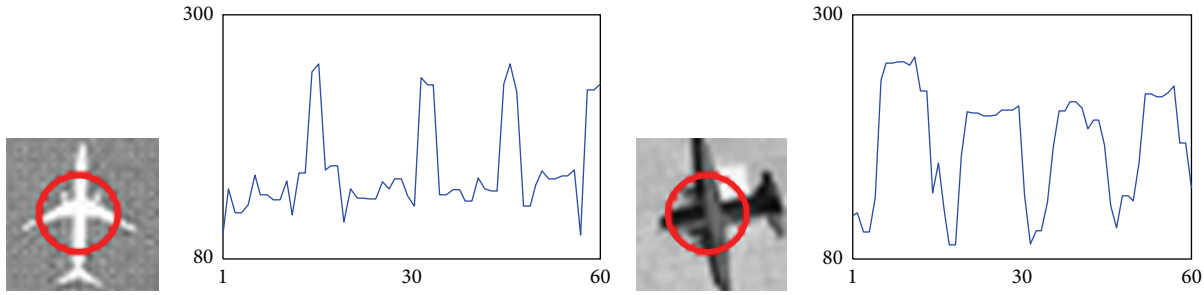


Figure 2: Illustration of the aircraft regular patterns. (a) Bright aircraft (32 × 32 pixels); (b) regular pattern of the bright aircraft; (c) dark aircraft (32 × 32 pixels); (d) regular pattern of the dark aircraft. The intensities along the red circle will change regularly from darkness to brightness and will repeat four times.

the image; then the classifier is used to determine whether the detection window contains an object or not. Kembhavi et al. [1] proposed a vehicle detector by using partial least square analysis to project high-dimensional features to a much lower dimensional subspace, and then the computational burden can be reduced. Sun et al. [17] presented an aircraft detector which applied a mapping strategy to encode the geometric information in the detection window. Zhang et al. [18] proposed a method to detect aircrafts by encoding the features of the rotated parts and objects. These methods rely on intensity-based features to capture the appearance of the object and obtain good results. However, all the pixels contained in the image are generally searched in these methods, thus resulting in a number of false positive detections. Moreover, due to the variability of color in a cluttered background, color cues have been ignored by most of these detection methods.

The proposed method improves the existing methods by exploiting circle-frequency filter (CF-filter) [3] to extract potential location of aircraft targets from the entire image. Consequently, most of the cluttered background regions are eliminated, and false alarms are therefore greatly reduced. Moreover, a multilevel feature model that incorporates both color and shape information is designed to identify aircrafts by checking potential locations. Our method includes a coarse-to-fine search process. The flow diagram of our method is shown in Figure 1.

Stage 1 (candidate location extraction by circle-frequency filter). Inspired by [3], CF-filter is applied to extract the candidate locations of aircraft. The goal of candidate extraction is to reduce the false positives, for example, man-made objects that may be detected as aircrafts, and false negatives.

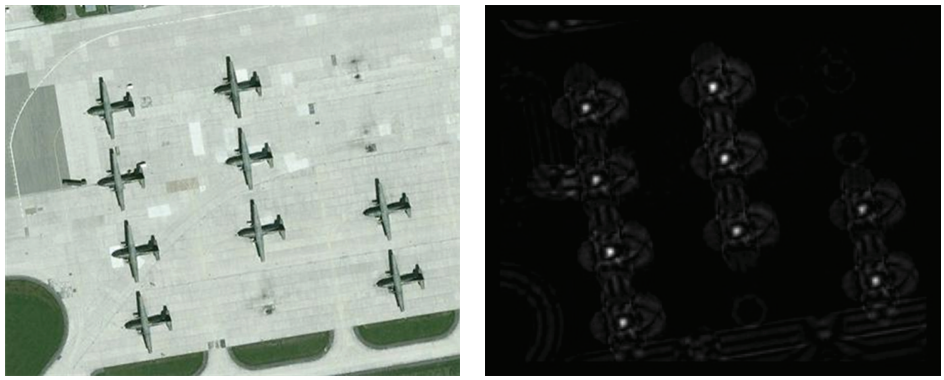
Stage 2 (aircraft identification by multilevel features). Robust Hue Descriptor [19] and Histogram of Oriented Gradients [20] are used to capture color and shape information of aircrafts, respectively. Meanwhile, a multilevel feature model is constructed to represent both local appearance and spatial layout of aircrafts.

