

Infrared and Visible Image Fusion for Face Recognition

Saurabh Singh^a, Aglika Gyaourova^a, George Bebis^a, and Ioannis Pavlidis^b

^aComputer Vision Laboratory, University of Nevada, Reno

^bVisual Computing Laboratory, University of Houston

ABSTRACT

Considerable progress has been made in face recognition research over the last decade especially with the development of powerful models of face appearance (i.e., eigenfaces). Despite the variety of approaches and tools studied, however, face recognition is not accurate or robust enough to be deployed in uncontrolled environments. Recently, a number of studies have shown that infrared (IR) imagery offers a promising alternative to visible imagery due to its relative insensitivity to illumination changes. However, IR has other limitations including that it is opaque to glass. As a result, IR imagery is very sensitive to facial occlusion caused by eyeglasses. In this paper, we propose fusing IR with visible images, exploiting the relatively lower sensitivity of visible imagery to occlusions caused by eyeglasses. Two different fusion schemes have been investigated in this study: (1) image-based fusion performed in the wavelet domain and, (2) feature-based fusion performed in the eigenspace domain. In both cases, we employ Genetic Algorithms (GAs) to find an optimum strategy to perform the fusion. To evaluate and compare the proposed fusion schemes, we have performed extensive recognition experiments using the Equinox face dataset and the popular method of eigenfaces. Our results show substantial improvements in recognition performance overall, suggesting that the idea of fusing IR with visible images for face recognition deserves further consideration.

Keywords: Face Recognition, Infrared, Visible, Fusion, Principal Component Analysis, Wavelets

1. INTRODUCTION

Several factors affect face recognition performance including pose variations, facial expression changes, occlusions, and most importantly, illumination changes. Previous studies have shown that IR imagery offers a promising alternative to visible imagery for handling variations in face appearance due to illumination changes more successfully^{1,2}. As a result, face recognition in the IR spectrum has the potential to offer simpler and more robust solutions, improving recognition performance in uncontrolled environments and deliberate attempts to obscure identity.³

Despite its robustness to illumination changes, however, IR imagery has several drawbacks including that it is sensitive to temperature changes in the surrounding environment, variations in the heat patterns of the face, and it is opaque to glass. In contrast to IR imagery, visible imagery is more robust to the above factors but very sensitive to illumination changes. This suggests that effective algorithms to fuse information from both spectra have the potential to improve face recognition performance. In the past, IR and visible image fusion has been successfully used for visualization purposes,⁴ especially in the remote sensing area. Sanjeevi et al.⁵ provide a review and comparisons among existing image fusion techniques for visualization purposes.

In this paper, we concentrate on the sensitivity of IR imagery to facial occlusion due to eyeglasses. Objects made of glass act as a temperature screen, completely hiding the parts located behind them. This can affect recognition performance significantly. In fact, our experimental results illustrate that face recognition performance in the IR spectrum degrades seriously when eyeglasses are present in the probe image but not in the gallery image and vice versa. To address this serious limitation of IR, we propose fusing IR with visible imagery. Visible imagery can suffer from highlights on the glasses under certain illumination conditions, but the problems are considerably less severe than with IR. Since IR and visible imagery capture intrinsically different characteristics of the observed faces, intuitively, a better face description could be found by utilizing the complimentary information present in the two spectra.

Two different fusion schemes have been investigated in this work. The first one is image-based, operates in the wavelet domain, and yields a fused image capturing important information from both spectra. The second one is feature-based, operates in the eigenspace domain, and yields a set of important eigenfeatures from both spectra. In both cases, we employ GAs to find an optimum strategy to perform the fusion. It should be noted that both schemes are different from fusion schemes reported in the recently literature where fusion takes place at the decision level (i.e., fusing the outputs of different classifiers^{6,7}). To evaluate and compare the proposed fusion schemes, we have performed extensive recognition experiments using the Equinox face dataset and the popular eigenface approach.⁸ It should be emphasized that the eigenface approach is used only for the purpose of testing the proposed fusion schemes (i.e., any other recognition technique could have been used for the same purpose).

The rest of the paper is organized as follows: In Section 2, we review the problem of face recognition in the IR spectrum. A brief review of wavelets and eigenfaces is given in Section 3. The fusion schemes in the wavelet and eigenspace domains are described in detail in Section 4 while the use of GAs to implement them is described in Section 5. The face dataset and experimental procedure used to evaluate and compare the proposed fusion schemes are given in Section 6. Experimental results are presented and discussed in Section 7. Finally, Section 8 contains our conclusions and plans for future research.

