Hierarchical Fusion of Multi Spectral Face Images for Improved Recognition Performance

Richa Singh, Mayank Vatsa and Afzel Noore[1]

Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV 26506

Abstract

This paper presents a two level hierarchical fusion of face images captured under visible and infrared light spectrum to improve the performance of face recognition. At image level fusion, two face images from different spectrums are fused using DWT based fusion algorithm. At feature level fusion, the amplitude and phase features are extracted from the fused image using 2D log polar Gabor wavelet. An adaptive SVM learning algorithm intelligently selects either the amplitude or phase features to generate a fused feature set for improved face recognition. The recognition performance is observed under the worst case scenario of using single training images. Experimental results on Equinox face database show that the combination of visible light and short-wave IR spectrum face images yielded the best recognition performance with an equal error rate of 2.86%. The proposed image-feature fusion algorithm also performed better than existing fusion algorithms.

Key words: Face recognition, image fusion, feature fusion, wavelet transform, support vector machine.

[1] Corresponding Author: Tel. +1 304 293 5263 ext. 2547 Fax: +1 304 293 8602
Email address: {richas, mayankv, noore}@csee.wvu.edu (R. Singh, M. Vatsa, A. Noore).

1 Introduction

A vast number of studies in face recognition have focused on improving the performance of face images captured under controlled environments using visible light. However, the challenges of improving the performance of face recognition under uninhibited and uncontrolled environments such as changes in facial expression, variations in pose, changes in ambient lighting, occlusions in face, and variations in disguise, have stimulated further on-going research [15], [19], [28]. Infrared (IR) images have been recently used to improve the performance of face recognition because IR face images capture unique heat signature patterns corresponding to the blood flow in internal anatomical structure such as veins and tissues. Recent studies have compared the performance of visible and IR-based face recognition under varying lighting conditions [11], [24], [25], over a passage of time [7], [8], and with disguises [18]. Other studies have attempted to improve the performance of face recognition by fusing the IR face image with visible face image at image level [13], [23], feature level [23], match score level [12], and at decision level [8].

All these studies require at least three face images for training. For commercial applications, it is feasible to obtain large training samples. However, law enforcement applications usually have to contend with limited availability of images or, in most cases, with only a single image for training. In this paper, we first propose a face recognition algorithm based on single training image based. It extracts amplitude and phase features from a face image to process the inherent unique characteristics of the face. These

features are independent of environmental constraints such as illumination and contrast. Further, the paper presents a novel approach for fusing information from face images captured under different light spectrum, at both the image and feature level. At image level fusion, the face images from two different spectrums, for example long-wave IR and visible spectrum, are fused using Discrete Wavelet Transform (DWT). Image level fusion algorithm captures salient features of both the images to recognize a person. In feature level fusion, we propose a learning based algorithm which uses Support Vector Machine (SVM) [9] to fuse two or more feature vectors from training samples that are labeled as good and bad. SVM fuses the phase and amplitude feature vectors extracted from the face image using 2D log polar Gabor wavelet. The proposed two-level hierarchical image and feature fusion significantly enhances the performance of face recognition.

The paper is divided into four sections. Section 2 describes the proposed face recognition algorithm, Section 3 presents the feature fusion and image fusion algorithms, and Section 4 presents the experimental results for evaluating the performance of face recognition and fusion algorithms. Results are also compared with other existing face recognition and fusion algorithms.

2 Face Recognition using Amplitude and Phase Features

In this section we describe the proposed face recognition algorithm which is based on the amplitude and phase information extracted from a face image. In a given image, first the

face is detected using triangle based face detection algorithm [10]. Let the size of detected face image $I(x,y)$ be N x N. Face image is transformed into its log polar form. The log polar transformation converts original face image $I(x,y)$ into $I(r,\theta)$ in which the angular coordinates are represented on the vertical axis and the logarithmic coordinates of radius are represented on the horizontal axis. More precisely, with respect to a center point (x_c, y_c),

$$r = \log\sqrt{(x-x_c)^2 + (y-y_c)^2}\ , \quad 0 \le r \le r_{\max}$$
$$\theta = \tan^{-1}\left(\frac{y-y_c}{x-x_c}\right), \quad 0 \le \theta \le 2\pi \qquad \textbf{(1)}$$

Here a crucial decision is the choice of center point (x_c, y_c). In a face image, there are very few points such as the center of mouth, eye locations, and nose tip which remain consistent and are easy to locate. The proposed face detection algorithm uses the coordinates of both eyes to detect a face image. The best choice is the center (x_c, y_c) of line joining the two eye coordinates and the proposed algorithm uses this point as the center point. The mapping process converts the scale change and rotation of face image into constant horizontal and cyclic vertical shifts in the log polar plane. This process eliminates the need for using multiple templates to cover scale and rotation variations of face image.

In the next step, polar face image is expressed into 2D Fourier domain,

$$I(\rho,\phi) = \frac{1}{\mu\lambda}\sum_r\sum_j I(r,\theta_j)\exp\left[-i\frac{2\pi r}{\mu}\rho + \frac{2\pi j}{\lambda}\phi\right] \qquad \textbf{(2)}$$

where, $\theta_j = {2\pi j}/{\lambda}$, $0 \le r < \mu$, $0 \le j < \lambda$, μ and λ are the radial and angular frequency resolutions respectively.

2D Fourier transform of face image is convolved with the 2D Fourier transform of log polar Gabor wavelet $G(\rho,\phi)$. Log polar Gabor is a form of log Gabor wavelet [4] which is based on polar coordinates and the dependency of directional independent variance (σ) on the polar coordinate is realized by a logarithmic scale. Thus the functional form of 2D log polar Gabor filter can be represented as,

$$G_{r_0,\theta_0}(r,\theta) = \exp\left(-2\pi^2\sigma^2\left[\left(\ln(r)-\ln\left(r_0\right)\right)^2 + \left(\ln(r)\sin\left(\theta-\theta_0\right)\right)^2\right]\right) \quad \textbf{(3)}$$

and the position of filter in the Fourier domain is defined by

$$r_{00} = \sqrt{2}, \qquad r_{0i} = 2^i r_{00}, \qquad \theta_{0i} = i\frac{2\pi}{N_\theta} \quad \textbf{(4)}$$

where r_{00} is the smallest possible frequency, N_θ is the number of filters on the unit circle, and at index L, σ_L and s_L are further defined by

$$\sigma_L = \frac{1}{\ln(r_0)\pi\sin\left({\pi}/{N_\theta}\right)}\sqrt{\frac{\ln 2}{2}} \quad \textbf{(5a)}$$

$$s_L = \frac{2\ln(r_0)\pi\sin\left({\pi}/{N_\theta}\right)}{\ln(2)}\sqrt{\frac{\ln 2}{2}} \quad \textbf{(5b)}$$

Inverse Fourier Transform (IFT) of the convolved face image is computed and the output $I_g(r,\theta)$ is a complex valued matrix containing the amplitude and phase information.

$$I_g(r,\theta) = IFT\left(\sum_r \sum_j I(\rho,\phi) * G(\rho,\phi)\right) \quad \textbf{(6)}$$