We implement the proposed method on remote sensing images collected from aerial and WorldView-2 satellite images. Our methods are also compared with two previously proposed object detection methods, and experimental results demonstrate the superior performance of the proposed method.

2. Extraction of Aircraft Candidate Locations

The candidate location extraction is comprised of two steps: circle-frequency filter (CF-filter) and candidate locations extraction. To facilitate discussion, we begin this section with a general introduction to CF-filter. We follow that with an introduction on how candidate locations can be extracted.

2.1. Circle-Frequency Filter. The candidate extraction is based on two common features of aircraft as follows. First, airplanes and their background have difference in brightness. Second, airplanes are comprised by four main bulges [3], that is, head, tail, and two wings. If a circle is located at the center of an aircraft, the intensities along the circle will change



(a) (b)

Figure 3: Illustration of CF-filter. (a) Original image; (b) CF-filtered result.

regularly from darkness to brightness and will repeat four times. Figures 2(a) and 2(c) are two examples of typical airplanes, that is, civil and military aircraft. In general, civil aircraft is brighter than background, while military aircraft is darker than background. Figures 2(b) and 2(d) illustrate the regular patterns of the two aircrafts.

Assume that f_i $(i = 0, \ldots, N-1)$ are pixel values along the circle centered at (x, y) with radius r, the CF-filter can be expressed as

$$f(x,y) = \left(\sum_{k=0}^{N-1} f_k \cos\left(8\pi\frac{k}{N}\right)\right)^2 + \left(\sum_{k=0}^{N-1} f_k \sin\left(8\pi\frac{k}{N}\right)\right)^2. \quad (1)$$

If the whole image is filtered by CF-filter, a bright spot will be generated at the center of an airplane. Figure 3 gives an example of a CF-filtered image, where the values of CF-filtered image are normalized to (0, 255). It can be seen from Figure 3 that only the center of airplane has a strong response. The CF-filter is rotation invariant; thus, airplanes with different layouts can be detected by CF-filter.

2.2. Aircraft Candidate Locations Extraction. As mentioned before, only the center of aircraft has a strong response in the CF-filtered image. Therefore, aircraft candidates can be extracted by removing low values of CF-filtered image. Here, Otsu's method [21] is used to calculate a global adaptive threshold T from CF-filtered image to remove low values, most of which are produced by clutters within background.

According to the specific size of aircraft candidate (bright spot), a simple shape analysis process is applied to eliminate obvious false candidates. The pixels remained are considered as true candidates; then these regions are forwarded to the second stage.

3. Identifying Aircrafts by Multilevel Features

The appearance of an aircraft can be better captured when color and shape information is combined together. In the second stage, two classes of feature are involved in the proposed solution: Robust Hue Descriptor and Histogram of Oriented Gradients. We build a multilevel feature model, which is akin to the image pyramid representation proposed by Lazebnik et al. [22], to represent both local appearance and spatial layout of an aircraft.

3.1. Robust Hue Descriptor. Color is an indispensable aspect in describing the world around us. However, in a cluttered background, the color representation of an object is greatly interfered by photometric variance; thus, most of existing methods avoid exploiting color information. To obtain a good color descriptor, Van De Weijer et al. [19] proposed robust hue (HUE) descriptor. Experiments show that this descriptor is reliable under photometric and geometrical changes in image retrieval and classification. The HUE descriptor is represented by a histogram over hue computed from the corresponding RGB values of each pixel according to

$$\text{hue} = \arctan\left(\frac{\sqrt{3}\,(R-G)}{R+G-2B}\right). \quad (2)$$

The HUE descriptor is invariant with respect to lighting geometry. In our implementation, the HUE descriptor has 6 dimensions.