2. REVIEW OF FACE RECOGNITION IN THE IR SPECTRUM

Prokoski⁹ presents an overview of identification in the IR spectrum. Below, we review several studies comparing the performance of visible and IR based face recognition. The effectiveness of visible versus IR was compared using several recognition algorithms by Wilder et al.⁶ Using a database of 101 subjects without glasses, varying facial expression, and allowing minor lighting changes, they concluded that there are no significant performance differences between visible and IR recognition across all the algorithms tested. They also concluded that fusing visible and IR decision metrics represents a viable approach for enhancing face recognition performance. In Selinger et al.¹⁰ and Socolinsky et al.,¹¹ several different face recognition algorithms were tested under various lighting conditions and facial expressions. Using radiometrically calibrated thermal imagery, they reported superior performance for IR-based recognition than visible-based recognition. The effect of lighting, facial expression, and passage of time between the gallery and probe images were examined by Chen et al.⁷ Although IR-based recognition outperformed visible-based recognition assuming lighting and facial expression changes, their experiments demonstrated that IR-based recognition degrades when there is substantial passage of time between the gallery and probe images. Using fusion strategies at the decision level based on ranking and scoring, they were able to develop schemes that outperformed either modality alone. IR has also been used recently in face detection by Dowdall et al.¹² This approach employs multi-band feature extraction and capitalizes on the unique reflectance characteristics of the human skin in the near-IR spectrum. Human skin exhibits an abrupt change in reflectance around 1.4 μm . This phenomenology allows for a highly accurate skin mapping by taking a weighted difference of the lower band near-IR image and the upper band near-IR image. This allowed for simple algorithmic-based face detection methods to perform extremely well.

3. BACKGROUND

3.1. Wavelets Review

Wavelets are a type of multi-resolution function approximation that allow for the hierarchical decomposition of a signal or an image.¹³ In particular, they decompose a given signal onto a family of functions with finite support. This family of functions is constructed by the translations and dilations of a single function called *mother wavelet*. The finite support of the *mother wavelet* gives exact time localization while the scaling allows extraction of different frequency components. The basic requirement of multi-resolution analysis is formulated by requiring a nesting of the spanned spaces as:

$$\cdots V_{-1} \subset V_0 \subset V_1 \cdots \subset L^2 \quad (1)$$

In space V_{j+1} , we can describe finer details than in space V_j . In order to construct a multi-resolution analysis, a scaling function ϕ is necessary, together with a dilated and translated version of it:

$$\phi_i^j(x) = 2^{\frac{j}{2}} \phi(2^j x - i). \quad i = 0, \dots, 2^j - 1. \quad (2)$$

The important features of a signal can be better described or parameterized, not by using $\phi_i^j(x)$ and increasing j to increase the size of the subspace spanned by the scaling function, but by defining a slightly different set of function $\psi_i^j(x)$ that span the difference between the spaces spanned by various scales of the scale function. These functions are the wavelets, which span the wavelet space W_j such that $V_{j+1} = V_j \oplus W_j$, and can be described as:

$$\psi_i^j(x) = 2^{\frac{j}{2}} \psi(2^j x - i). \quad i = 0, \dots, 2^j - 1. \quad (3)$$

Different scaling functions $\phi_i^j(x)$ and wavelets $\psi_i^j(x)$ determine various wavelet transforms. In this paper, we use the Haar wavelet which is the simplest to implement and computationally the least demanding. Furthermore, since Haar basis forms an orthogonal basis, the transform provides a non-redundant representation of the input images. The Haar scaling function is given by:

$$\phi(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The Haar wavelet is defined as:

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \leq x < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Wavelets capture visually plausible features of the shape and interior structure of objects. Features at different scales capture different levels of detail. Coarse scale features encode large regions while fine scale features describe smaller, local regions. All these features together disclose the structure of an object in different resolutions.

3.2. Eigenfaces Review

The eigenface approach uses Principal Components Analysis (PCA), a classical multivariate statistics method, to linearly project face images in a low-dimensional space.⁸ This space is spanned by the principal components (i.e., eigenvectors corresponding to the largest eigenvalues) of the distribution of the training images. Specifically, representing each image $I(x, y)$ as a $N \times N$ vector Γ_i , first the average face Ψ is computed:

$$\Psi = \frac{1}{R} \sum_{i=1}^R \Gamma_i \quad (6)$$

where R is the number of faces in the training set. Next, the difference Φ of each face from the average face is computed: $\Phi_i = \Gamma_i - \Psi$. Then the covariance matrix is estimated by:

$$C = \frac{1}{R} \sum_{i=1}^R \Phi_i \Phi_i^T = AA^T, \quad (7)$$

where, $A = [\Phi_1 \Phi_2 \dots \Phi_R]$. The eigenspace can then be defined by computing the eigenvectors μ_i of C . Since C is very large ($N \times N$), computing its eigenvector will be very expensive. Instead, we can compute ν_i , the eigenvectors of $A^T A$, an $R \times R$ matrix. Then μ_i can be computed from ν_i as follows⁸:

$$\mu_i = \sum_{j=1}^R \nu_{ij} \Phi_j, \quad j = 1 \dots R. \quad (8)$$

Usually, we only need to keep a smaller number of eigenvectors R_k corresponding to the largest eigenvalues. Given a new image, Γ , we subtract the mean ($\Phi = \Gamma - \Psi$) and compute the projection:

$$\tilde{\Phi} = \sum_{i=1}^{R_k} w_i \mu_i. \quad (9)$$

where $w_i = \mu_i^T \Gamma$ are the coefficients of the projection. We refer to $\{w_i\}$ as eigenfeatures.

In the eigenface approach, each image is represented as low-dimensional feature vector, containing the coefficients of the projection of the image in the eigenspace. Recognition is then performed by matching the coefficients of an unknown face image (i.e., probe) to the coefficients of a set of known images (i.e., gallery).⁸

4. FUSION SCHEMES

We describe below in detail the two fusion schemes considered in this work. We assume that each face is represented by a pair of images, one in the IR spectrum and one in the visible spectrum. Both images have been normalized prior to fusion to ensure similar ranges of values (see Section 6).

