IMPROVING INDEPENDENT COMPONENT ANALYSIS USING SUPPORT VECTOR MACHINES FOR MULTIMODAL IMAGE FUSION

[1]D. Egfin Nirmala, [1]A. Bibin Sam Paul and [2]V. Vaidehi

[1]Department of Electronics and Communication Engineering, B. S. Abdur Rahman University, Chennai, India
[2]Department of Information technology, Madras Institute of Technology, Anna University, Chennai, India

ABSTRACT

The objective of this study is to combine multiple images of a scene acquired by different sensors to create a new image with all important information from the input images. Recent studies show that bases trained using Independent Component Analysis (ICA) is effective in multisensor fusion and has improved performance over traditional wavelet approaches. In the ICA based fusion, the coefficients of the input images are combined simply by selecting the coefficients with maximum magnitude. But this method resulted in fused images with poor contrast, due to the distortion introduced in constant background areas. The performance of ICA based fusion can be greatly improved by using a region based approach with intelligent decision making in order to choose the significant regions in the source images. Hence, a new region based image fusion algorithm for combining visible and Infrared (IR) images using Independent Component Analysis and Support Vector Machines (SVM) is proposed. Region based joint segmentation of the source images is carried out in the spatial domain and important features of each region are computed in spatial and transform domain. A Support Vector Machine is trained to select the regions from the source images with significant features and the corresponding ICA coefficients are combined to form the fused ICA representation. The proposed algorithm is applied to different sets of multimodal images to validate the robustness of the algorithm and compared with some standard image fusion methods. The fusion results demonstrate that the proposed scheme performs better than the state-of-the-art image fusion methods and show a significant improvement in Entropy, Petrovic and Piella evaluation metrics.

Keywords: Image Fusion, Independent Component Analysis, Joint Segmentation, Support Vector Machines, Training Images

1. INTRODUCTION

In recent years, with the availability of low cost imaging sensors, multiple sensors are used in applications such as robotics, surveillance, defence, medical imaging and remote sensing in order to improve the system performance. In such multisensor systems, redundant and complementary information from different sensors are combined to form a fused image that contains all important information from the input images (Lewis *et al.*, 2007). Image fusion enhances the perception of a scene, by integrating multiple sensor data (Mitianoudis and Stathaki, 2007). The fusion process creates more accurate fused images suitable for carrying out further processing (Nirmala *et al.*, 2011). The multiple sensors deployed to capture images of a scene may be of the same type or different. In multimodal fusion the images from different sensors such as visible and Infrared (IR) are combined (Saha *et al.*, 2013). Recently, merging visual and infrared images is commonly

Corresponding Author: D. Egfin Nirmala, Department of Electronics and Communication Engineering,
B. S. Abdur Rahman University, Chennai, India

used in many surveillance and tracking applications (Drajic and Cvejic, 2007). The fused images obtained by this multimodal fusion, have enhanced information than the individual sensors, which improves detection.

Image fusion algorithms are broadly classified into two categories namely, pixel based and feature based. In pixel level fusion, the raw pixels obtained directly from imaging sensors are combined based on certain fusion rules. Feature based fusion, extracts some features from the source images and performs fusion based on these features. This is usually achieved by using a region-based scheme. In the recent past, many researchers have demonstrated the effectiveness of Multiresolution Transforms (MR) in image fusion applications (Blum and Liu, 2005). A popular multiresolution transform is the Discrete Wavelet Transform (DWT) (Mallat, 1998). Li et al. (1995) first employed DWT for image fusion. In recent years, many pixel-level image fusion algorithms using MR transforms are proposed and successfully applied for fusion (Ellmauthaler et al., 2013).

Recently region-based methods are widely used in majority of the applications, as the objects and the region details are more important than the individual pixels. Hence, the object and region details are incorporated for fusion (Piella, 2002). In these methods the input images are divided into several regions and some important features of each region are determined. These features are used to decide the input image from which a particular region is to be selected in the fused representation. Region based methods are more flexible in adapting to intelligent fusion rules than pixel based approaches (Wan et al., 2009). Also these methods circumvent the common problems of noise sensitivity and blurring associated with pixel based schemes. A region based fusion technique using the Dual Tree Complex Wavelet Transform (DT-CWT) is proposed by Lewis et al. (2007). In this, only one region feature Shannon entropy computed from DT-CWT coefficients is used as the activity measure for selecting regions from the input images. The fusion performance of this region based method is found to be comparable to that of pixel-level fusion method using DT-CWT. Piella (2002) presented a standard region based multiresolution scheme that outperforms other multiresolution fusion techniques. But the implementation was in the preliminary stage and the results need to be optimized.

In image processing applications, it is required is to find a suitable representation for image data and are generally based on linear transformations such as Fourier, Cosine, Haar and wavelet transforms. Each of these transforms has some advantages over the other. If a linear transformation is estimated from the data to be processed itself, it could adapt to any type of data (Jutten and Herault, 1991). As Independent Component Analysis (ICA) is such a fundamental linear transformation in which the bases are determined from the data itself, in the proposed work ICA is used for image representation. The ICA bases are very effective for image fusion, which can outperform wavelet based schemes (Mitianoudis and Stathaki, 2007). In the recent past, several methods were implemented in which ICA is successfully applied for image fusion in transform domain (Cvejic et al., 2007). Mitianoudis and Stathaki (2007) used bases trained using Independent Component Analysis (ICA) as a tool for fusion and applied this algorithm for fusing visual and IR images. In the fused images produced by this method, the important objects from the infrared images were blurred. Recently, Cvejic et al. (2007) have proposed a region based method using ICA bases in which entropy of regions is used as a priority measure. The authors used an adaptive approach for the reconstructing the fused image using the Piella quality index (Piella and Heijmans, 2003). Though this approach resulted in improved performance, it adds to a high computational overhead and consumes more time.

Region based image fusion involves feature extraction and selection process. This requires a suitable classifier that uses multiple features of the input images to identify the most important regions (Lu and Weng, 2007). The Support Vector Machine (SVM) is a classifier employed successfully in various applications and is found to outperform conventional classifiers (Li et al., 2004). SVM a supervised classification method can handle large input data and feature sets (Luo et al., 2012) and hence suitable for image processing applications. A pixel based method for fusing multifocus images using Discrete Wavelet Frame Transform (DWFT) and SVM is proposed by Li et al. (2004) in which a trained SVM classifier chooses the pixels from the source images that have the highest activity at each coefficient location. Chen et al. (2008) presented Empirical Mode Decomposition (EMD) with SVM for merging multifocus images. Fusion results of these two methods demonstrate that the use of SVM classifier resulted in an improved performance over DWFT and EMD fusion.

It is evident from the related work, that ICA with SVM is more appropriate for combining multiple images. In this study, the combination of ICA and SVM

is used for multimodal image fusion, to improve the ICA based fusion. This method employs bases trained using Independent Component Analysis and Support Vector Machine classifier to choose regions with significant features from the segmented source images. This algorithm takes into account, multiple region-based measurements in spatial and transforms domains, to decide the significance of the various regions in the input images. The decision making to choose or discard a region, based on the region features is performed by a trained SVM. The region with highest priority is always chosen and the corresponding ICA coefficients form the fused ICA representation. The inverse transformation is applied to this fused ICA representation to get the fused image.

The rest of this study is organized as follows: In Section 1.1 the basics of ICA and SVM are discussed. In Section 2 the proposed fusion method is described in detail. The fusion results are presented in Section 3 and Section 4 provides the conclusion the paper.

1.1. Background

The proposed study employs Independent Component Analysis as the linear transformation tool for image analysis. Segmentation is performed to divide the source images into meaningful regions and a Support Vector Machine classifier is used to identify the most important regions based on the region properties. The algorithm then combines the ICA coefficients of the selected regions to obtain the fused representation. In this section, the theoretical background of the proposed study is discussed.