3.2. Histogram of Oriented Gradients. Histograms of Oriented Gradients (HOG) [20] are used to capture the spatial distribution of gradients. Since the HOG operator has been proven to be robust and reliable for representing the shape of an object, it is therefore involved into our feature model to describe the shape information of aircrafts. In Dalal's work, the detection window of the HOG operator is split into small cells, in which a 9-bin histogram of gradient orientations is calculated. Every 2×2 cells are grouped together to form a block; thus, each block has a 36-bin histogram feature. Four blocks of features are then concatenated to form the HOG descriptor. Maji et al. [23] proposed a Spatial Histogram of Oriented Gradients (SPHOG) descriptor. Compared to Dalal's HOG descriptor, Maji's descriptor computes the oriented edge energy response by the magnitude of odd elongated oriented filters. Experiments show that, while combining with intersection kernel SVM (IKSVM),

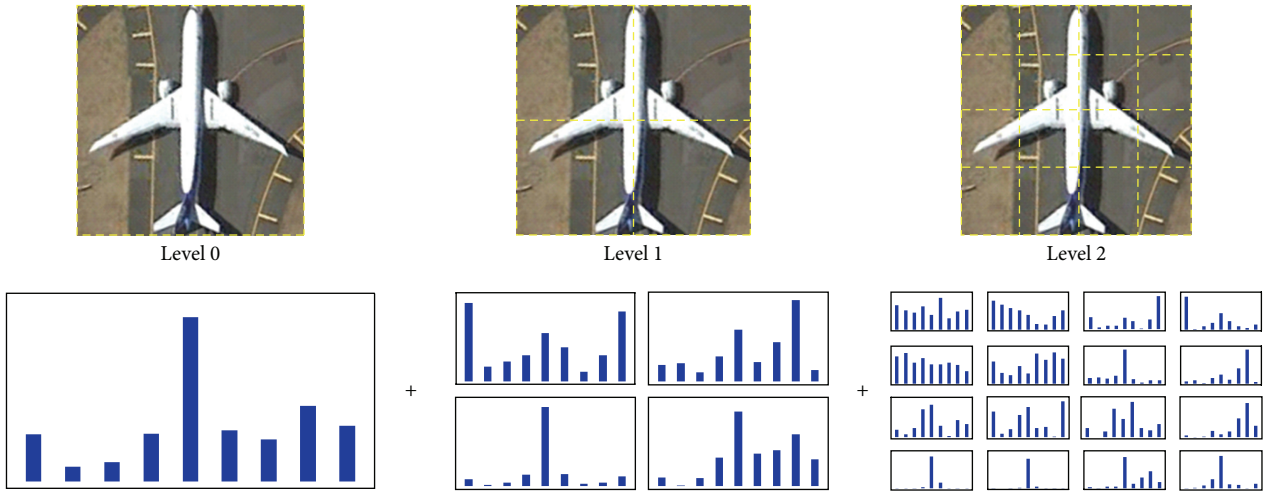


Figure 4: Multilevel feature representation.

the SPHOG descriptor outperforms the original single scale HOG. Hence, we follow Maji's method and use the odd elongated oriented filters to compute oriented edge energy and SPHOG descriptor.

3.3. Multilevel Feature Model. We build a multilevel feature model to represent the color and shape information of aircrafts. The detection window W_i is decomposed into a sequence of increasingly finer spatial cells by repeatedly doubling the number of decompositions. For cell C_i, the feature representation is obtained by concatenation:

$$C_i = \left[\mathrm{HUE}_i, \mathrm{HOG}_i\right]. \tag{3}$$

The feature vector for each cell is a 15-dimensional vector (HUE descriptor has 6 dimensions, and HOG descriptor has 9 dimensions). The number of pixels assigned to a specified color bin or gradient bin is the sum over those contained in the four cells it is divided into at a finer level. Hence, the multilevel feature model is a pyramid representation. The final features are the concatenations of all the 15-dimensional feature vectors. As shown in Figure 4, the cell at level l has 2^l cells along each dimension. Therefore, level 0 is represented by a 15-dimensional vector, level 1 by a 60-dimensional vector, and so forth. Total features of the detection window are a vector with dimensionality $15 \times \sum_{l \in L} 4^l$. Through a multilevel representation, the feature model captures both the local image appearance and its spatial layout. In our implementation, we use a feature model with three levels, yielding a total of 21 cells.