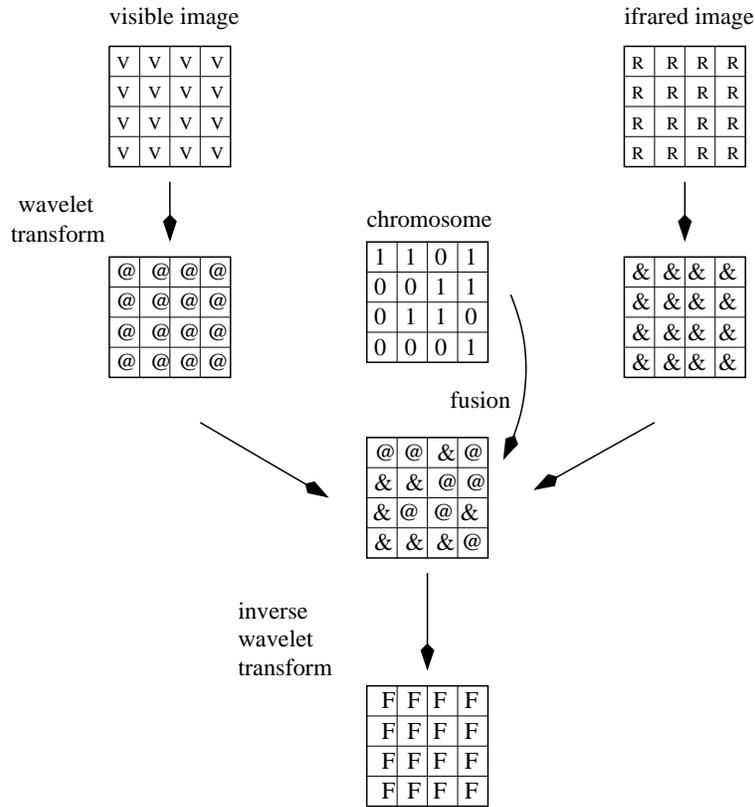


Figure 1. Fusion in the wavelet domain.

4.1. Image-based Fusion in the Wavelet Domain

The goal in this scheme is to compute a fused image from the IR and visible images, hopefully capturing the most important information from each spectrum. Since IR images have much lower resolution compared to visible images, we have considered fusing multi-resolution representations of the IR and visible images. Specifically, the slow heat transfer through the human body causes natural low resolution IR images of human face. Using pixel by pixel fusion between IR and visible images would not preserve spatial information. In contrast, fusing

multi-resolution image representations would allow features with different spatial extends to be fused at the resolution that they are most salient.

The first step in this scheme is to compute a multi-resolution representation for the IR and visible images. This is done using Haar wavelets.¹³ The result is a set of wavelet coefficients for each image. To fuse the two images, we mix the wavelet coefficients by selecting a subset of coefficients from the IR image and the rest from the visible image. The key question is how to combine the coefficients from each image. Obviously, using un-weighted averages is not appropriate since it assumes that the two spectra are equally important and, even further, that they have the same resolution. Several experiments fusing images in the wavelet domain have been reported in.¹⁴ Perhaps, the most intuitive approach is picking the coefficients with maximum absolute value.¹⁵ The higher the absolute value of a coefficient is, the higher is the probability that it encodes salient image features. Our experiments using this approach showed poor performance.

Here, we employ GAs to decide which wavelet coefficients to select from each spectrum. The fused images are then computed by applying the inverse Haar wavelet transform on the selected wavelet coefficients. Fig. 1 illustrates the main steps of this fusion scheme. Recognition is performed by applying the eigenface approach on the fused images. In this case, the eigenspace is computed from the fused images and each face is represented by projecting its fused image(s) in this space.

4.2. Feature-based Fusion in the Eigenspace Domain

This scheme considers an alternative strategy where instead of fusing the IR and visible images explicitly, we fuse the eigenfeatures computed separately from those images. Specifically, first we compute two eigenspaces, one using the IR face images and one using the visible face images. Then, each face is represented by two sets of eigenfeatures, the first computed by projecting the IR face image in the IR-eigenspace, and the second computed by projecting the visible face image in the visible-eigenspace. The goal of fusion in this case is to combine important information from each eigenspace by selecting a subset of eigenfeatures from the IR-eigenspace and the rest from the IR-eigenspace. GAs are used again to decide which eigenfeatures to select and from which eigenspace. Fig. 2 illustrates the main steps of this fusion scheme.