Amplitude feature is computed from the matrix $I_g(r,\theta)$ using Equation 7. Phase information is extracted using Equation 8 and then quantized to obtain the binary phase feature or template using Equation 9 which is used for recognition. The amplitude and phase features are shown in Fig. 1.

$$F_A(r,\theta) = \sqrt{\left(\text{Re}\, I_g(r,\theta)\right)^2 + \left(\text{Im}\, I_g(r,\theta)\right)^2} \quad \textbf{(7)}$$

$$P(r,\theta) = \tan^{-1}\left(\frac{\text{Im}\, I_g(r,\theta)}{\text{Re}\, I_g(r,\theta)}\right) \quad \textbf{(8)}$$

$$F_P(r,\theta) = \begin{cases} [1,\ 1] & \textit{if} \quad 0^0 < \text{Re}\,(P(r,\theta)) \le 90^0 \\ [0,1] & \textit{if} \quad 90^0 < \text{Re}\,(P(r,\theta)) \le 180^0 \\ [0,0] & \textit{if} \quad 180^0 < \text{Re}\,(P(r,\theta)) \le 270^0 \\ [1,0] & \textit{if} \quad 270^0 < \text{Re}\,(P(r,\theta)) \le 360^0 \end{cases} \quad \textbf{(9)}$$

Fig. 1. (a) Face image (b) Amplitude features, (c) Phase features

To match the two amplitude feature templates, F_{A1} and F_{A2} ,

a) The amplitude features are divided into *z* number of frames, each of size *k* x *l*.

b) Corresponding frames from the two amplitude templates are matched using correlation distance CD_A^i given in Equation 10.

$$CD_A^i = \frac{F_{A1}{}^i \otimes F_{A2}{}^i}{k\ l}, \quad i = 1, 2, \ldots, z \tag{10}$$

c) The amplitude matching score MS_A is calculated using Equation 11.

$$MS_{IA} = \begin{cases} MS_{IA} + 1 & if \quad CD_A^i \geq T_{IA} \\ MS_{IA} & if \quad CD_A^i < T_{IA} \end{cases}$$
$$MS_A = \frac{MS_{IA}}{z} \tag{11}$$

where T_{IA} is the frame matching threshold, MS_{IA} is the intermediate amplitude matching score, and *z* is the number of frames.

d) A match occurs if the amplitude matching score, MS_A, is greater than the matching threshold, T_A.

To match the two phase feature templates, F_{P1} and F_{P2},

a) The phase templates are divided into *m* frames, each of size *p* x *q*.

b) Corresponding frames from the two phase templates are matched using hamming distance given in Equation 12.

$$HD_P^i = \frac{\sum F_{P1}{}^i \oplus F_{P2}{}^i}{p\,q}, \quad i = 1, 2, \ldots, m \tag{12}$$

where $F_{P1}{}^{i}$ and $F_{P2}{}^{i}$ are the i^{th} frames for the two phase templates, and HD_P^i is the corresponding hamming distance measure.

c) The phase matching score MS_P is calculated using Equation 13,

$$MS_{IP} = \begin{cases} MS_{IP} + 1 & if \quad HD_P^i \le T_{IP} \\ MS_{IP} & if \quad HD_P^i > T_{IP} \end{cases}$$
$$MS_P = \frac{MS_{IP}}{m} \qquad (13)$$

where T_{IP} is the frame matching threshold, MS_{IP} is the intermediate phase matching score, and m is the number of frames.

d) A match occurs if the phase matching score, MS_P, is greater than the matching threshold T_P.

The advantage of the proposed face recognition algorithm is that the amplitude and phase features extracted from the filter response of face image are invariant to frequency, scale and orientation of the filter. Here we transform the problem of recognizing a face into an efficient test of statistical dependence operating on amplitude variation and statistical independence operating on phase variation. Also, Venkatesh and Owens [26] show that the phase is independent of the overall magnitude of face image and provides invariance to the changes in illumination and contrast. Log polar transformation eliminates the necessity of using multiple templates to handle the variations due to scale and rotation. Thus the amplitude features in the proposed face recognition algorithm are invariant to frequency, scale, and orientation; while the phase features are invariant to frequency, scale, orientation, illumination and contrast.

3 Fusion at Image level and Feature Level for Enhanced Performance

In [20], it has been suggested that biometric information fusion can be used to enhance the recognition performance. Information fusion can be performed at five levels namely data or image level, feature level, match score level, decision level, and rank level. Researchers have proposed several algorithms [5], [14], [20] for fusion at match score level and decision level. Kumar [16], Ross [21], and Singh [23] proposed algorithms for feature level fusion, and Chang [6] and Singh [23] proposed algorithms for image level fusion. In this section, we present algorithms for both feature and image level fusion in face biometrics. The two levels of fusion are further combined hierarchically to increase the recognition performance.

3.1 Fusion of Features from Single Spectrum Face Image

Feature fusion in biometrics has been addressed by several researchers [21], [23] in the literature. In [21], feature fusion is performed by concatenating two feature vectors. Singh *et al.* [23] fuse the eigen vectors of visible and IR face image using genetic algorithms. In the proposed approach, a learning based algorithm is used to fuse the amplitude and phase features from a face image. There are several learning algorithms based on neural networks which can be trained to learn and identify features. However, one of the major drawbacks with neural network based approach is choosing a mapping function between the input data space and the output feature space. The mapping function may be inefficient in determining the local minima consistently. When a large set of

training data is used, it could lead to overfitting and hence poor generalization affecting the matching performance. Learning algorithms have several parameters which are controlled heuristically, making the system difficult and unreliable to use. Also, traditional multilayer neural networks suffer from the existence of multiple local minima solutions. To alleviate these inherent limitations, we use Support Vector machine (SVM) based learning algorithms for feature fusion. SVM training always finds a global minimum by choosing a convex learning bias. In classical SVM, the goal is to minimize the probability of making an error. Furthermore, the parameters of SVM can be set to penalize the errors asymmetrically by assigning costs to different errors to minimize the expected misclassification cost. This approach, known as dual υ-Support Vector Machines (2υ-SVM) [9], can also address the difficulties that arise when the class frequencies in training data do not accurately reflect the true prior probabilities of the classes. Considering all these properties and advantages, our proposed algorithm uses 2υ-SVM [9]. 2υ-SVM is applied as a two class problem, classifying good feature class and bad feature class. Proposed by Chew *et al*. [9], 2υ-SVM can be expressed as follows:

Let $\{x_i, y_i\}$ be a set of *n* data vectors with $x_i \in \Re^d$, $y_i \in +1, -1$ and $i = 1,...,n$. x_i is the i^{th} data vector that belongs to a binary class y_i. The objective of training 2υ-SVM is to find the hyper-plane that separates any two classes with the widest margins, i.e.,

$$\omega(x) + b = 0$$

subject to $$y_i(\omega\,\varphi(x) + b) \geq \rho - \psi_i,\ \psi_i \geq 0, \tag{14}$$

to minimize $$\frac{1}{2}\|\omega\|^2 - \sum_i C_i(\upsilon\rho - \psi_i)$$

where ρ is the position of the margin and υ is the error parameter. φ is the mapping function to map the data space to the feature space. It is used to provide generalization to the decision function that may not be a linear function of the training data. $C_i(\upsilon\rho - \psi_i)$ is the cost of errors, ω is normal vector, b is bias, and ψ_i is the slack variable for classification errors. υ is the error parameter that can be calculated using υ_+ and υ_-, error parameters for training of positive and negative classes respectively.