1.1.1. Image Analysis using ICA

In image analysis, an image is represented as a combination of basis images and is based on standard bases such as wavelet and cosine. But in ICA, the bases are estimated from the input data itself. In order to generate these bases some training images having statistical properties similar to the input images to be combined are required. For an $M_1 \times N_2$ image f(x,y), an image patch is defined by Equation 1:

$$f_w(u,v) = w(u,v)f(u_0 - (N/2) + u, v_0 - (N/2) + v) \quad (1)$$

for all $u,v \in [0,N-1]$, where w(u,v) is a N×N window centered at the pixel (u_0, v_0) Each N×N image patch $f_w(u,v)$ is expressed as a linear combination of P basis image patches $b_j(u,v)$ as given by:

$$f_w(u,v) = \sum_{j=1}^{P} s_j b_j(u,v) \quad (2)$$

where, s_j are scalar values, given by (Cvejic *et al.*, 2007):

$$s_j = \langle f_w(u,v)\, b_j(u,v) \rangle$$

Several N×N training patches are chosen randomly from the training set using a rectangular window and each N×N image patch is converted into a vector $\underline{f}_w$. The resulting vector is represented in terms of the basis vectors $\underline{b}_j$ as follows Equation 3:

$$\underline{f}_w(t) = \sum_{j=1}^{P} s_j(t)\underline{b}_j \quad (3)$$

where 't' is the patch index and P is the number of image patches.

Equation (2) can be written as Equation 4 and 5:

$$\underline{f}_w(t) = B\underline{s}(t) \quad (4)$$

$$\underline{s}(t) = B^{-1}\underline{f}_w(t) = A\underline{f}_w(t) \quad (5)$$

where, s(t) = $[s_1(t),s_2(t),…,s_p(t)]^T$, B the synthesis kernel given by $B = [\underline{b}_1,\underline{b}_2,...,\underline{b}_P]^T$ and $A = B^{-1} = [\underline{a}_1,\underline{a}_2,...,\underline{a}_P]^T$ the analysis kernel. The purpose is to determine a set of *L* basis vectors that are statistically independent. Therefore to estimate the $L<N^2$ uncorrelated basis vectors, Principal Component Analysis (PCA) is used. This helps in dimensionality reduction (Hyvarinen, 1999a), as reducing the dimension lowers the computational costs. The input data correlation matrix given by Equation 6:

$$C_d = \varepsilon\left(\underline{f}_w\underline{f}_w^{\ T}\right) \quad (6)$$

where, ε is the expectation operator. PCA is the eigen decomposition of the matrix C_d. After applying PCA, the L eigen vectors with L largest eigen values are selected. Let V^R be the reduced PCA matrix of dimension $L \times N^2$ and the transformed patches are given by Equation 7:

$$\underline{Z}(t) = V^R\underline{f}_w(t) \quad (7)$$

Following the pre-processing step using PCA, the basis vectors $\underline{s}_j$ that are statistically independent are selected by optimizing cost function negentropy. This optimization results in Fast ICA algorithm proposed by Hyvarinen and Oja (2000) and the update rule is given by Equation 8 and 9:

$$\underline{a}_i^+ = \varepsilon\{\underline{a}_i\varphi(\underline{a}_i^T\underline{z})\} - \varepsilon\{\underline{a}_i\varphi'(\underline{a}_i^T\underline{z})\}\underline{a}_i, \ 1 \le i \le K \tag{8}$$

$$A \leftarrow A(A^TA)^{-\frac{1}{2}} \tag{9}$$

where, $\underline{a}_i$ are the projection vectors, $\varphi(x) = -\dfrac{\delta G(x)}{\delta x}$ and $G(x)$ is any non-quadratic function. In this study, $G(x)$ is chosen as given by Equation 10:

$$G(x) = \frac{1}{\beta}\log(\cosh(\beta x)) \tag{10}$$

where, β is a constant, $1 \le \beta \le 2$ (Hyvarinen, 1999b).

The ICA update rule given in Equation (8) is iterated in a chosen neighborhood until $\underline{a}_i$ converges. The selected random patches $\underline{f}_k(t)$ are now converted to the ICA domain representation $\underline{s}_k(t)$ as given by Equation 11:

$$s_k(t) = T\{f_k(t)\} = A\underline{f}_k(t) \tag{11}$$

Thus the input image patches $\underline{f}_k(t)$ of the source images can be converted to the ICA domain representations $s_k(t)$.

1.1.2. Support Vector Machines

A SVM is a classification method that discriminates between two classes, by fitting an Optimal Separating Hyperplane (OSH) (Gunn, 1998). It actually maximises the margin between two classes of training data. With $\{x_j\}$ as a set of input training samples having n training points in a 'N' dimensional space, the output y_j with $y_j \in \{-1, 1\}$, the training data is given by Equation 12:

$$\{x_j, y_j\}, \ j = 1, 2, \ldots n, \quad x_j \in R^N, \ y_j \in \{-1, 1\} \tag{12}$$

The SVM performs a mapping of the training patterns to a feature space from the input space and constructs a hyperplane that separates the various classes with maximum margin. This is a quadratic programming problem (Gunn, 1998) and involves maximizing Equation 13:

$$L_s = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \kappa(x_i, x_j) \tag{13}$$

where, $\kappa(x_i, x_j)$ is the SVM kernel function that performs the mapping, α_i 's are support vectors, $0 \le \alpha_i \le C$ and $\sum_{i=1}^{n}\alpha_i y_i = 0$ for $i = 1, 2, \ldots, n$, C is a regularization parameter to be defined by the user. Higher the value assigned to C, higher the penalty assigned to the margin errors.

One of the popular kernels used in SVM is the Radial Basis Function (RBF) kernel, which has a parameter known as Gaussian width, σ. In this study, Gaussian Radial Basis Function (GRBF) is used and is given by Equation 14:

$$\kappa(x_i, x_j) = e^{-\frac{\|x_i, x_j\|^2}{2\sigma^2}} \tag{14}$$

where, x_i and x_j denote the training patterns given. Interested readers may refer to Gunn (1998) for more details.

To perform classification, the input data is divided into two sets: training and test. From the input data, several features or attributes are computed which form the input variables. The training set at each instance consists of a target value called class labels and the observed input variables. In the training phase, the SVM is trained using the training data generated. In the testing phase, the trained SVM determines the target values of the test data based on its attributes.

1.1.3. Region Based Fusion

Region based fusion methods employ segmentation in order to divide an image into several disjoint regions (O'Callaghan and Bull, 2005). The input images to be combined are segmented using a suitable method into regions of different sizes. For each region one or more important properties/features can be computed. These features are used in deciding whether a region in the fused representation should be from the infrared image or visible image.

The segmentation process employed should produce the optimum number of regions with all salient objects included, as more number of regions increases the time taken to generate the fused image. For fusion, it is better to use joint segmentation (Lewis et al., 2007) as the joint segmentation map will produce the same number of regions in each source image.

2. PROPOSED FUSION METHOD

Consider that there are M registered source images $f_k(x,y)$ of size M×N to be combined. Divide the given source images into all possible patches of size N×N and convert the patches into a vector $f_k(t)$ using lexicographic ordering. Transform these representations to the ICA domain $S_k(t)$. This method employs segmentation in spatial domain to divide the source images into non overlapping regions. For each region, various features are extracted and a SVM classifier is trained to select the regions from the visual or IR input images based on these features. **Figure 1** shows the block schematic of the proposed image fusion method.

2.1. Training ICA Bases

From the set of training images which contains visual and IR images, 10,000 rectangular patches are randomly taken. These patches are transformed to the ICA domain, using the estimated analysis kernel *A* as given in (11). This training process is required to be done only once. After training the desired bases, the estimated transform is used for fusion of images of similar type.