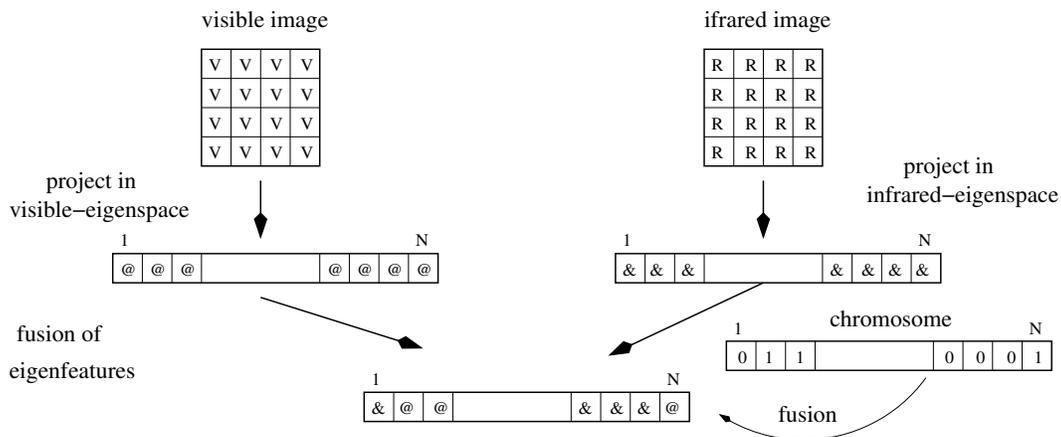


Figure 2. Fusion in the eigenspace domain.

5. EVOLUTIONARY IR AND VISIBLE FUSION

Deciding which wavelet coefficients or eigenfeatures to select from each spectrum is essentially a search problem. We use GAs to address this issue. GAs are a class of randomized, parallel search optimization procedures inspired by the mechanisms of natural selection, the process of evolution.¹⁶ They were designed to efficiently search large, non-linear, poorly-understood search spaces. In the past, GAs have been used in target recognition,¹⁷ object recognition,¹⁸ face detection/verification,^{19,20} and feature selection.^{21,22}

GAs operate iteratively on a population of structures, each of which represents a candidate solution to the problem, encoded as a string of symbols (i.e., chromosome). A randomly generated set of such strings forms the initial population from which the GA starts its search. Three basic genetic operators guide this search: selection, crossover and mutation. Evaluation of each string is based on a fitness function which is problem-dependent. The fitness function determines which of the candidate solutions are better. Selection probabilistically filters out poor solutions and keeps high performance solutions for further investigation. Mutation is a very low probability operator that plays the role of restoring lost genetic material. Crossover in contrast is applied with high probability. It is a randomized yet structured operator that allows information exchange between the strings.

Our decision to use GAs for fusion was based on several reasons. First, the search spaces for the fusion tasks under consideration are very large. In the past, GAs have demonstrated good performance when searching large spaces. Much work in the genetic and evolutionary computing communities has led to growing understanding of why they work well and plenty of empirical evidence to support this claim.^{23,24} Second, the problem of fusion appears to have many suboptimal solutions. Although GAs do not guarantee to find a global optimum solution, they have the ability to search through very large search spaces and come to nearly optimal solutions fast. Their ability for fast convergence is explained by the *schema theorem* (i.e., short-length bit patterns in the chromosomes with above average fitness, get exponentially growing number of trials in subsequent generations¹⁶). Third, they suitable for parallelization and linear speedups are the norm, not the exception.²⁵ Finally, we have applied GAs in related problems in the past with good success.^{21,22}

Below, we describe in more detail the encoding schemes, fitness evaluation functions, and genetic operators used for fusing IR with visible information in the wavelet and eigenspace domains.

Encoding: In the case of fusion in the wavelet domain, the chromosome is a bit string whose length is determined by the number of wavelet coefficients in the image decomposition. Each bit in the chromosome is associated with a wavelet coefficient at a specific location. The value of a bit in the chromosome determines whether the corresponding wavelet coefficient is selected from the IR (e.g., 0) or from the visible spectrum (e.g., 1) (see Fig. 1). In the case of fusion in the eigenspace domain, the chromosome is also a bit string whose length is determined by the number of eigenvectors. Here, we use the first 100 eigenvectors from each space (see Section 6), thus, the chromosome has length 100. Each bit in the chromosome is associated with an eigenfeature at a specific location. The value of a bit in the chromosome determines whether a particular eigenfeature is selected from the visible image (i.e., 1) or the IR image (i.e., 0) (see Fig. 2).

Fitness Evaluation: Each individual in a generation represents a possible way to fuse IR with visible information. To evaluate its effectiveness, we perform the fusion based on the information encoded by this individual and perform recognition using the eigenface approach. Recognition accuracy is computed using a validation dataset (see Section 6) and is used to provide a measure of fitness. Upon convergence, the best chromosome found is kept and used to evaluate performance on a test set.

Initial Population: In general, the initial population is generated randomly, (e.g., each bit in an individual is set by flipping a coin). In this way, however, we will end up with a population where each individual contains the same number of 1's and 0's on average. To explore subsets of different numbers of wavelet coefficients or eigenfeatures chosen from each domain, the number of 1's for each individual is generated randomly. Then, the 1's are randomly scattered in the chromosome. We used population sizes between 100 and 200 and 100 generations.

Selection: Our selection strategy was cross generational. Assuming a population of size N , the offspring double the size of the population and we select the best N individuals from the combined parent-offspring population.²⁶

Crossover: In general, we do not know how different wavelet coefficients depend on each other. If dependent coefficients are far apart in the chromosome, it is more probable that traditional 1-point crossover, will destroy the schemata. To avoid this problem, uniform crossover is used here. The crossover probability used in our experiments was 0.96.

Mutation: Mutation is a very low probability operator which flips the values of randomly chosen bit. The mutation probability used here was 0.02.