$$\upsilon = \frac{2\upsilon_+\upsilon_-}{\upsilon_+ + \upsilon_-}, \qquad 0 < \upsilon_+ < 1 \quad and \quad 0 < \upsilon_- < 1 \tag{15}$$

Error penalty C_i is calculated as,

$$C_i = \begin{cases} C_+, & if \quad y_i = +1 \\ C_-, & if \quad y_i = -1 \end{cases} \tag{16}$$

where,

$$\begin{aligned} C_+ &= \left[n_+ \left(1 + \frac{\upsilon_+}{\upsilon_-} \right) \right]^{-1} \\ C_- &= \left[n_- \left(1 + \frac{\upsilon_-}{\upsilon_+} \right) \right]^{-1} \end{aligned} \tag{17}$$

and n_+ and n_- are the number of training points for the good and bad classes respectively. 2υ-SVM training can be formulated as,

$$\max_{(\alpha_i)}\left\{-\frac{1}{2}\sum_{i,j}\alpha_i\,\alpha_j\,y_i\,y_j K\left(x_i,x_j\right)\right\}$$

$$0 \le \alpha_i \le C_i \qquad \textbf{(18)}$$

where,

$$\sum_i \alpha_i\, y_i = 0$$

$$\sum_i \alpha_i \ge \upsilon$$

$i, j \in 1, \ldots, n$ and the kernel function is

$$K\left(x_i, x_j\right) = \varphi\left(x_i\right)\varphi\left(x_j\right) \qquad \textbf{(19)}$$

2υ-SVM is initialized and optimized using iterative decomposition training [9], which leads to reduced complexity of 2υ-SVM. If the size of data vectors is n, then the complexity without optimization is $O(n^2)$ and with optimization is $O(n)$ [9]. In this section, we consider the fusion of phase and amplitude features extracted from a face image. The proposed fusion algorithm, shown in Fig. 2, is described as follows:

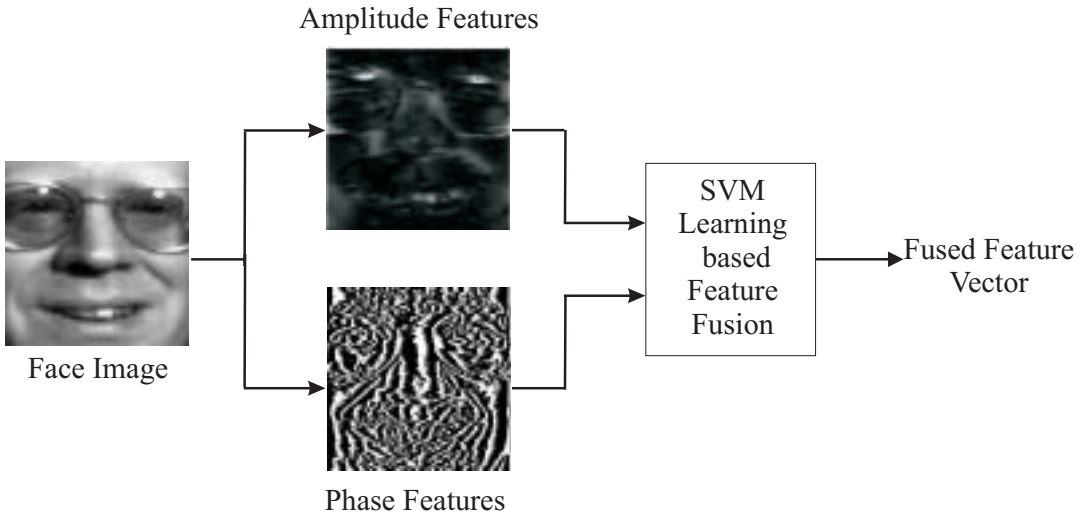

Fig. 2. Block diagram for feature level fusion

Let amplitude features F_A and phase features F_P be extracted from a face image. Both the features are divided into windows of size 3 x 3 to calculate the weighted average window based activity. A total of n activity levels are computed and provided as input to the 2υ-SVM which is then trained to determine whether the coefficient from the amplitude feature set F_A or the phase feature set F_P should be used. At any position (x, y), if the features of F_A are classified as "good" and the SVM classification margin of F_A is greater than the SVM classification margin of F_P, then output $O(x, y)$ of the learning algorithm is 1 otherwise the output is -1. As shown in Equation 20, amplitude and phase features are selected depending on the output $O(x, y)$.

$$FF(x,y) = \begin{cases} F_A(x,y), & if \quad O(x,y) > 0 \\ F_P(x,y), & if \quad O(x,y) < 0 \end{cases} \qquad \textbf{(20)}$$

where FF is the fused feature vector. Furthermore, to match the two fused feature vectors, FF_1 and FF_2, the correlation based matching technique is applied. The features are first divided into m frames, each of size k x l. The correlation distance, CD_F^i, between two corresponding frames is computed using Equation 21. Using the frame matching threshold T_{IF}, the intermediate matching score MS_{IF} for the frames and the final matching score MS_F is calculated using Equation 22. A person is said to be matched if MS_F is greater than the fused feature vector matching threshold, T_F.

$$CD_F^i = \frac{FF_1^i \otimes FF_2^i}{k\ l}, \quad i = 1, 2, \ldots, m \qquad \textbf{(21)}$$

$$MS_F = \begin{cases} MS_{IF} + 1 & if \quad CD_F^i \geq T_{IF} \\ MS_{IF} & if \quad CD_F^i < T_{IF} \end{cases}$$
$$MS_F = \frac{MS_{IF}}{m} \tag{22}$$

3.2. Fusion of Features from Multi Spectral Face Images

To enhance the performance of face recognition, the amplitude and phase features of face images from different spectrum are fused as shown in Fig. 3. Features are extracted from two single-spectrum face images using the algorithm described in Section 2. For each image, the extracted amplitude and phase features are fused using the algorithm described in Section 3.1 to generate two fused feature vectors FF_V and FF_{IR} in visible and infrared spectrum respectively. These fused feature vectors are then combined using Equation 25, to generate an optimal fused feature vector, FF_{V-IR},

$$FF_V = \begin{cases} F_{A-V}, & if \quad O_1(x,y) > 0 \\ F_{P-V}, & if \quad O_1(x,y) < 0 \end{cases} \tag{23}$$

$$FF_{IR} = \begin{cases} F_{A-IR}, & if \quad O_2(x,y) > 0 \\ F_{P-IR}, & if \quad O_2(x,y) < 0 \end{cases} \tag{24}$$

$$FF_{V-IR} = \begin{cases} FF_V, & if \quad O_3(x,y) > 0 \\ FF_{IR}, & if \quad O_3(x,y) < 0 \end{cases} \tag{25}$$

where $O_1(x,y), O_2(x,y)$ and $O_3(x,y)$ are the outputs of the three trained 2υ-SVM. These feature vectors are further matched using the correlation based matching described in

Equations 21 and 22. The SVM based feature fusion approach can handle images that have variations in occlusion and expression.