2.2. Segmentation

In this study, joint segmentation of source images is performed using combined morphological-spectral unsupervised image segmentation algorithm proposed by O'Callaghan and Bull (2005). The first step of this spatial domain algorithm uses textured and non textured regions of the input image to perform the initial segmentation. The texture features are generated from the details sub-bands of the Dual-Tree Complex Wavelet Transform (DT-CWT). Then a Gaussian gradient function is applied to all scales and combined with the intensity gradient information to form the final gradient with all perceptual edges in the images. This is followed by Watershed transformation of the perceptual gradient. The second stage uses a clustering algorithm to group the primitive regions provided by the watershed algorithm (O'Callaghan and Bull, 2005).

Once the joint segmented map is obtained, it is applied to both visual and IR source images separately to divide them into different segments, so that the individual segments can be compared independently. In order to separate the various segments, all the regions formed by the segmentation process in each source images are labeled.
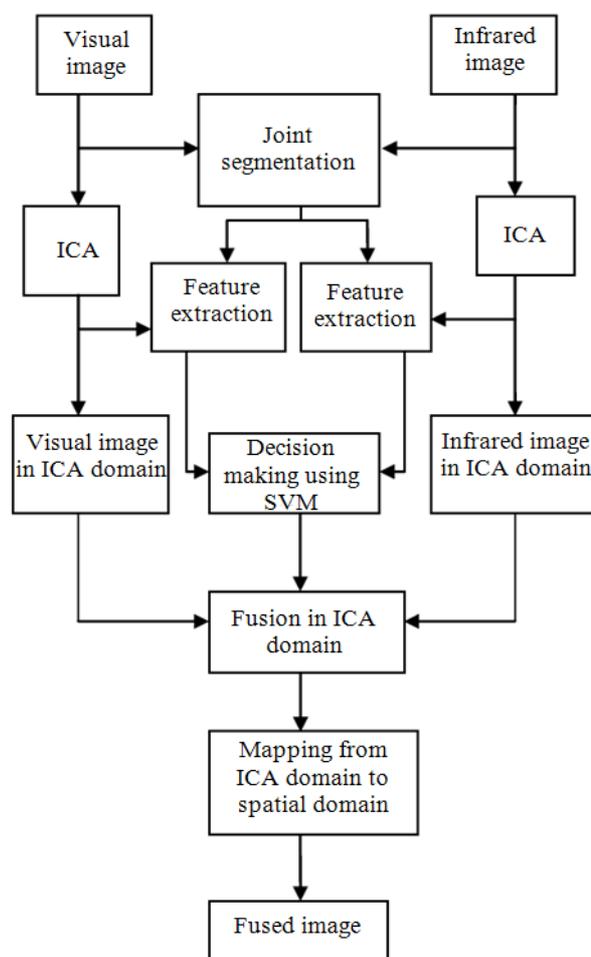

Fig. 1. Image fusion using the proposed method

2.3. Feature Extraction and SVM Training

The segmentation process is followed by feature extraction, in which various region features are computed in order to determine the importance of each region in the source images. In this study, the features extracted in DT-CWT domain, ICA transformation domain and spatial domain are used in order to improve the performance of the fusion process. Using more number of features is of great significance and seven features proved to be good measures of importance are used in this method.The average energy, entropy and standard deviation of a region are generally good measures in deciding its importance. In the first stage of the segmentation process, DT-CWT is used to extract the various texture features. The DT-CWT coefficients $d_{k(\theta,1)}(m,n)$ generated in segmentation are

used for computing the features energy, entropy and standard deviation for all regions in the input images and are given by:

- Energy Equation 15:

$$E(S_{ik}) = \sum_{\forall\theta,\forall l,(m,n)\in S_{ik}} d^2_{k(\theta,l)}(m,n) \qquad (15)$$

- Entropy Equation 16:

$$H(S_{ik}) = - \sum_{\forall\theta,\forall l,(m,n)\in S_{ik}} d_{k(\theta,l)}(m,n)\log_2 d_{k(\theta,l)}(m,n) \qquad (16)$$

- Standard Deviation Equation 17:

$$SD(S_{ik}) = \frac{1}{|S_{ik}|}\sqrt{\sum_{\forall\theta,\forall l,(m,n)\in S_{ik}}\left[d_{k(\theta,l)}(m,n)-\overline{d}_{k(\theta,l)}(m,n)\right]^2} \qquad (17)$$

where, S_{ik} is the ith region in image k, $d^2_{k(\theta,l)}(m,n)$ are the detail coefficients, $\overline{d}_{k(\theta,l)}(m,n)$ are the mean of the detail coefficients of a region of the DT-CWT, l is the decomposition level, θ is the orientation of the sub-band.

There are different activity measures that can be computed from the ICA coefficients to define the significance of a region. Some such activity measures for fusion were proposed by Piella (2002). The ideas proposed by the author are modified in this study to extract region features from ICA coefficients. The following two features for each region are computed from the ICA coefficients.

- The mean absolute value

The mean absolute value (L_1-norm) of the ICA coefficients of each region arranged as a vector is given by Equation 18:

$$M(S_{ik}) = \|C_{ik}(j)\|_1, \quad i=1,2,....,N \qquad (18)$$

where, N is the total number of segmented regions.

- Variance Equation 19:

$$Var(S_{ik}) = \sum_{j\in S_{ik}}^{N}\left[C_{ik}(j)-\overline{C}_{ik}(j)\right]^2 \qquad (19)$$

where, $C_{ik}(j)$ are the ICA coefficients of the ith region in image k arranged as a vector. Large values of M(S) and $Var(S_{ik})$ correspond to increased activity in that region.

Region contrast and Sharpness are two image features that are of greater visual importance. Region sharpness is a measure of the edge information a region can convey. Regions with a high contrast with respect to their surroundings correspond to increased activity (Lu and Weng, 2007). Hence, the following two features are calculated for all the regions of the input images in the spatial domain:

- Contrast Equation 20:

$$CO(S_{ik}) = \sum_{p=1}^{M}\sum_{q=1}^{N}(p-q)^2 S_{ik}(p,q) \qquad (20)$$

where, $S_{ik}(p,q)$ are the (p,q) th entry in the ith region of image k in spatial domain.

- Sharpness Equation 21:

$$SH(S_{ik}) = \frac{\sqrt{(G_x^2(S_{ik}(p,q))+G_y^2(S_{ik}(p,q)))}}{MN} \qquad (21)$$

where, G_x and G_y are the gradients along 'x' and 'y' directions for the ith region.

Thus seven important features computed in spatial and transform domains are used in this algorithm to quantify the importance of regions for making decisions, but additional features can be easily incorporated. To train the SVM a set of training images both visual and IR is required, which have similar content and statistical properties as that of the inputs to be fused. The training images are segmented into regions using the segmentation algorithm described in section 3.2. Some regions from the segmented images of different sizes are randomly selected for training and the seven features given above are computed. The inputs to the SVM are the seven features obtained in this section. The SVM is trained to determine whether a region from visual or IR image should be included in the fused image. The SVM output is positive (+1) if a region in the visual image has any four of the feature values greater than the corresponding IR image feature values and negative (-1) otherwise (supervised learning). The SVM is thus trained to select important regions from the visual or IR images using the extracted features.

2.4. Fusion Process

The SVM training is followed by the testing phase in which fusion is performed based on the SVM output. If

positive, the ICA coefficients for the corresponding region will be taken the visual image and vice versa. The SVM always picks the regions with most important features and the coefficients corresponding to the selected regions form the fused ICA domain representation. The synthesis kernel B is used to generate the fused image from the fused ICA coefficients.

3. RESULTS AND DISCUSSION

In this study, region based fusion of different sets of multimodal images were carried out with the ICA-SVM framework and the results are presented. Before the fusion process, the ICA bases were trained using 2000 image patches of size 8×8 taken from a set of eight visual and eight IR images of similar content from the image fusion site (www.imagefusion.org). Then PCA is applied to the selected patches and 32 important bases were selected based on the largest eigen values.