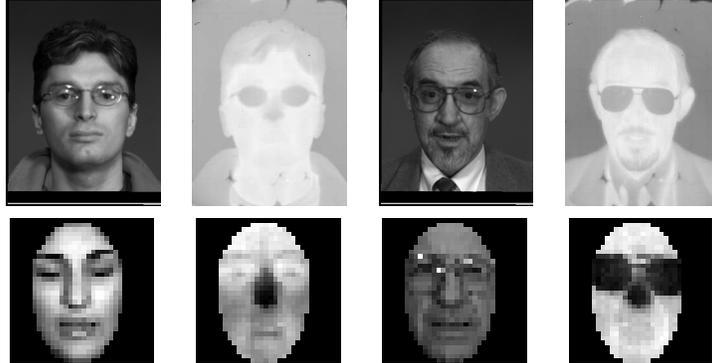


Figure 3. Examples of visible and IR image pairs (first row) and preprocessed images (second row).

6. FACE DATASET AND EXPERIMENTAL PROCEDURE

In our experiments, we used the face database collected by Equinox Corporation under DARPA’s HumanID program.²⁷ Specifically, we used the long-wave infrared (LWIR) (i.e., $8\mu\text{-}12\mu$) and the corresponding visible spectrum images from this database. The data was collected during a two-day period. Each pair of LWIR and visible light images was taken simultaneously and co-registered with 1/3 pixel accuracy (see Fig. 1). The LWIR images were radiometrically calibrated and stored as grayscale images with 12 bits per pixels. The visible images are also grayscale images represented with 8 bits per pixel. The size of the images in the database is 320×240 pixels.

The database contains frontal faces under the following scenarios: (1) three different light direction - frontal and lateral (right and left); (2) three facial expression - ”frown”, ”surprise” and ”smile”; (3) vocals pronunciation expressions - subjects were asked to pronounce several vocals from which three representative frames were chosen; and (4) presence of glasses - for subjects wearing glasses, all of the above scenarios were repeated with and without glasses. Both IR and visible face images were preprocessed prior to experimentation by following a procedure similar to that described in.^{10,11} The goal of preprocessing was to align and scale the faces, remove background, and account for some illumination variations (see Fig. 1). For comparison purposes, we have attempted to evaluate our fusion schemes using a similar experimental protocol to that given in.^{10,11} Our evaluation methodology employs a training set (i.e., used to compute the eigenfaces), a gallery set (i.e., set of persons enrolled in the system), a validation set (i.e., used in the fitness evaluation of the GA), and a test set (i.e., probe image set containing the images to be identified).

For training, we used 200 images, randomly chosen from the entire Equinox database. For recognition, we used the Euclidean distance and the first 100 principal components as in Selinger et al.¹⁰ and Socolinsky et al.¹¹ Recognition performance was measured by finding the percentage of the images in the test set, for which the top match is an image of the same person from the gallery. To mitigate for the relatively small number of images in the database, the average error was recorded using a three-fold cross-validation procedure. In particular, we split each dataset used for testing randomly three times by keeping only 75% of the images for testing purposes and the rest 25% for validation purposes. To account for performance variations due to random GA initialization, we averaged the results over three different GA runs for each test, choosing a different random seed each time. Thus, we performed a total of 9 runs for each gallery/test set experiment.

Following the terminology of Selinger et al.¹⁰ and Socolinsky et al.,¹¹ we created the following test sets: EA (expression frames, all illuminations), EL (expression frames, lateral illuminations), and EF (expression frames, frontal illumination). The inclusion relation among these sets is as follows: $EA = EL \cup EF$. Measuring the effect of eyeglasses was done by using the EA test set. There are 90 subjects with a total of 1266 pairs of images in the EA dataset. From them, 43 subjects wear glasses making a total of 822 images. For experimentation, we created the following test sets: EG (expression frames with glasses, all illuminations), EnG (expression frames without glasses, all illuminations), EFG (expression frames with glasses, frontal illumination), ELG (expression

Table 1. Averages and standard deviations for the eyeglasses experiments using fusion in the wavelet domain. The columns represent the gallery set and the rows represent the test set. The first entry in each cell shows the average performance and standard deviation from the visible images, the second entry is from the IR images, and the third entry is from the fused images. The bottom entry shows the minimum and maximum recognition performances from the three cross-validation runs achieved when using the fused images. Test scenarios for which the test and the gallery sets had common subsets were not performed.