3.3 Fusion of Multi Spectral Face Images

Image level fusion for the IR and visible face images has been proposed by various researchers [12], [13], [23]. These algorithms are generally based on learning techniques such as genetic algorithms which are computationally intensive. In this paper, we propose an algorithm to fuse the visible and IR face image using DWT with mother wavelet Daubechies 9/7 [2]. This mother wavelet is chosen because it is among the best filters for wavelet based operations such as image compression when operating in a distributed environment [27].

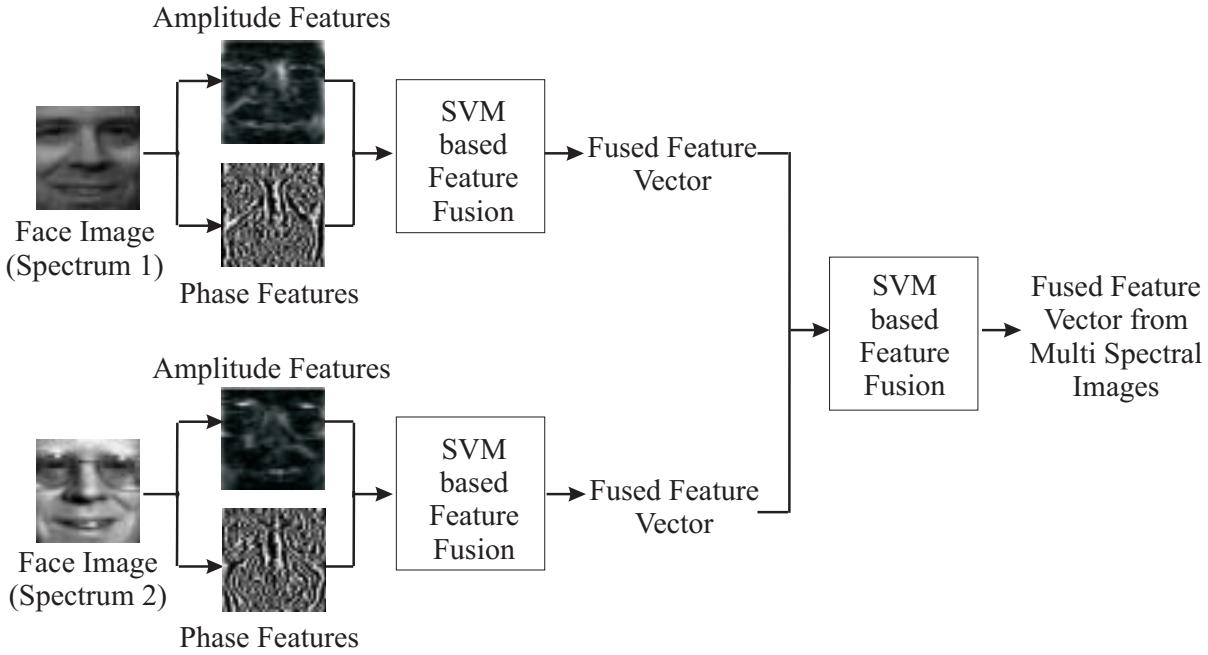

Fig. 3: Feature fusion from multi spectral face images

Since the visible and IR imaging devices usually have different spatial resolutions and viewpoints, face images may exhibit different pose and size. Before fusion, face images should be normalized and aligned. Face region is first detected using the triangle based face detection algorithm [10]. Preprocessing is then performed using affine transformation, which also handles the rotation, translation and scaling. Let $(x_1, y_1), (x_2, y_2)$, and (x_3, y_3) be the eyes and mouth coordinates respectively in the visible image and $(x_1', y_1'), (x_2', y_2')$, and (x_3', y_3') be the corresponding coordinates in the IR face image. We calculate a transformation matrix *TM* to align the two detected face images,

$$TM = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} \times \begin{bmatrix} x_1' & x_2' & x_3' \\ y_1' & y_2' & y_3' \\ 1 & 1 & 1 \end{bmatrix}^{-1} \quad \textbf{(26)}$$

which gives

$$TM = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \end{bmatrix} \quad \textbf{(27)}$$

This transformation is generalized to the whole image,

$$\begin{aligned} x &= u_1 x' + v_1 y' + w_1 \\ y &= u_2 x' + v_2 y' + w_2 \end{aligned} \quad \textbf{(28)}$$

Let I_V and I_{IR} be the preprocessed visible and IR face images. First, the pixel values of I_V and I_{IR} are transformed in the range of (0, 1). Single level DWT is then applied on these images to obtain the detail and approximation wavelet bands for both the images. Let

$I_{LL-V}, I_{LH-V}, I_{HL-V}$ and I_{HH-V} be the four bands from visible face image and $I_{LL-IR}, I_{LH-IR}, I_{HL-IR}$ and I_{HH-IR} be the corresponding bands from IR face image as shown in Fig. 4. To preserve the features from both the images, coefficients from approximation band of I_V and I_{IR} are averaged,

$$I_{LL-F} = mean(I_{LL-V}, I_{LL-IR}) \quad \textbf{(29)}$$

where I_{LL-F} is the approximation band of the fused image. For the three detailed bands, each band is divided into windows of size 3 x 3 and the sum of absolute value of all the pixels in each window is calculated. Binary decision maps, *DM*, are generated for all the three detail bands using Equations 30, 31, and 32. A value '1' is assigned if the window from visible image has a value greater than the corresponding window from the IR image; otherwise a value '0' is assigned.

$$DM_{LH} = \begin{cases} 1 & if \quad \max_{3\times3}(I_{LH-V}) > \max_{3\times3}(I_{LH-IR}) \\ 0 & if \quad \max_{3\times3}(I_{LH-V}) < \max_{3\times3}(I_{LH-IR}) \end{cases} \quad \textbf{(30)}$$

$$DM_{HL} = \begin{cases} 1 & if \quad \max_{3\times3}(I_{HL-V}) > \max_{3\times3}(I_{HL-IR}) \\ 0 & if \quad \max_{3\times3}(I_{HL-V}) < \max_{3\times3}(I_{HL-IR}) \end{cases} \quad \textbf{(31)}$$

$$DM_{HH} = \begin{cases} 1 & if \quad \max_{3\times3}(I_{HH-V}) > \max_{3\times3}(I_{HH-IR}) \\ 0 & if \quad \max_{3\times3}(I_{HH-V}) < \max_{3\times3}(I_{HH-IR}) \end{cases} \quad \textbf{(32)}$$

where DM_{LH}, DM_{HL}, and DM_{HH} are the binary decision maps for three detail bands, shown in Fig. 5.