The training images are segmented using the segmentation algorithm discussed in section 3.2 and the joint segmentation maps were formed. For SVM training, 100 regions of various sizes were randomly selected from the visual and IR images training data set, with 50 from each. The seven features discussed in the previous section were computed for the selected regions in the visible and IR images and are represented as a 7-dimensional feature vector. This feature vector is fed to a SVM for classification. The same image set used for training ICA bases were used to train the SVM too. The training patterns for the SVM were only a small portion of the full image data.

Experiments in this study were carried out in an i3, 370M CPU having 4 GB RAM using MATLAB. The Bioinformatics toolbox with the GRBF kernel in MATLAB is used for SVM. The regularization parameter C is set to 2000, as the performance of the SVM classifier is found to be is stable for large values of C (Li *et al*., 2004). The value of σ for the GRBF kernel is chosen as the scatter radius of the training samples. The ICA bases and SVM were trained offline only once using randomly selected samples and thereafter used constantly for fusion.

The multimodal images used in this study are: surveillance UN camp images from the image fusion site (www.imagefusion.org), OTCBVS images available in the site (www.cse.ohiostate.edu/otcbvs-bench) and another pair of visible and IR surveillance images.

The results obtained for the source image sets, by the proposed method were compared with fusion algorithms

such as standard ICA (Mitianoudis and Stathaki, 2007), Laplacian Pyramid (LAP) (Burt and Andelson, 1983), Discrete wavelet transform (Li *et al*., 1995), DT-CWT (Lewis *et al*., 2007) and Region based ICA (Cvejic *et al*., 2007) in order demonstrate the improvement in performance. In standard ICA, the source images were combined by the fusion scheme proposed by Mitianoudis and Stathaki (2007). The LAP and DWT based schemes were demonstrated to have good performance and hence included in the comparison. For LAP, DWT, DT-CWT 3-levels of decomposition is used and the transform coefficients with the maximum absolute value were selected to form the fused representation. In Region based ICA, fusion is performed following the algorithm given by Cvejic *et al*. (2007) using entropy as the priority measure for selecting regions.

The effectiveness of the proposed study in multimodal fusion is tested using the standard objective fusion metrics Entropy, Standard Deviation (SD), Mutual Information (MI), Petrovic metric (Q^f_{AB}) (Xydeas and Petrovic, 2000) and Piella metric (Q_{piella}) (Piella and Heijmans, 2003). Mutual Information as a fusion metric is a measure of degree of dependence between the fused image and the input images. Petrovic metric (Q^f_{AB}) evaluates the amount of edge details contained in the fused image which have been transferred from the inputs. The quality measure Q_{piella}, estimates the relative amount of details transferred to the fused image, from the source images. Larger the values of these metrics better the performance.

Once the ICA bases are generated and the SVM is trained, fusion of input images can performed. The ICA coefficients for the source images to be fused were obtained using (11). Joint segmentation of the source images were performed in the spatial domain. **Figure 2** shows the segmentation outputs obtained with the UN camp source images. For all regions in both visible and IR images, the seven region features were computed. Then the trained SVM is used to make intelligent decisions and fusion is performed based on the SVM output.

Figure 3a and b show the UN camp visual and IR source images, **Fig. 3c-h** shows the fused images using standard ICA, LAP, DWT, DT-CWT, Region based ICA and the proposed method. The person appears brighter, in all methods tested except in standard ICA (Mitianoudis and Stathaki, 2007). It is observed from the fused images that only the proposed method is able to transfer the fine details present in the visible image to the fused image. The trees and the full fence detail in the visible image are clearly seen in the proposed scheme than any other method tested.
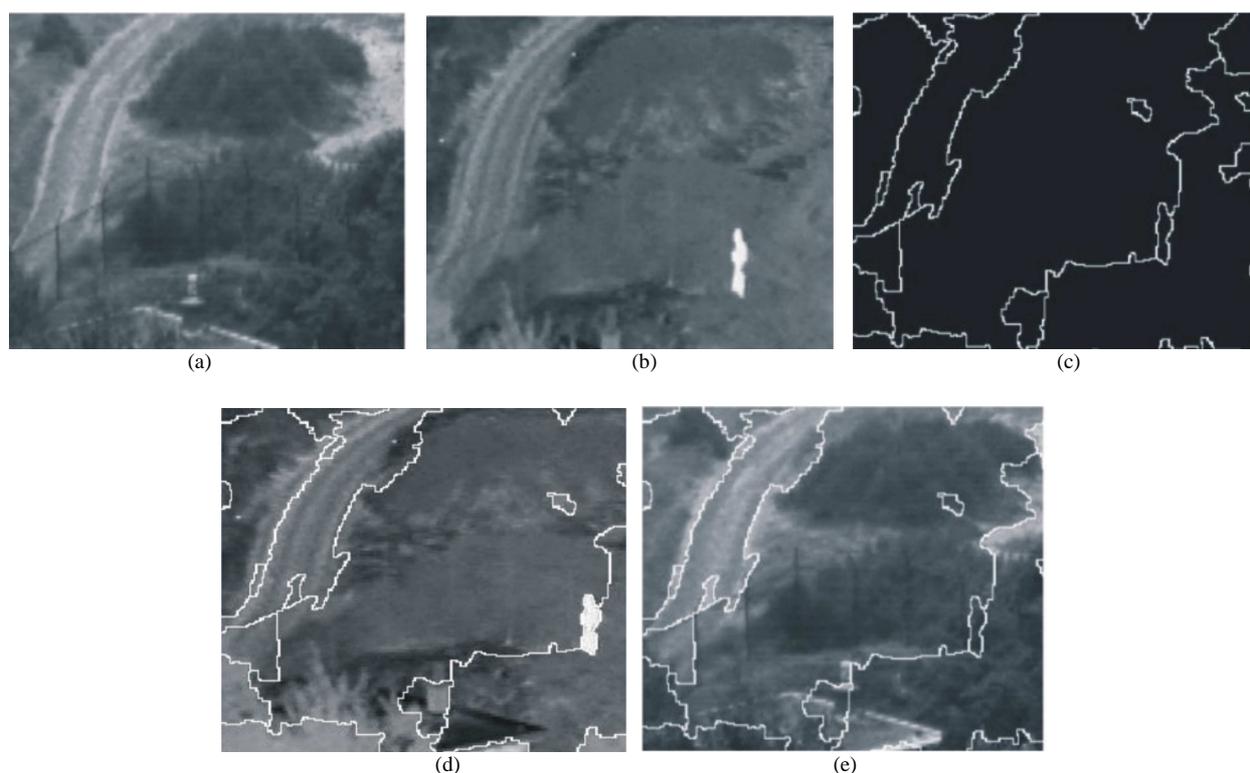
Fig. 2. Joint segmentation results of UN camp image. (a)Visual input image (b) IR input image. (c) Joint segmented map (d) Segmentation Map applied to Visual Image (e) Segmentation Map applied to IR Image

Also, visual comparison show that the fused image in **Fig. 3h** is superior to Standard ICA (Mitianoudis and Stathaki, 2007), region based ICA (Cvejic *et al.*, 2007) and other multiresolution methods compared. The comparison results based on the objective fusion metrics for the UN camp image are given in **Table 1**. The proposed method scores the highest MI value of 7.1, while region based ICA scores 4.162. All the multiresolution methods included for comparison scores low values for mutual information. This indicates that more useful details from the input images are available in the fused image of the proposed method. Also the proposed method scores high in terms of the two important fusion metrics Q^f_{AB} and Q_{piella} than all other methods. Even a small difference of 0.01 in the two metrics Q^f_{AB} and Q_{piella}, is considered significant for quality rating (Wan *et al.*, 2009).