	EG	ELG	EFG	EnG	ELnG	EFnG
EG	×	×	×	(84.8, 1.4) (15.1, 1.0) (92.5, 1.3) 91.0 – 93.6	(84.8, 1.4) (13.1, 1.0) (88.9, 1.4) 88.0 – 90.5	(64.3, 1.7) (21.7, 1.0) (82.1, 3.1) 79.2 – 85.4
ELG	×	×	(71.4, 1.0) (99.6, 0.3) (93.2, 3.0) 90.2 – 94.2	(85.8, 0.7) (16.2, 0.3) (92.3, 2.0) 90.2 – 94.2	(85.8, 0.7) (14.2, 0.4) (92.7, 0.6) 92.0 – 93.0	(56.0, 1.2) (22.4, 0.4) (83.9, 1.3) 82.5 – 84.8
EFG	×	(78.3, 1.1) (100, 0) (97.9, 0.6) 97.4 – 98.5	×	(83.7, 3.9) (14.5, 0.6) (91.7, 1.3) 90.9 – 93.1	(50.7, 1.7) (13.0, 0) (77.1, 3.1) 74.3 – 80.4	(82.6, 4.0) (22.1, 0.6) (92.2, 2.0) 90.2 – 94.2
EnG	(79.4, 1.1) (2.6, 0.2) (84.9, 0.8) 84.3 – 85.8	(72.0, 2.2) (2.4, 0.2) (81.6, 0.3) 81.2 – 81.8	(60.2, 2.0) (17.4, 0.5) (98.0, 0.1) 97.9 – 98.1	×	×	×
ELnG	(82.0, 1.7) (2.7, 0.5) (84.2, 1.9) 82.1 – 85.7	(82.0, 1.7) (2.5, 0.3) (84.1, 1.4) 82.5 – 85.2	(52.6, 0.8) (17.7, 0.6) (96.7, 1.4) 95.1 – 97.8	×	×	(73.0, 1.9) (98.4, 0.6) (96.9, 0.6) 96.2 – 97.5
EFnG	(78.6, 1.6) (2.1, 0) (85.6, 3.2) 83.2 – 89.2	(56.8, 1.1) (2.1, 0) (80.7, 2.9) 78.3 – 83.9	(78.6, 1.6) (18.6, 1.2) (87.0, 2.1) 85.2 – 89.3	×	(71.9, 1.2) (100, 0) (97.8, 1.0) 97.2 – 99.0	×
number of images	398	260	138	424	281	143

frames with glasses, lateral illumination), EFnG (expression frames without glasses, frontal illumination), ELnG (expression frames without glasses, lateral illumination). The inclusion relations among these sets are as follows: $EG = ELG \cup EFG$, $EnG = ELnG \cup EFnG$ and $EG \cap EnG = \emptyset$.

7. EXPERIMENTAL RESULTS

Our experimental results illustrate clearly that IR is robust to illumination changes but performs poorly when glasses are present in the gallery set but not in the test set and vice versa. Considerable improvements in recognition performance were achieved in this case by fusing IR with visible images both in the wavelet (see Table 1) and eigenspace (see Table 2) domains. The improvements were even greater when, in addition to eyeglasses, the test and the gallery set contained images taken under different illuminations. For example, in the EFG/ELnG test case using fusion in the wavelet domain, recognition performance was improved by 46% compared to recognition using visible images and by 82% compared to recognition using IR images (see Table 1). Between the two fusion schemes tested, fusion in the wavelet domain yielded higher recognition performance overall.

Let us take a closer look of the results shown in Tables 1 and 2. The horizontal and vertical double lines through the center of each table divide it into four quadrants (i.e., *I* to *IV*, starting from the upper-right corner and moving counterclockwise). Each quadrant represents experiments testing some specific difference between the gallery and the test sets: (1) experiments in quadrant *I* evaluate the effect of eyeglasses being present in the probe but not in the gallery; (2) experiments in quadrant *III* evaluate the effect of eyeglasses being present in the gallery but not in the probe; (3) experiments along the off-diagonals within each of these two quadrants represent tests where the illumination conditions between the gallery and probe sets are the same; (4) experiments in quadrants *II* and *IV* evaluate the effect illumination changes only.

By considering the above four cases, several interesting conclusions can be made. As expected, face recognition success based on IR images is not influenced by lighting conditions. However, IR yielded very low success when eyeglasses were present in the gallery but not in the probe and vice-versa (cases (1) and (2)). The success of visible-based face recognition was relatively insensitive to subjects' wearing glasses (cases (1) and (2)). Lighting conditions had big influence on the success of face recognition in the visible domain (case (3)). The success

Table 2. Averages and standard deviations for the eyeglasses experiments with fusion in the eigenspace domain. Our notation is similar to that used in Table 2

	EG	ELG	EFG	EnG	ELnG	EFnG
EG	×	×	×	(84.8, 1.4) (15.1, 1.0) (94.4, 1.37) 92.9 – 95.5	(84.8, 1.4) (13.1, 1.0) (88.4, 0.8) 87.6 – 89.2	(64.3, 1.7) (21.7, 1.0) (70.3, 2.2) 68.6 – 72.7
ELG	×	×	(71.4, 1.0) (99.6, 0.3) (97.3, 0.6) 96.7 – 97.8	(85.8, 0.7) (16.2, 0.3) (91.1, 1.3) 89.8 – 92.4	(85.8, 0.7) (14.2, 0.4) (92.9, 0.8) 92.0 – 93.4	(56.0, 1.2) (22.4, 0.4) (61.4, 0.0) 61.4 – 61.4
EFG	×	(78.3, 1.1) (100, 0) (100.0, 0.0) 100.0 – 100.0	×	(83.7, 3.9) (14.5, 0.6) (93.1, 2.1) 90.9 – 95.0	(50.7, 1.7) (13.0, 0) (73.1, 1.4) 72.1 – 74.6	(82.6, 4.0) (22.1, 0.6) (90.4, 2.4) 87.7 – 92.0
EnG	(79.4, 1.1) (2.6, 0.2) (85.8, 0.7) 85.3 – 86.7	(72.0, 2.2) (2.4, 0.2) (77.0, 0.2) 76.9 – 77.2	(60.2, 2.0) (17.4, 0.5) (67.8, 0.0) 67.8 – 67.8	×	×	×
ELnG	(82.0, 1.7) (2.7, 0.5) (84.1, 3.1) 80.5 – 86.1	(82.0, 1.7) (2.5, 0.3) (84.6, 1.8) 82.6 – 86.3	(52.6, 0.8) (17.7, 0.6) (58.0, 1.0) 57.0 – 59.1	×	×	(73.0, 1.9) (98.4, 0.6) (98.0, 0.3) 97.8 – 98.4
EFnG	(78.6, 1.6) (2.1, 0) (85.5, 3.0) 82.1 – 88.0	(56.8, 1.1) (2.1, 0) (69.0, 3.4) 65.2 – 71.5	(78.6, 1.6) (18.6, 1.2) (87.0, 1.8) 85.3 – 88.8	×	(71.9, 1.2) (100, 0) (98.2, 1.4) 96.8 – 99.6	×
number of images	398	260	138	424	281	143