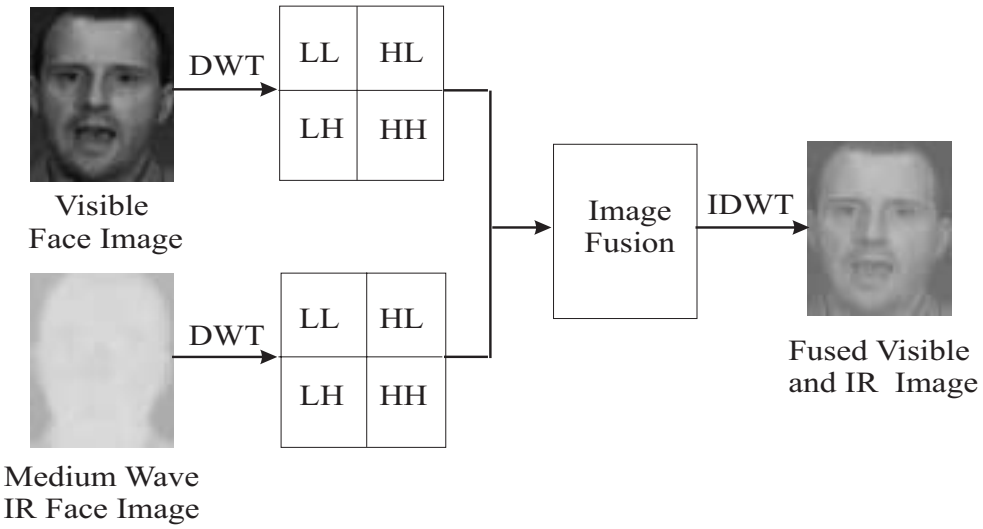

Fig. 4. Multispectral image level fusion of visible and medium-wave infrared face images

Fig. 5. Decision maps for the horizontal, vertical and diagonal bands

Based on these decision maps, the detailed bands for the fused image I_{LH-F}, I_{HL-F} and I_{HH-F} are generated. Inverse DWT is then applied on the four fused bands to generate the fused image.

$$I_F = IDWT(I_{LL-F}, I_{LH-F}, I_{HL-F}, I_{HH-F}) \qquad \textbf{(33)}$$

To validate the image fusion algorithm, recognition performance of the fused face image is calculated. Amplitude and phase features are extracted from the fused image and matched independently as described in Section 2. Face images captured using different

light spectrum yields complementary information that is useful for identification. For improved recognition performance, the fusion of this distinct multi-spectral information into a single image requires that the same image be captured using different light spectrum.

For an image of size N x N, the computational complexity of DWT and IDWT is $O\left(\frac{N^2}{4}\right)$[1]. The fusion process has a complexity of $O(N \log N)$. Hence the overall complexity of the proposed image fusion algorithm is $O\left(\frac{N^2}{4}\right)$ when $N > 8$. This shows that the image fusion algorithm is fast and can be used for real time applications.

3.4 Integration of Image Fusion and Feature Fusion

The algorithms proposed for both image level fusion and feature level fusion are combined to further enhance the recognition performance. Fig. 6 shows the block diagram to hierarchically combine both the image and feature fusion techniques. Face images from different spectrums, e.g. short-wave IR and long-wave IR, are fused using the image fusion algorithm described in Section 3.3. The amplitude and phase features extracted from the fused image are then fused using the feature fusion algorithm described in Section 3.1 to generate a composite feature vector. Correlation technique described in Section 3.1 is used to match the two fused feature vectors.

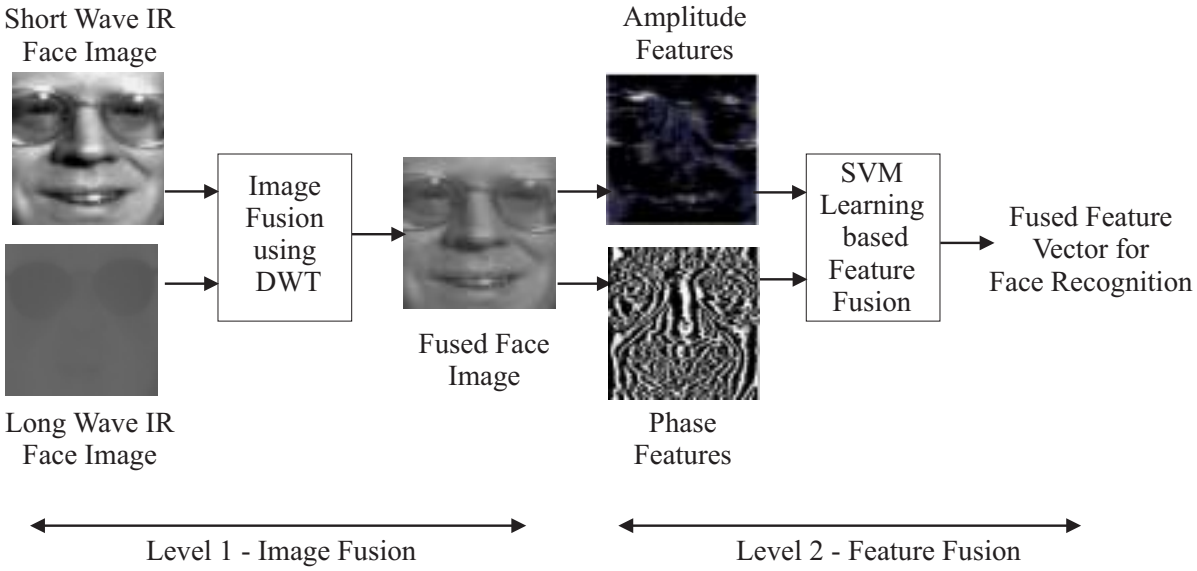

Fig. 6. Two-level hierarchical fusion of multi spectral face images

4 Experimental Results

Experiments are performed on the face database from Equinox Corporation [29]. The database contains images captured in long-wave, medium-wave and short-wave infrared light along with the images in the visible spectrum. Visible spectrum captures the electromagnetic energy in the range 0.4 - 0.7 μm, and infrared spectrum comprises of thermal and reflectance wavebands in the range of 0.7 - 14.0 μm. Short-wave infrared (SWIR) which is in the range of 0.9 - 2.4 μm is reflective. Thermal IR spectrum is divided into two parts, medium-wave infrared (MWIR) from the spectral range of 3.0 - 5.0 μm and long-wave infrared (LWIR) from 8.0 - 14.0 μm. Emitted radiation depends on the temperature and emissivity of the material. There are images from 91 individuals in the database with variations in facial expression, illumination, and occlusion due to glasses for all four light spectrums. The results are divided into two parts. First part

presents the performance of the proposed face recognition algorithm for different variations on the images from four light spectrums. We also present the recognition performance with variation in number of training images. Second part presents the performance of face recognition with the image and feature fusion algorithms.

4.1 Performance of Face Recognition Algorithm

Face is detected from the image using the triangle based face detection algorithm [10]. Performance of the proposed face recognition algorithm is then computed for different variations in all four spectrums. Fig. 7 shows an example of the amplitude and phase features for images in all four spectrums.