In this paper, SVM with histogram intersection kernel (IKSVM) [23] is applied as a classifier. For feature vectors $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}_+^n$, the intersection kernel $k(\mathbf{x}, \mathbf{y})$ can be expressed as

$$k\left(\mathbf{x}, \mathbf{y}\right) = \sum_{i=1}^{n} \min\left(x\left(i\right), y\left(i\right)\right), \tag{4}$$

and the classification is based on evaluating

$$h\left(\mathbf{x}\right) = \sum_{l=1}^{m} \alpha_l y_l k\left(\mathbf{x}, \mathbf{x}_l\right) + b. \tag{5}$$

Maji et al. [23] noticed that the summation in (5) can be reformed as

$$\begin{aligned} h\left(\mathbf{x}\right) &= \sum_{i=1}^{n} \alpha_l y_l \left(\sum_{l=1}^{m} \min\left(x\left(i\right), x_l\left(i\right)\right)\right) + b \\ &= \sum_{i=1}^{n} \left(\sum_{l=1}^{m} \alpha_l y_l \min\left(x\left(i\right), x_l\left(i\right)\right)\right) + b \\ &= \sum_{i=1}^{n} h_i\left(x\left(i\right)\right) + b. \end{aligned} \tag{6}$$

Consider the functions $h_i(s)$ at fixed point i; $\overline{x}_l(i)$ represents the sorted values of $x_l(i)$ in increasing order with corresponding values of α and labels given by $\overline{\alpha}_l$ and $\overline{y}_l$. Let r be the largest integer, such that $\overline{x}_r(i) \le s$; then we can get

$$h_i\left(s\right) = \sum_{l=1}^{m} \alpha_l y_l \min\left(s, x_{l,i}\right) \tag{7}$$

$$= \sum_{1 \le l \le r} \overline{\alpha}_l \overline{y}_l \overline{x}_l\left(i\right) + s \sum_{r < l \le m} \overline{\alpha}_l \overline{y}_l \tag{8}$$

$$= A_i\left(r\right) + sB_i\left(r\right), \tag{9}$$

where $A_i(r) = \sum_{1 \le l \le r} \overline{\alpha}_l \overline{y}_l \overline{x}_l$, $B_i(r) = \sum_{r < l \le m} \overline{\alpha}_l \overline{y}_l$. It is obvious that (8) is piecewise linear and the function A_i and B_i are independent of the input data. Therefore, s can be precomputed by finding the position of $s = x(i)$ in the sorted list. Maji et al. [23] noticed that the support distributions in each dimension tend to be smooth and concentrated. Therefore, $h(x)$ can be approximated by simpler functions, and the prediction can be accelerated. In this paper, $h_i(s)$ is computed by a look-up table with a piecewise constant approximation.