Figure **4a and b** show a visible image and an IR image of the same scene. **Figure 4c-h** show the fused images of the methods tested. Visual inspection of the fused images shows that DT-CWT and the two region based methods are only able to transfer the target (the gun)

information from the input IR image to the fused images. But DT-CWT based fusion (Lewis *et al.*, 2007) resulted in a fused image with the background blurred. The fused image of the ICA-SVM method in **Fig. 4h** has taken the entire background information from the visual image and object (the gun) information alone from the IR image. The fusion performance results for the source images shown in **Fig. 4a and b** are presented in **Table 2**. The proposed method scores slightly higher than region based ICA proposed by Cvejic *et al.* (2007) in terms of all the fusion metrics, but significantly outperforms the individual multiresolution methods in terms of Entropy, MI, Q^f_{AB} and Q_{piella}. **Figure 5a and b** show OTCBVS multimodal source images and the fused images are presented in **Fig. 5c-h**. It is clearly seen that the fused images obtained by the proposed method contain more details than any other methods tested. The fused image obtained by the proposed ICA-SVM method in **Fig. 5h** has taken most of the significant information from the source images and scores higher than the tested methods in terms of the performance metrics. The comparison results for the OTCBVS image is listed in **Table 3**.

Table 1. Performance metrics for UN camp Image

Fusion method	Performance metrics				
	Entropy	SD	MI	Q^f_{AB}	Q_{piella}
ICA	6.23	22.65	0.61	0.41	0.86
LAP	6.85	30.47	2.22	0.49	0.91
DWT	6.84	30.47	2.55	0.40	0.81
DT-CWT	6.83	32.99	2.98	0.49	0.91
Region based ICA	6.53	30.41	4.16	0.57	0.92
Proposed scheme	7.18	31.41	7.17	0.61	0.95

Table 2. Performance metrics of multimodal Images in **Fig. 4 a and b**

Fusion method	Performance metrics				
	Entropy	SD	MI	Q^f_{AB}	Q_{piella}
ICA	7.41	50.64	3.64	0.44	0.92
LAP	7.78	70.19	2.91	0.58	0.94
DWT	7.75	68.18	2.96	0.50	0.92
DT-CWT	7.53	68.37	3.84	0.53	0.93
Region based ICA	7.64	67.43	5.68	0.60	0.95
Proposed scheme	7.98	69.96	7.57	0.66	0.96

Table 3. Performance metrics for OTCBVS images

Fusion method	Performance metrics				
	Entropy	SD	MI	Q^f_{AB}	Q_{piella}
ICA	7.02	50.42	3.44	0.42	0.84
LAP	7.49	69.32	4.98	0.71	0.93
DWT	7.49	69.19	4.50	0.69	0.91
DT-CWT	7.31	56.81	4.06	0.55	0.86
Region based ICA	7.43	63.71	7.64	0.72	0.95
Proposed scheme	7.53	68.73	7.93	0.75	0.96

Comparing the fusion evaluation metrics it is found that the proposed method has resulted in a fused image with a MI value of 7.93 a much higher score than any other method tested.

Experimental results show that the proposed region-based ICA-SVM scheme is able to extract most of the important information in the input images.

The proposed algorithm performs better than the ICA based fusion algorithms for the tested multimodal images, with higher scores in objective performance metrics. Also it performs much better than the various multiresolution methods compared.Visual comparison of the fused images shows that the proposed method is superior to the standard ICA (Mitianoudis and Stathaki, 2007), region based ICA method (Cvejic *et al*., 2007) and also other multiresolution transform based fusion methods.The subjective quality of the fused images is improved significantly by the use of intelligent decision making using SVM.