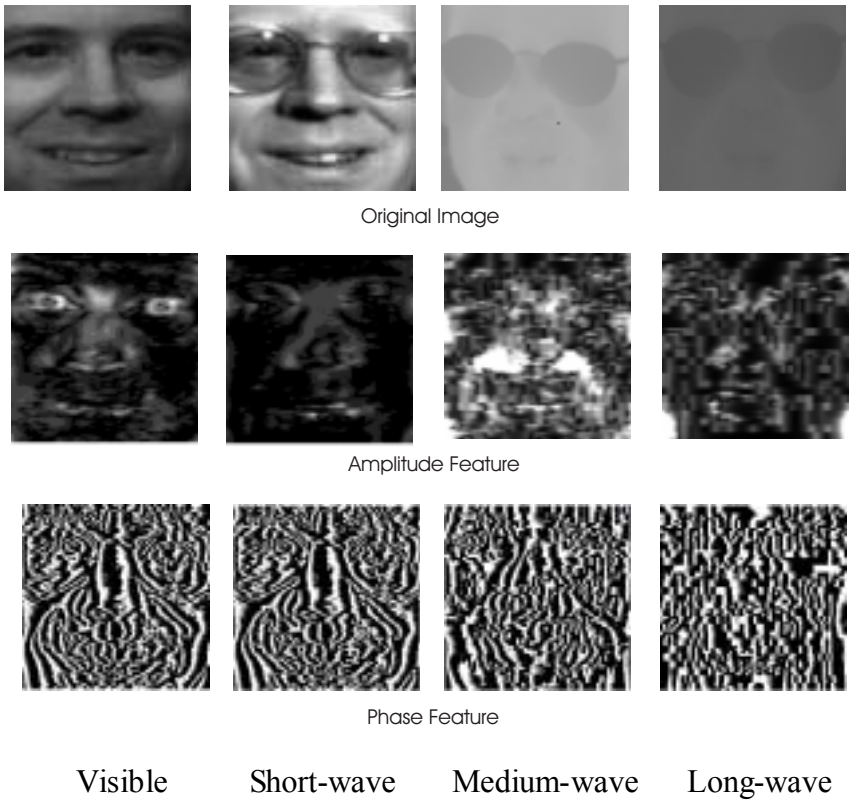

Fig. 7. Face image from four spectrums with amplitude and phase features

Figs. 8 and 9 show the Receiver-Operating Characteristic (ROC) curves using amplitude and phase features respectively for face recognition. These two ROC plots show that the IR light spectrum contains more invariant features that are useful for face recognition compared to the visible light spectrum. In IR light spectrum, the performance of long-wave and medium-wave spectrum images are nearly the same, whereas the best results are obtained with short-wave IR images.

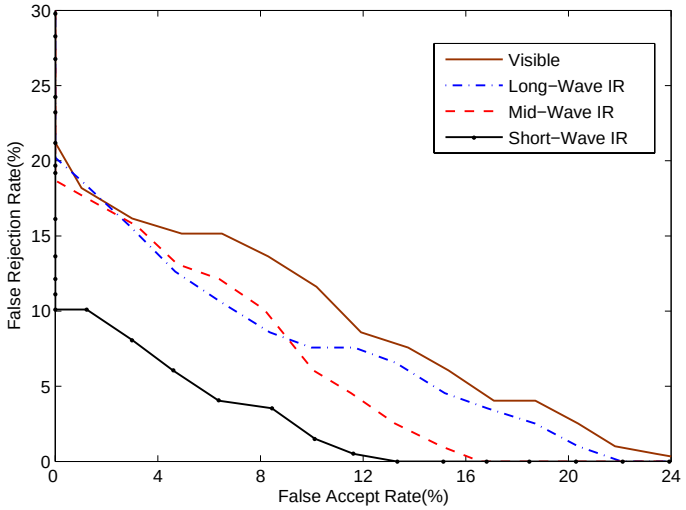

Fig. 8. ROC for amplitude features

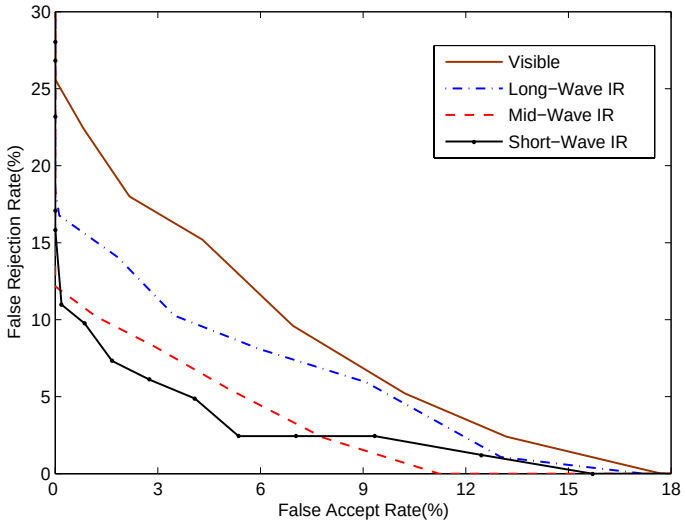

Fig. 9. ROC for phase features

To evaluate the performance of recognition algorithm with different variations, we further divide the database into three parts: images containing variation in expression, variation in illumination, and variation in occlusion due to glasses. The performance is compared using equal error rate (EER) which indicates the value for which the false acceptance rate is equal to the false rejection rate. Lower value of equal error rate indicates higher accuracy of the recognition algorithm. We used the equal error rate for the group with minor variations in expression as the baseline reference to compare and analyze other experimental results. Table 1 shows the equal error rates of all four spectrums with different variations. For the complete visible spectrum database, an EER of 12.72% and 11.58% is achieved using amplitude and phase features respectively. The algorithm is also found to perform better than Principle Component Analysis (PCA) [3] and Linear Discriminant Analysis (LDA) based algorithms [3]. Similarly, for the three IR spectrums, amplitude and phase based face recognition algorithm performs three to four times better than the PCA and LDA based face recognition algorithms. With long-wave IR (LWIR) and medium-wave IR (MWIR) face images, the equal error rate is higher when occlusions are present. This shows that the IR face images do not provide additional information when glasses are used, which is theoretically true because IR images are opaque to glass. The results show that for visible and short wave IR images, illumination causes the equal error rate of amplitude features to increase compared to the phase features. Also, the results show that long wave and medium wave IR spectrum are invariant to illumination variations since they capture the thermal property of the face image. Table 1 also shows that in all cases, minimum equal error rate is obtained with short-wave IR (SWIR) face images.

In specific applications such as law enforcement, only one image per person may be available for training the face recognition system [17], [22], [30]. Most of the face recognition algorithms may not work well in such a scenario because they require large number of example images per class so that the intraclass variability could be considered against the interclass similarity. We evaluated the performance of proposed face recognition algorithm for such worst case scenario using short-wave IR face images. Fig. 10 shows the identification accuracy of face recognition on varying the number of training images from seven to one. With seven training images, the identification accuracy of amplitude and phase features is nearly 100%. However as the number of training images decrease, the accuracy also decreases.

Table 1

Comparison of performance with variations in image features under different light spectrum.