4. Experimental Results and Evaluation

4.1. Data Collection and Training. From the publicly available Google Earth service and WorldView-2 satellite images, we

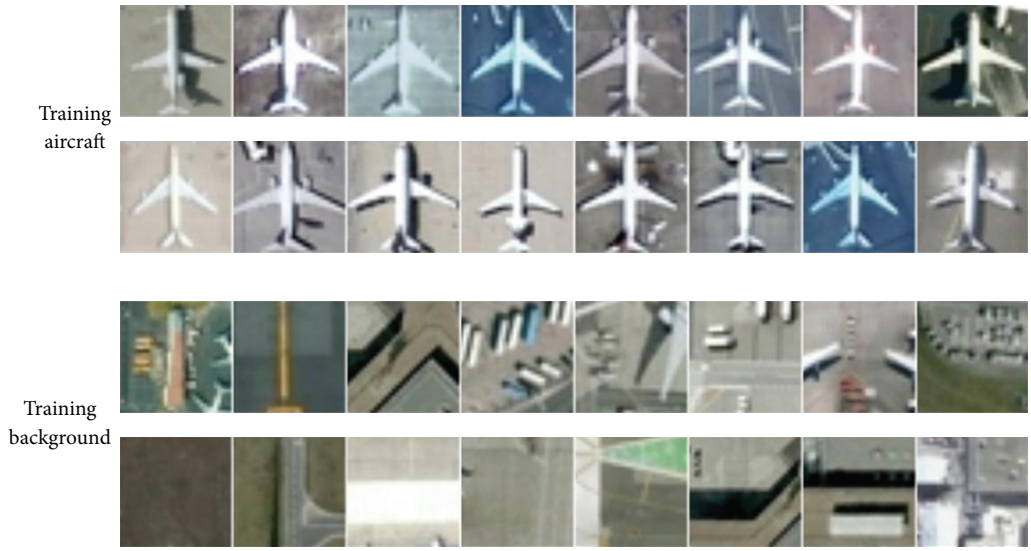


Figure 5: Examples of positive and negative image patches of training dataset.

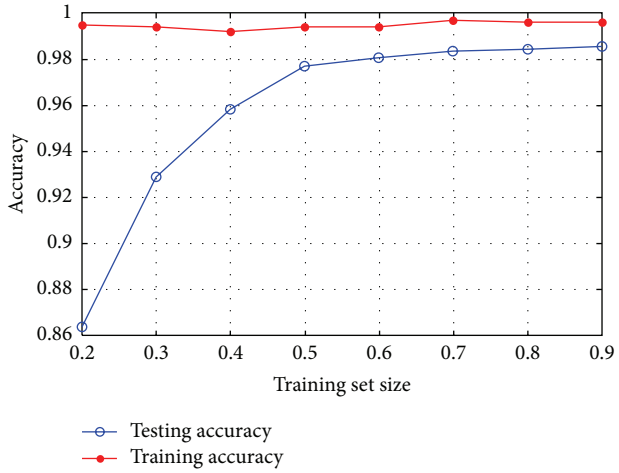


Figure 6: Cross-validation results.