of face recognition based on fused images was similar in all four quadrants. This implies that we were able to achieve relative insensitivity to both eyeglasses and variable illumination.

Fusion in the wavelet domain led to higher recognition performance compared to recognition in the visible spectrum, however, this was not always the case with recognition performed using IR images, especially when recognition performance using IR was higher than 97%. The main reason that fusion was not always better than IR is because illumination effects present in the visible images were not completely disregarded from the fused images. We have noticed that in all the cases where IR performed better than fusion, the illumination direction in the gallery set was different from that in the test set, assuming that glasses were both present or absent from the gallery and test sets (e.g., ELG/EFG, EFG/ELG, ELnG/EFnG, and EFnG/ELnG). This can be confirmed by observing the reconstructed fused images shown in Fig. 4, as well as their first eigenfaces shown in Fig. 5. The fused images have higher resolution compared to IR images, however, they are also affected by illumination present in the visible images. The first few eigenfaces of the fused images show that they still encode illumination direction.

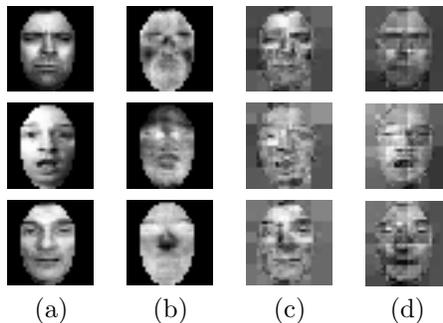


Figure 4. Fusion in the wavelet domain. (a) Visible and (b) IR images, (c) fused images in the case of lateral illumination without glasses, (d) fused images in the case of lateral illumination with glasses.

Fusion in the eigenspace domain always led to higher recognition performance compared to both recognition in the visible and IR domains with only one exception (i.e., EFG/ELG test case). This exception could be explained again by the difference in illumination directions between the gallery and test sets. For comparison



Figure 5. Fusion in the wavelet domain – the first few eigenfaces of several fused images. The second and third eigenfaces show clear influence of right and left lateral illumination.

purposes, we reconstructed several visible and IR images from the ELnG set using only the subset of eigenvectors selected from each spectrum (see Fig. 6). The reconstructed visible images reveal clearly that they contain information about illumination direction.

Although fusion in the eigenspace domain improved recognition performance more consistently compared to fusion in the wavelet domain, the overall recognition performance using wavelet-based fusion is lower. Significant differences in recognition performance can be noticed in the following cases: EFG/EnG (recognition rate is 30% higher using wavelet-based fusion), EFG/ELnG (38% higher), EFnG/ELG (23% higher), EFnG/EG (12% higher), and ELG/EFnG (11% higher). We attribute the higher recognition accuracy using fusion in the wavelet domain to the more powerful eigenfeatures computed using the fused images.

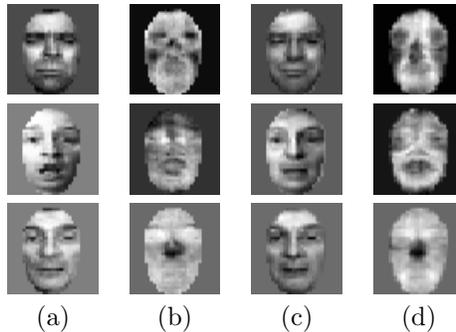


Figure 6. Fusion in the eigenspace domain. (a) Visible and (b) IR images assuming lateral illumination; (c) reconstructed visible images and (d) reconstructed IR images using the eigenvectors selected from each domain only

8. CONCLUSIONS AND FUTURE WORK

We presented and compared two different fusion schemes for combining IR and visible light images for the purposes of face recognition. The first scheme is image-based, operates in the wavelet domain, and yields a fused image capturing important information from both spectra. The second scheme is feature-based, operates in the eigenspace domain, and yields a vector containing important eigenfeatures from both spectra. Both schemes aim at improved and robust recognition performance across variable lighting and presence/absence of eyeglasses. Between the two schemes, fusion in the wavelet domain demonstrated significant improvements in recognition performance overall.

Future work includes considering more effective fusion schemes (e.g., weighted averages of wavelet coefficients or eigenfeatures) and more powerful fitness functions (i.e., add extra terms to control the number of coefficients selected from different bands of each spectrum or different ranges of eigenvectors). These might help to overcome some of the problems mentioned in the previous section (e.g., illumination effects carried over to the fused images) and improve fusion overall.