Algorithms	**Variation**	**Equal Error Rate (%)**			
		Visible Spectrum	**Long-wave IR**	**Mid-wave IR**	**Short-wave IR**
PCA [3]	Expression	35.83	31.04	27.43	21.55
	Illumination	35.86	31.26	27.89	21.58
	Occlusion	35.92	32.79	28.90	21.61
	All variations	35.87	31.96	28.32	21.59
LDA [3]	Expression	31.22	26.13	23.83	18.33
	Illumination	31.37	26.69	23.99	18.32
	Occlusion	31.39	27.09	24.57	18.34
	All variations	31.24	26.71	24.04	18.33
Proposed Amplitude Feature	Expression	12.69	10.22	10.13	6.01
	Illumination	13.07	10.36	10.21	6.22
	Occlusion	12.65	10.97	10.32	6.03
	All variations	12.72	10.45	10.18	6.06
Proposed Phase Feature	Expression	11.53	9.36	8.27	4.60
	Illumination	11.57	9.39	8.31	4.60
	Occlusion	11.64	10.18	8.64	4.62
	All variations	11.58	9.92	8.46	4.61

A comparison with PCA [3] and LDA [3] has also been performed by varying the number of training images. On decreasing the number of training images from seven to one, the accuracy of PCA algorithm is reduced by 46.95%, and the accuracy of LDA algorithm is reduced by 44.21%. Comparatively, the amplitude and phase based algorithms maintain a higher level of accuracy at 96.87% and 97.34% respectively resulting in a decrease of only 3%. To determine the recognition performance for each light spectrum, the single training image belonging to the same light spectrum is used. These results show that the proposed face recognition algorithm is able to reduce the interclass similarity and intraclass variability even with single training face image.

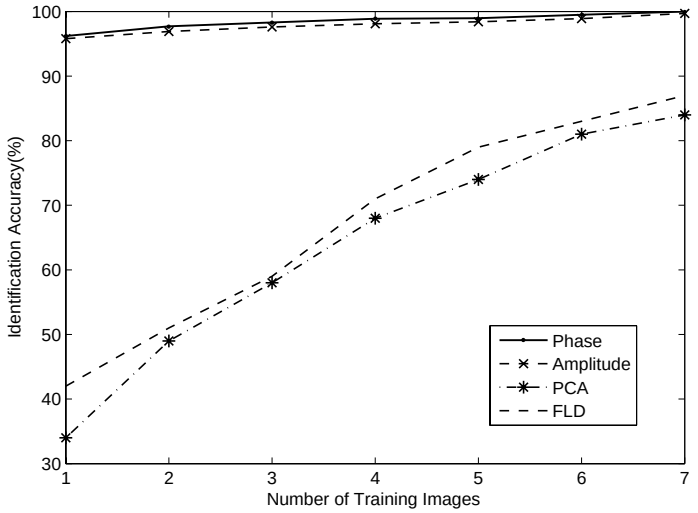

Fig. 10. Accuracy of algorithms on varying the number of training images

4.2 Performance of Image and Feature Fusion Algorithms

Previous studies on face recognition used only the visible and long-wave IR face images [8], [12], [13], [23], [24], [25]. In this research, we extend the study using our proposed approach to all possible pairs of images captured from different light spectrums.

Experiments for validating the fusion algorithms are carried out for four fusion schemes described in Sections 3.1, 3.2, 3.3 and 3.4:

Experiment 1: Feature fusion of amplitude and phase features extracted from a single spectrum face image.

Experiment 2: Feature fusion of amplitude and phase features extracted from multi spectral face images.

Experiment 3: Image fusion of two face images of different spectrum.

Experiment 4: Image and feature fusion of two face images of different spectrum.

These experiments are performed using one training image per spectrum, e.g. one short-wave IR and one visible face image are used for information (image or feature) fusion. The fused image or feature vector obtained is used for training the texture based recognition algorithm. To determine the recognition performance for combinations of light spectrum and information fusion, the single image used for training also has similar multi-spectral fused image characteristics. In experiments 2, 3 and 4, images from two different spectrum are fused. There are six possible pairs of light spectrum combinations.

1. Long-wave IR with medium-wave IR (L-M)
2. Long-wave IR with short-wave IR (L-S)
3. Long-wave IR with visible (L-V)
4. Medium-wave IR with short-wave IR (M-S)
5. Medium-wave IR with visible (M-V)
6. Short-wave IR with visible (S-V)

Table 2 summarizes the equal error rates for these experiments. For cases when the features are extracted and fused from only one image (experiment 1), short-wave IR gives the best results with EER of 4.01% followed by visible spectrum face image, with the EER of 4.20%. The performance of long-wave and medium-wave IR are close to each other with an EER of 8.26% and 7.88% respectively. Feature fusion with multi spectral face images (experiment 2) performs best for the combination of short-wave IR and visible spectrum, with an EER of 3.27%. Further, for fusion of information from two images (experiments 3 and 4), the combination of short-wave IR and visible spectrum performs best with the EER ranging from 2.86% to 3.59%.

Table 2

EER (%) for image and feature fusion algorithms using Equinox database

Algorithms	EER of Single Spectrum Images				EER of Fused Images					
	LWIR	MWIR	SWIR	Visible	L-M	L-S	L-V	M-S	M-V	S-V
Feature Fusion – Single Spectrum	8.26	7.88	4.01	4.20	--	--	--	--	--	--
Feature Fusion – Multi Spectral	--	--	--	--	3.99	3.81	3.94	3.62	3.78	3.27
Image Fusion (Amplitude Features)	--	--	--	--	6.23	5.33	5.26	3.98	4.72	3.59
Image Fusion (Phase Features)	--	--	--	--	5.95	4.01	4.91	3.83	4.68	3.51
Image and Feature Fusion	--	--	--	--	3.21	3.16	3.18	3.01	3.09	2.86

For the fusion experiments, amplitude and phase features extracted from the short-wave IR and visible spectrum gives the best results. Visible spectrum images provide detailed information of the features which leads to comparatively high intraclass variability and hence high false rejection rate; whereas the short-wave IR is associated with reflected radiation and is reflective in nature. This radiation is for the most part invisible to the human eye but is useful for recognition. Differences in appearance between the visible

and short-wave IR are due to the reflective properties. In fusion, these properties of short-wave IR are fused with the properties of visible image to obtain better performance.

4.3 Comparison of Proposed Algorithm with Existing Fusion Algorithms

Experiments are also carried out to compare the performance of the proposed image and feature fusion algorithm with other existing fusion algorithms. These include image fusion using concatenation by Chang [6], feature fusion using concatenation technique by Ross [21] (referred as Ross1), and feature and match score fusion by Ross [21] (referred as Ross2). The dataset for this comparison consists of short-wave IR and visible face images from the Equinox database. For every individual, one face image from each spectrum is used for training and rest of the images is used for testing.