Experiments were also carried out by varying the number of training patches to show the variation in the fusion metrics Q^f_{AB} and Q_{piella} with number of training patches. The number of training patches is varied from 100 to 10000 and the fusion metrics Q^f_{AB} and Q_{piella} were determined. The performance plots of the three ICA based methods namely Standard ICA method Region based ICA and Region based ICA with SVM for UN camp images and OTCBVS images are presented in **Fig. 6 and 7**. The plots of Petrovic metric (Q^f_{AB}) and Piella metric (Q_{piella}) versus the number of training patches for the UN camp image are shown in **Fig. 6(a) and (b)** and for OTCBVS images is given in **Fig. 7(a) and (b)**. It is found that the Petrovic and Piella fusion metrics increase with the number of training patches. Similar performance is observed for the multimodal source images shown in **Fig. 4(a) and (b)**. This shows that the quality of the fused images would be further improved by increasing the number of ICA training patches.
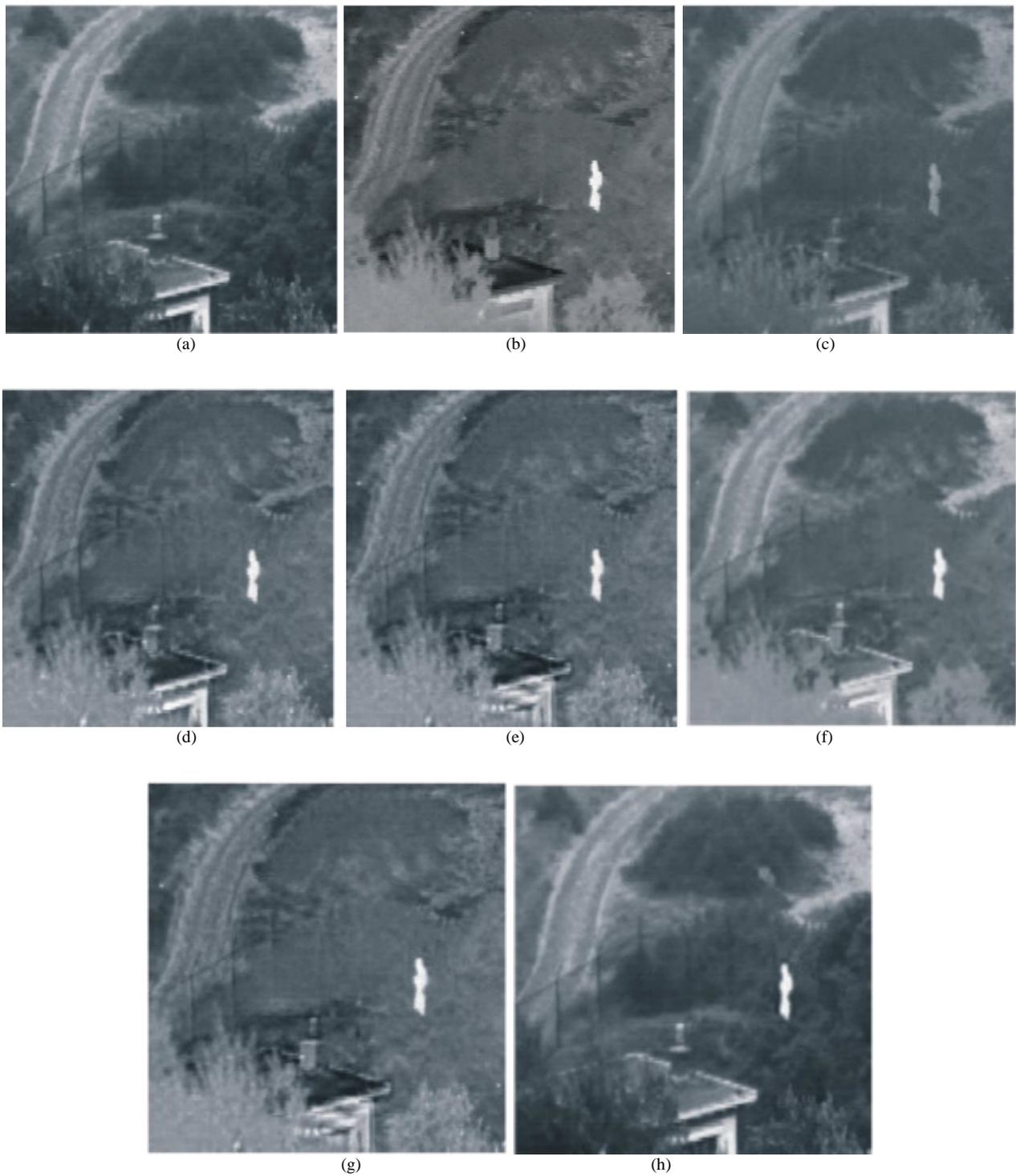
Fig. 3. UN camp Source Images (270×360, 256 levels) and fusion results (a) Source image (visual), (b) Source Image (IR) (c)-(h) fused images, (c) ICA, (d) LAP, (e) DWT, (f) DT-CWT, (g) Region based ICA without SVM, (h) Proposed method
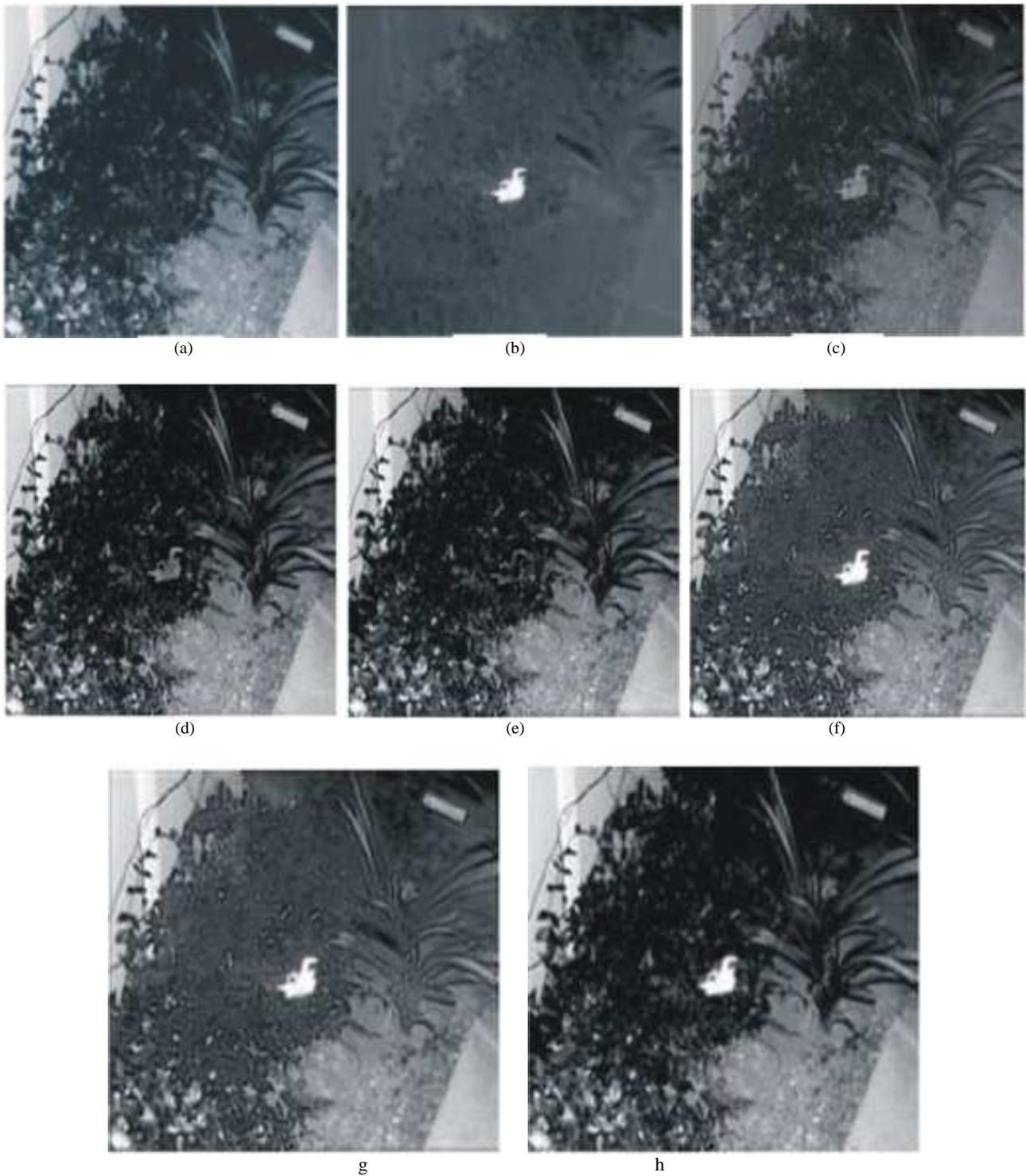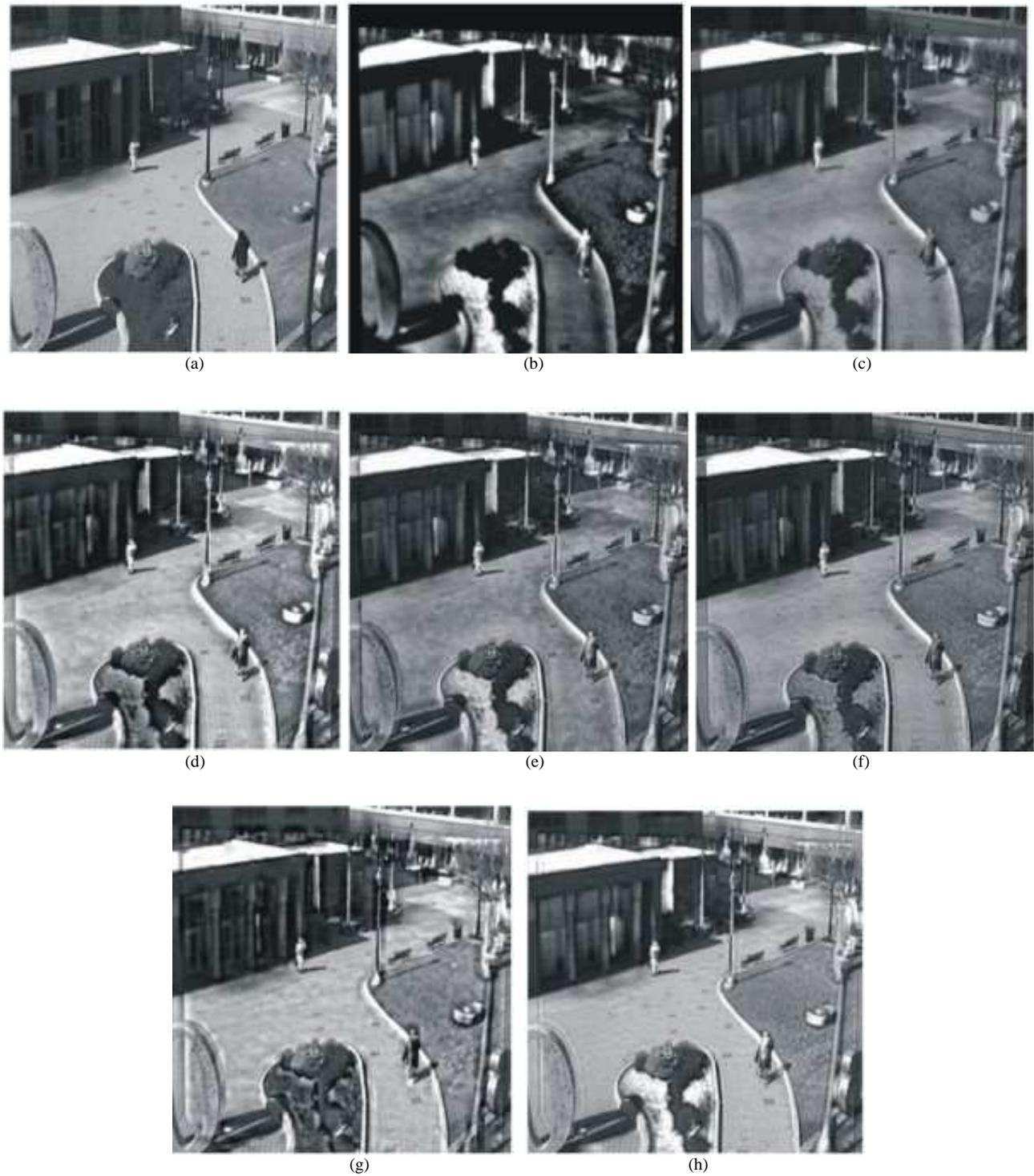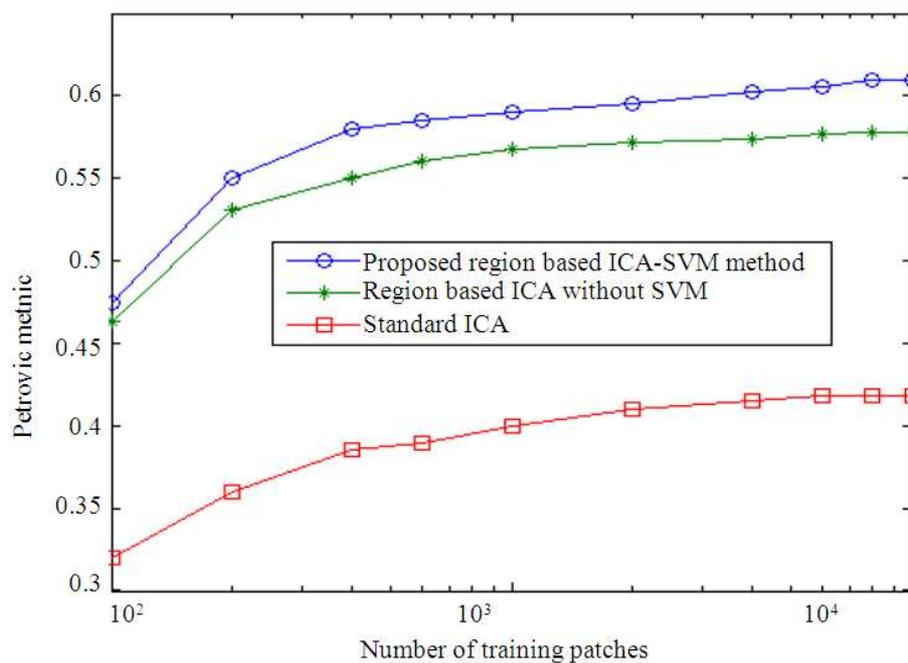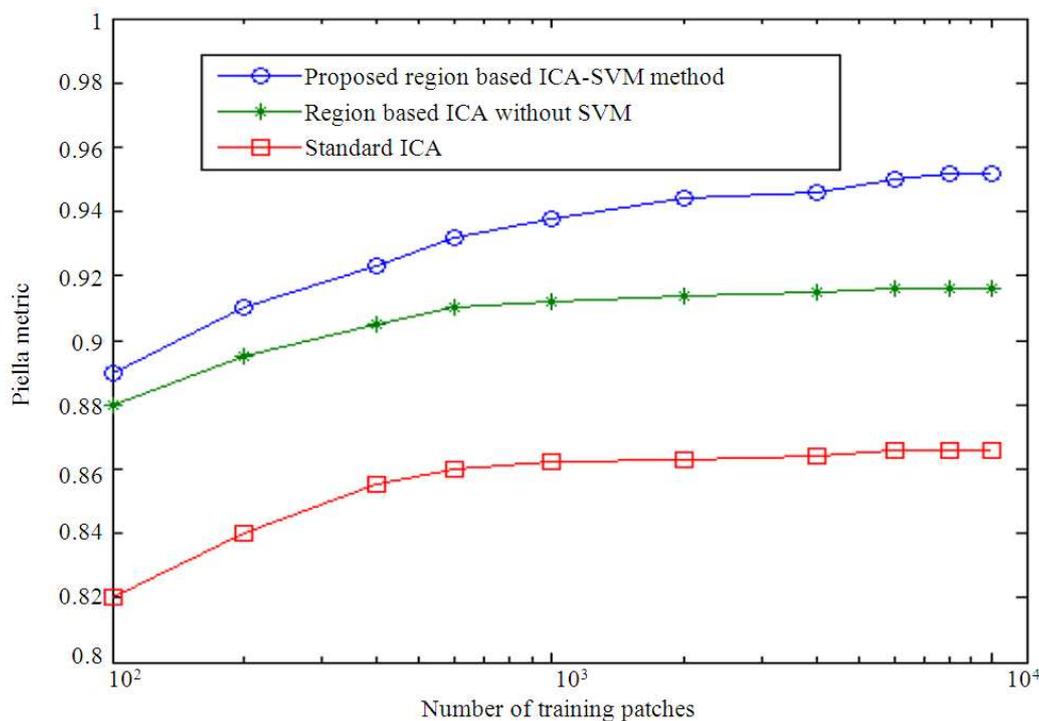
Fig. 4. Multimodal Source Images (220 × 320, 256 levels) and fusion results, (a) Source image (visual), (b) Source image (IR), (c)-(h) fused images, (c) ICA,(d) LAP, (e) DWT, (f) DT-CWT, (g) Region based ICA without SVM, (h) Proposed method