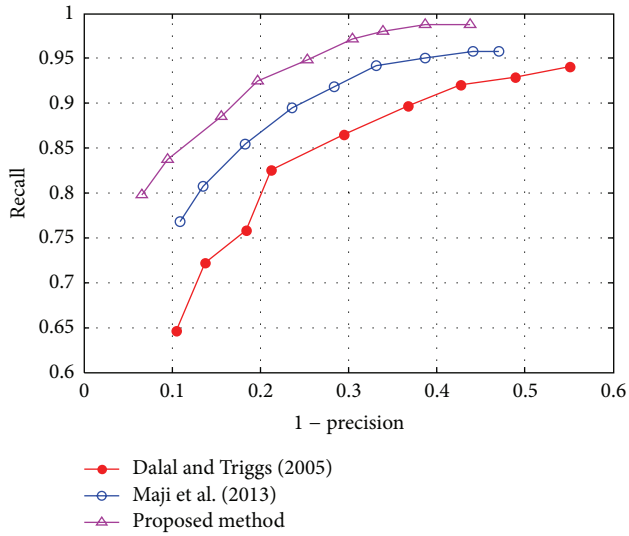


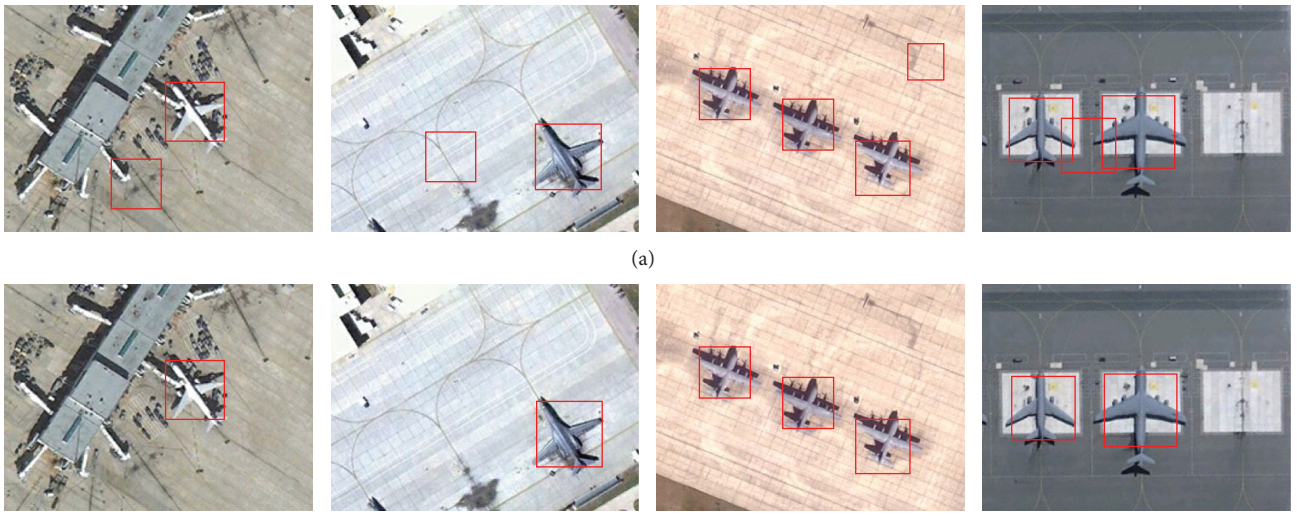
Figure 7: Performance of the three detectors.

collected 914 positive image patches (aircrafts) and 1953 negative image patches. Negative image patches are randomly chosen from the airport backgrounds. Each image patch has a size of 32×32 pixels. Figure 5 shows parts of training image patches of the dataset. We build our testing set by collecting 214 images of other airports containing 597 aircrafts in China, Germany, and France. Each image contains multiple instances of aircrafts with different orientations and sizes.

The IKSVM [23] is applied as a classifier in our method. To find an optimal training set size, we carried out cross-validation. A portion of the training set is selected to train the IKSVM classifier, and then the performance of the classifier is tested by the remaining images in the training set. This process is repeated, and average accuracy is calculated. The average training accuracy and testing accuracy versus training set size are illustrated in Figure 6. The training accuracy is high for all training set sizes, indicating that the multilevel feature model is effective to distinguish the images in the training set. The testing accuracy tends to be stable at a high rate when the number of training images is more than half of training set, which indicates that our method is capable of representing aircraft from a small training set.

4.2. Quantitative Evaluation. We manually label the aircrafts appearing in all testing images as a ground truth. When a detection to be marked is a true positive, more than 50% of it must be detected. M denotes the total number of aircrafts in testing images. The recall-precision curve (RPC) is chosen to exhibit the tradeoff between recall and precision. Recall and 1 − precision are defined as

$$\begin{aligned} \text{Recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})}, \\ 1 - \text{Precision} &= \frac{\text{FP}}{(\text{TP} + \text{FP})}, \end{aligned} \tag{10}$$



(a)

(b)

Figure 8: The effectiveness of CF-filter. (a) Detection results without CF-filter; (b) detection results with CF-filter.

where TP (true positive) denotes the number of true detected aircrafts, FN (false negative) is the number of missed detections, and FP (false positive) is the number of false detected aircrafts. Therefore, recall denotes the number of true detected aircrafts divided by the total number of aircrafts in testing images and 1 − precision denotes the number of false detected aircrafts divided by the elements detected. The ratio recall/(1 − precision) is used to represent the performance of the algorithm.

Since the size of aircraft in the testing set is unknown, the testing images are scanned at multiple scales. At each scale, the radius r and the number of pixels N in CF-filter are set to 8 and 60, respectively. After candidate extraction, we use the sliding window approach to detect aircrafts. The size of the sliding window is 32×32 pixels, and the step is set to be 8 pixels.

We compared our method with two classical methods. The first method is proposed by Dalal and Triggs [20], in which a linear Support Vector Machine is exploited to classify aircrafts and background by HOG features. The second method is Maji's approach [23], where the IKSVM is applied as a classifier and spatial histograms of oriented gradients (SPHOG) are calculated for the classifier. From Figure 7, we can observe that the proposed method outperforms the other two detectors. At the same recall level, the precision of the proposed method is the best, which means that the false alarms rate of the proposed method is the lowest. While at the same precision level, the proposed method detects more true positives.