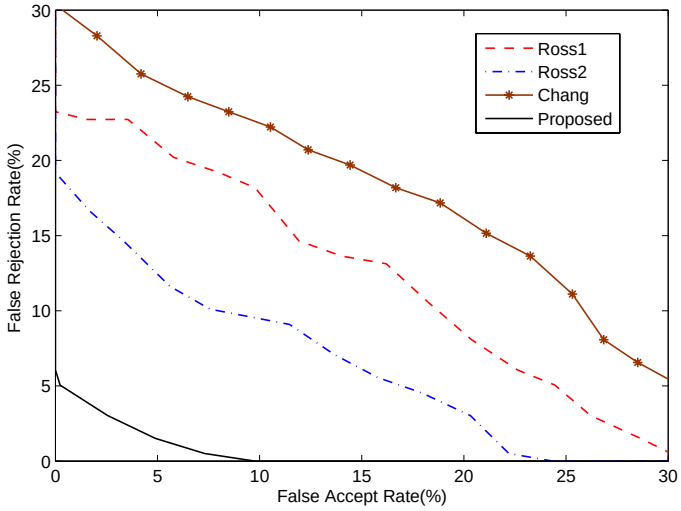

Fig. 11. ROC plot comparing the performance of proposed algorithm with fusion algorithms

During training and testing, we fuse the information from visible and SWIR face images which have similar expression, occlusion and lighting variations. Fig. 11 shows the ROC curves for the existing fusion algorithms along with the proposed image and feature fusion algorithm. Equal error rate for fusion technique by Chang is 17.63%, by Ross1 is 14.10%, by Ross2 is 9.41% and with the proposed algorithm is 2.86%.

5 Conclusion

In this paper, we proposed algorithms to fuse the information from multi spectral face images. Fusion is performed at both the image and feature level to generate a fused feature vector which is used for recognizing a face image. DWT based image fusion algorithm is proposed to fuse the information from multi spectral face images at image level. Feature fusion algorithm uses a learning algorithm based on SVM which intelligently selects and fuses good features from two feature sets. A hierarchical combination of the two fusion algorithms is then used to generate a fused feature vector. Features extracted from face image are the amplitude and phase features obtained using 2D log polar Gabor wavelet. The algorithms are evaluated using the Equinox multi spectral face database which consists of face images from thermal IR, reflective IR and visible spectrum. The experiments are performed using single training image to simulate the scenario when limited number of images are available for training. Experimental results show that the combination of short-wave IR and visible face images performs best in comparison to other combinations with an EER of 2.86% and is much better than the performance of face recognition using only visible images. Experimental results also

show that the proposed image and feature fusion algorithm performs better than the existing fusion algorithms.

Acknowledgement

The authors wish to acknowledge Equinox Corporation for providing the multi spectral face database for this research. The authors thank the reviewers for their valuable comments and suggestions for improving the quality and technical content of this paper.

References

[1] M. Alam, W. Badawy, V. Dimitrov and G. Jullien, An efficient architecture for a lifted 2D biorthogonal DWT, Journal of VLSI Signal Processing, 40 (2005) 335–342.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using the wavelet transform, IEEE Transaction on Image Processing, 1 (2) (1992) 205-220.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711-720.

[4] J. Bigun, and J. M. Du Buf, N-folded symmetries by complex moments in Gabor space and their applications to unsupervised texture segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1) (1994) 80-87.

[5] E. Bigun, J. Bigun, B. Duc, and S. Fischer, Expert conciliation for multi modal person authentication systems by bayesian statistics, in: Proceedings of International Conference on Audio- and Video-based Person Authentication (1997) 311-318.

[6] K. Chang, K. Bowyer, V. Barnabas, and S. Sarkar, Comparison and combination of ear and face images in appearance based biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1160-1165.

[7] X. Chen, and P. Flynn, and K. Bowyer, PCA-based face recognition in infrared imagery: Baseline and comparative studies, in: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (2003) 127-134.

[8] X. Chen, P. Flynn and K. Bowyer, IR and visible light face recognition, Computer Vision and Image Understanding 99 (3) (2005) 332-358.

[9] H. G. Chew, C.C. Lim, and R.E. Bogner. An implementation of training dual-nu Support Vector Machines, Optimization and Control with Applications. Qi, Teo, and Yang, (Editors) Kluwer (2004).

[10] L. Chiunhsiun, and F. Kuo-Chin, Triangle-based approach to the detection of human face, Pattern Recognition 34 (6) (2001) 1271-1284.

[11] J. Dowdall, I. Pavlidis, and G. Bebis, Face detection in the near-IR spectrum, Image and Vision Computing 21 (7) (2003) 565-578.

[12] A. Gyaourova, G. Bebis, and I. Pavlidis, Fusion of infrared and visible images for face recognition, Lecture Notes in Computer Science 3024 (2004) 456-468.

[13] J. Heo, S. Kong, B. Abidi, and M. Abidi, Fusion of visual and thermal signatures with eyeglass removal for robust face recognition, in: Proceedings of the IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR (2004) 94-99.

[14] J. Kittler, G. Matas, K. Jonsson, and M. Sanchez, Combining evidence in personal identity verification systems, Pattern Recognition Letters 18 (9) (1997) 845-852.

[15] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, Recent advances in visual and infrared face recognition - A Review, Journal of Computer Vision and Image Understanding 97 (1) (2005) 103-135.

[16] A. Kumar, D. C. M. Wong, H. Shen, and A. K. Jain, Personal verification using palmprint and hand geometry biometric, in: Proceedings of the International Conference on Audio- and Video-based Person Authentication (2003) 668-675.

[17] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, Pattern Recognition Letters 26 (2) (2005) 181-191.

[18] I. Pavlidis and P. Symosek, The imaging issue in an automatic face disguise detection system, in: Proceedings of the IEEE Workshop on Computer Vision beyond the Visible Spectrum: Methods and Applications (2000) 15-24.

[19] F. Prokoski, History, Current status, and future of infrared identification, in: Proceedings of the IEEE Workshop on Computer Vision beyond the Visible Spectrum: Methods and Applications (2000) 5-14.

[20] A. Ross and A. K. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (13) (2003) 2115-2125.

[21] A. Ross and R. Govindarajan, Feature level fusion using hand and face biometrics, in: Proceedings of the SPIE Conference on Biometric Technology for Human Identification II (2005) 196-204.

[22] R. Singh, M. Vatsa and A. Noore, Textural feature based face recognition for single training images, IEE Electronics Letters, 41 (11) (2005) 23-24.

[23] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, Infrared and visible image fusion for face recognition, in: Proceedings of the SPIE Defense and Security Symposium (Biometric Technology for Human Identification) 5404 (2004) 585-596.

[24] D. Socolinsky, and A. Selinger, A comparative analysis of face recognition performance with visible and thermal infrared imagery, in: Proceedings of the International Conference on Pattern Recognition 4 (2002) 217-222.

[25] D. Socolinsky, A. Selinger, and J. Neuheisel, Face recognition with visible and thermal infrared imagery, Journal of Computer Vision and Image Understanding 91 (2003) 72-114.

[26] S. Venkatesh and R.A.Owens, Energy feature detection scheme, in: Proceedings of the International Conference on Image Processing (1989) 553-557.

[27] J. D. Villasenor, B. Belzer, and J. Liao, Wavelet filter evaluation for image compression, IEEE Transaction on Image Processing, 4 (8) (1995) 1053-1060.

[28] J. Wilder, J. Phillips, C. Jiang, and S. Wiener, Comparison of visible and infra-red imagery for face recognition, in: Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (1996) 182-187.

[29] www.equinoxsensors.com/products/HID.html

[30] L. Zhang, and D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (3) (2006) 351 – 363.