Fig. 5. OTCBVS Source Images (270 × 360, 256 levels) and fusion results, (a) Visual Image; (b) IR Image; and (c)-(h) are the fusion results (c) ICA, (d) LAP, (e) DWT, (f) DT-CWT, (g) Region based ICA without SVM, (h) Proposed method
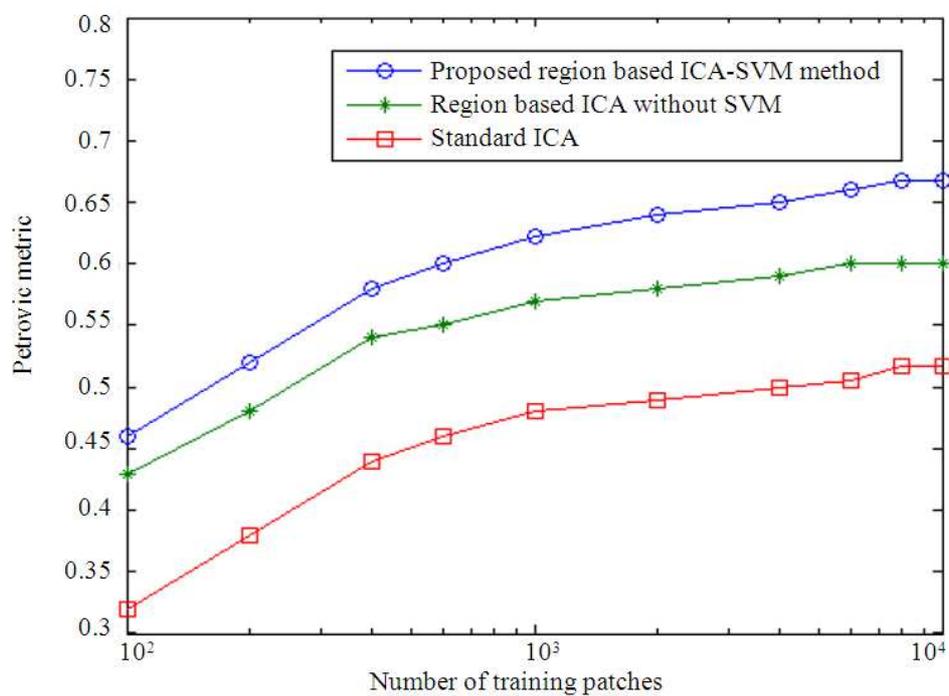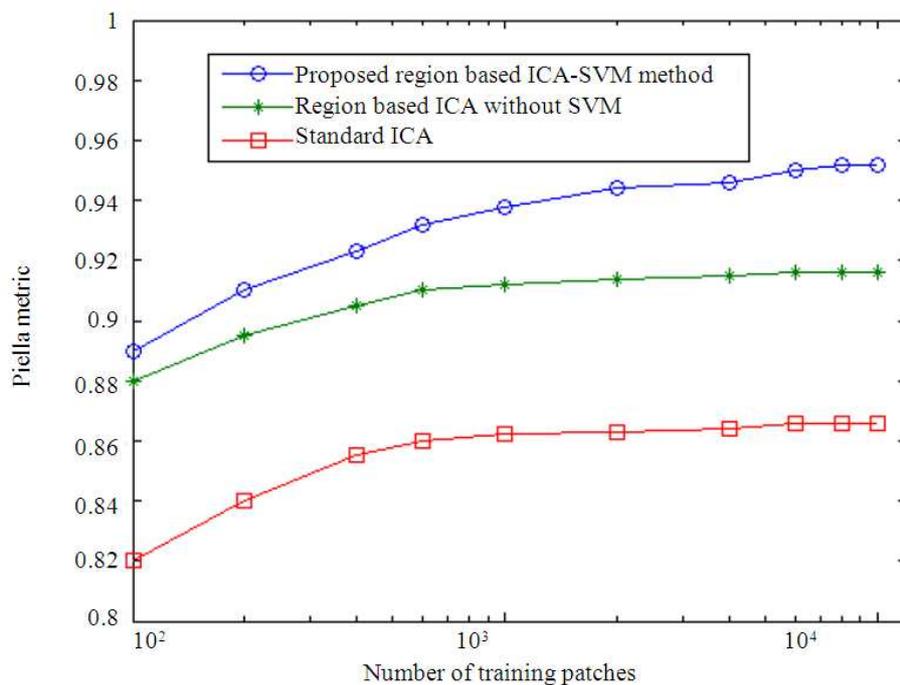
(a)