In order to test the impact of CF-filter, our original method is compared to the modified method without CF-filter, see Figure 8. The first row is the experimental results without CF-filter and the second row is the results with CF-filter. It is obvious that false positives of our original method are much less than the modified method. This is because CF-filter can identify the shape of aircrafts in cluttered background.

5. Conclusion and Future Work

In this paper, we proposed a coarse-to-fine method for aircraft detection in VHR remote sensing images. CF-filter is applied to extract the candidate locations of aircrafts in the coarse stage. In the fine stage, a multilevel feature model is designed to capture both the local appearance and spatial layout of aircrafts. Meanwhile, Robust Hue Descriptor and Histogram of Oriented Gradients are utilized to describe the color and shape information of aircrafts. The experiment results show the good performance of the proposed method.

Many satellite images provide more than red, green, and blue channels; other channels such as infrared, coastal blue, and red edge are often acquired. In the future work, we will study how to utilize more spectral information to represent aircrafts.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this Paper.

Acknowledgments

The authors would like to thank Dr. S. Maji for kindly providing IKSVM [23]. The authors would also like to thank Dr. Y. Yang from the Ocean University of China for her helpful discussion and the anonymous reviewers for their helpful comments and suggestions. This work was partially supported by the 973 Program (2010CB327900) and the National Science Fund for Distinguished Young Scholars (no. 61125206).

References

[1] A. Kembhavi, D. Harwood, and L. S. Davis, "Vehicle detection using partial least squares," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 33, no. 6, pp. 1250–1265, 2011.

[2] G. Liu, X. Sun, K. Fu, H. Wang, and Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior, *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 573–577, 2013.

[3] H. Cai and Y. Su, "Airplane detection in remote-sensing image with a circle-frequency filter," in *International Conference on Space Information Technology*, vol. 5985 of *Proceedings of SPIE*, Wuhan, China, November 2005.

[4] B. Salehi, Y. Zhang, and M. Zhong, "Automatic moving vehicles information extraction from single-pass worldView-2 imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 135–145, 2012.

[5] J. Leitloff, S. Hinz, and U. Stilla, "Vehicle detection in very high resolution satellite images of city areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2795–2806, 2010.

[6] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3446–3456, 2010.

[7] L. Chen, S. Zhao, W. Han, and Y. Li, "Building detection in an urban area using lidar data and QuickBird imagery," *International Journal of Remote Sensing*, vol. 33, no. 16, pp. 5135–5148, 2012.

[8] A. O. Ok, C. Senaras, and B. Yuksel, "Automated Detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, 2013.

[9] Z. Lei, T. Fang, H. Huo, and D. Li, "Rotation-invariant object detection of remotely sensed images based on texton forest and hough voting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1206–1217, 2012.

[10] Y. Li, X. Sun, H. Wang, H. Sun, and X. Li, "Automatic target detection in high-resolution remote sensing images using a contour-based spatial model," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 5, pp. 886–890, 2012.

[11] J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 3, pp. 236–248, 2007.

[12] H. G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2097–2111, 2008.

[13] G. E. Atteia and M. J. Collins, "On the use of compact polarimetry SAR for ship detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 80, no. 6, pp. 1–9, 2013.

[14] M. T. Alonso, C. Lopez-Martinez, J. J. Mallorqui, and P. Salembier, "Edge enhancement algorithm based on the wavelet transform for automatic edge detection in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 222–235, 2011.

[15] N. Renard and S. Bourennane, "Improvement of target detection methods by multiway filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2407–2417, 2008.

[16] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of target-detection algorithms based on adaptive three-dimensional filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1383–1395, 2011.

[17] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 1, pp. 109–113, 2012.

[18] W. Zhang, X. Sun, K. Fu, C. Wang, and H. Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model," *IEEE Geoscience and Remote Sensing Letters*, 2013.

[19] J. Van De Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 150–156, 2006.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, June 2005.

[21] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 2169–2178, June 2006.

[23] S. Maji, A. C. Berg, and J. Malik, "Efficient classification for Additive Kernel SVMs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66–77, 2013.