Further consideration should also be given to the existence of many optimal solutions found by the GA. Although optimal in the training phase, these solutions showed different recognition performances when used for testing. In investigating these solutions, we were not able to distinguish any pattern in the content of the chromosomes that might have revealed why some chromosomes were better than others. On the average, half of the wavelet coefficients or eigenfeatures were selected from the visible spectrum and the other half from the

IR spectrum. The use of larger validation sets and more selective fitness functions might help to address these issues more effectively.

Additional issues for future research include considering fitness approximation schemes²⁸ to reduce the computational requirements of fitness evaluation, and investigating the effect of environmental (e.g., temperature changes), physical (e.g., lack of sleep) and physiological conditions (e.g., fear, stress) to IR performance.

REFERENCES

1. Wolff L., Socolinsky D. and and Eveland C., “Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery,” in *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, (Hawaii), 2001.
2. Jain, A. and Bolle, R. and Pankanti, S., *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, 1999.
3. Pavlidis, I. and Symosek, P., “The imaging issue in an automatic face/disguise detection system,” in *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp. 15–24, 2000.
4. Scheunders, P., “Local mapping for multispectral image visualization,” *Image and Vision Computing* **19**(13), pp. 971–978, 2001.
5. Sanjeevi, S., Vani, K., Lakshmi, K., “Comparison of conventional and wavelet transform techniques for fusion of {IRS-1C} {KISS-III} and {PAN} images,” in *Proc. ACRS*, **1**, pp. 140–145, 2001.
6. Wilder, J. and Phillips, J. and Jiang, C. and Wiener, S., “Comparison of visible and infra-red imagery for face recognition,” in *2nd International Conference on Automatic Face and Gesture Recognition*, pp. 182–187, (Killington), 1996.
7. Chen, X. and Flynn, P. and Bowyer, K., “Pca-based face recognition in infrared imagery: Baseline and comparative studies,” in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, (Nice, France), 2003.
8. Turk, M. and Pentland, A.P., “Eigenfaces for recognition,” *CogNeuro* **3**(1), pp. 71–96, 1991.
9. Prokoski, F., “History, current status, and future of infrared identification,” in *IEEE Workshop on Computer Vision Beyond the Visible Spectrum*, (Hilton Head), 2000.
10. Selinger, A. and Socolinsky, D., “Appearance-based facial recognition using visible and thermal imagery: A comparative study,” tech. rep., Equinox Corporation n.02-01, 2002.
11. Socolinsky, D. and Selinger A., “Comparative study of face recognition performance with visible and thermal infrared imagery,” in *International Conference on Pattern Recognition*, pp. 217–222, 2002.
12. Dowdall, J. and Pavlidis, I. and Bebis, G., “Face detection in the near-ir spectrum,” *Image and Vision Computing* **21**(7), pp. 565–578, 2001.
13. Chui, C.K., *An introduction to wavelets*, Academic Press, 1992.
14. Chipman, L.J. and Orr, T.M., “Wavelets and image fusion,” in *IEEE International Conference on Image Processing*, **3**, pp. 248–251, 1995.
15. Li, H. and Manjunath, B.S. and Mitra, S.K., “Multisensor image fusion using the wavelet transform,” in *IEEE International Conference on Image Processing*, **1**, pp. 51–55, (Austin, Texas), 1994.
16. Goldberg, D.E., *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, 1989.
17. Katz, A. and Thrift, P., “Generating image filters for target recognition by genetic learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(9), 1994.
18. Bebis, G. and Louis, S. and Varol, Y. and Yfantis, A., “Genetic object recognition using combinations of views,” *IEEE Transactions on Evolutionary Computing* **6**(2), pp. 132–146, 2002.
19. Swets, D. and Punch, B., “Genetic algorithms for object localization in a complex scene,” in *IEEE International Conference on Image Processing*, pp. 595–598, 1995.
20. Bebis, G. and Uthiram, S. and Georgiopoulos, M., “Face detection and verification using genetic search,” *International Journal of Artificial Intelligence Tools* **9**(2), pp. 225–246, 2000.
21. Sun, Z. and Yuan, X. and Bebis, G. and Louis, S., “Genetic feature subset selection for gender classification: A comparison study,” in *IEEE Workshop on Applications of Computer Vision*, 2002.

22. Sun, Z. and Bebis, G. and Miller, R., "Boosting object detection using feature selection," in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2003.
23. Fogel, D., *Evolutionary Computation, Toward a New Philosophy of Machine Intelligence*, IEEE Press, 1992.
24. Koza, J., *Genetic Programming*, MIT Press, 1993.
25. Cantu-Paz, E., *Efficient and Accurate Parallel Genetic Algorithms and Evolutionary Computation*, Kluwer Academic Publishers, 2000.
26. Eshelman, I., "The chc adaptive search algorithm: How to have safe search when engaging in non-traditional genetic recombination," in *The Foundation of Genetic Algorithms Workshop*, pp. 265–283, 1989.
27. <http://www.equinoxsensors.com/products/HID.html>.
28. Jin, Y. and Sendhoff, B., "Fitness approximation in evolutionary computation - a survey," in *Genetic and Evolutionary Computation Conference*, 2002.