(b)

Fig. 6. Performance of UN camp images (a) Petrovic metric versus number of training patches (b) Piella metric versus number of training patches

(a)

(b)

Fig. 7. Performance of OTCBVS images (a) Petrovic metric versus number of training patches (b) Piella metric versus number of training patches

4. CONCLUSION

In this study, a novel image fusion framework using Independent Component Analysis and Support Vector Machines is introduced and its performance in multimodal image fusion is investigated. This region based approach employs segmentation in spatial domain and uses several statistical properties of regions such as energy, entropy and contrast, to make intelligent decisions by the SVM. Based on the SVM outputs, regions are selected from visible or IR images to form the fused representation. Experiments were conducted with number of multimodal images and the fused images obtained are with best visual quality compared to other tested methods. Results obtained show that the proposed multimodal scheme outperforms the standard ICA, individual multiresolution methods and region based fusion ICA in terms of the various fusion metrics.The fused images obtained by the proposed ICA-SVM method contain more useful information than the original multimodal images and hence subsequent processing tasks are expected to produce accurate results. The proposed method may appear to increase the computational load since the ICA bases are to be estimated and the SVM is to be trained. But these two functions are carried out only once and can be separately trained offline using the same training set and used for fusion of any sets of multimodal images. Hence, this method does not increase the computational cost. The proposed method can be extended further by applying it to different image types such as medical, remote sensing and multifocus images with more region features.

5. REFERENCES

Blum, R. and Z. Liu, 2005. Multi-Sensor Image Fusion and its Applications. 1st Edn., Taylor and Francis, London, UK., ISBN-10: 0849334179, pp: 528.

Burt, T. and E.H. Andelson, 1983. The laplacian pyramid as a compact image code. IEEE Trans. Commun., 31: 532-540. DOI: 10.1109/TCOM.1983.1095851

Chen, S., R. Zhang, H. Su, J. Tian and L. Yang, 2008. Improving empirical mode decomposition using support vector machines for multifocus image fusion. Sensors J., 8: 2500-2508. DOI: 10.3390/s8042500

Cvejic, N., D. Bull and N. Canagarajah, 2007. Region-based multimodal image fusion using ICA bases. IEEE Sensors J., 7: 743-751. DOI: 10.1109/JSEN.2007.897683

Drajic, D. and N. Cvejic, 2007. Adaptive fusion of multimodal surveillance image sequences in visual sensor networks. IEEE Trans. Consumer Electronics, 53: 1455-62. DOI: 10.1109/TCE.2007.4429237

Ellmauthaler, A., C.L. Pagliari and E.A.B. Da Silva, 2013. Multiscale image fusion using the undecimated wavelet transform with spectral factorization and nonorthogonal filter banks. IEEE Trans. Image Proc., 22: 1005-1017. DOI: 10.1109/TIP.2012.2226045

Gunn, S.R., 1998. Support vector machines for classification and regression. Technical Report, University of Southamptonn.

Hyvarinen, A. and E. Oja, 2000. Independent component analysis: Algorithms and applications. Neural Netw., 13: 411-430, DOI: 10.1016/S0893-6080(00)00026-5

Hyvarinen, A., 1999a. Survey on independent component analysis. Neural Comput. Surveys, 2: 94-128. DOI: 10.1007/978-3-540-30110- 3_4

Hyvarinen, A., 1999b. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw., 10: 626-634. DOI: 10.1109/72.761722

Jutten, C. and J. Herault, 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Process., 24: 1-10. DOI: 10.1016/0165-1684(91)90079-X

Lewis, J.L., R.J. O'Callaghan, S.G. Nikolov, D.R. Bull and N. Canagarajah, 2007. Pixel-and region-based image fusion with complex wavelets. Inform. Fusion, 8: 119-130. DOI: 10.1016/j.inffus.2005.09.006.

Li, H., B.S. Manjunath and S.K. Mitra, 1995. Multisensor image fusion using the wavelet transform. Graph. Models Image Proc., 57: 235-245. DOI: 10.1006/gmip.1995.1022

Li, S., J.T. Kwok, I.W. Tsang and Y. Wang, 2004. Fusing images with different focuses using support vector machines. IEEE Trans. Neural Netw., 15: 1555-1561. DOI: 10.1109/TNN.2004.837780

Lu, D. and Q. Weng, 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens., 28: 823-870. DOI: 10.1080/01431160600746456

Luo, Y., M. Liao, J. Yan and C. Zhang, 2012. A Multi-features fusion support vector machine method (MF-SVM) for classification of mangrove remote sensing image. J. Comput. Inform. Mation, Syst., 8: 323-334.

Mallat, S.G., 1998. A Wavelet Tour of Signal Processing. 2nd Edn., Academic Press, San Diego, ISBN-10: 0124666051, pp: 577.

Mitianoudis, N. and T. Stathaki, 2007. Pixel-based and Region-based Image Fusion schemes using ICA bases. Inform. Fusion, 8: 131-142. DOI: 10.1016/j.inffus.2005.09.001

Nirmala, D.E., B.S. Paul and V. Vaidehi, 2011. A novel multimodal image fusion method using shift invariant discrete wavelet transform and support vector machines. Proceedings of the International Conference on Recent Trends in Information Technology, Jun. 3-5, IEEE Xplore Press, Chennai, Tamil Nadu, pp: 932-937. DOI: 10.1109/ICRTIT.2011.5972405

O'Callaghan, R.J. and D.R. Bull, 2005. Combined morphological-spectral unsupervised image segmentation. IEEE Trans. Image Proc., 14: 49-63. DOI: 10.1109/TIP.2004.838695

Piella, G. and H. Heijmans, 2003. A new quality metric for image fusion. Proceedings of the International Conference on Image Processing, Sept. 14-17, IEEE Xplore Press, pp: 173-176. DOI: 10.1109/ICIP.2003.1247209

Piella, G., 2002. A region-based multiresolution image fusion algorithm. Proceedings of the International Conference on Information Fusion, Jul. 8-11, IEEE Xplore Press, Annapolis, pp: 1557-1564. DOI: 10.1109/ICIF.2002.1021002

Saha, A., G. Bhatnagar and Q.M.J. Wu, 2013. Mutual spectral residual approach for multifocus image fusion. Digital Signal Proc., 23: 1121-1135. DOI: 10.1016/j.dsp.2013.03.001

Wan, T., N. Canagarajah and A. Achim, 2009. Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients. IEEE Circ. Syst. Soc., 11: 624-633. DOI: 10.1109/TMM.2009.2017640

Xydeas, C.S. and V. Petrovic, 2000. Objective image fusion performance measure. Instit. Eng. Technol., 36: 308-309. DOI: 10.1049/el:20000267