

Face recognition by fusing thermal infrared and visible imagery

George Bebis^{a,*}, Aglika Gyaourova^a, Saurabh Singh^a, Ioannis Pavlidis^b

^a Computer Vision Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, NV 89557, USA

^b Visual Computing Laboratory, University of Houston, Houston, TX, USA

Received 5 October 2005; received in revised form 20 January 2006; accepted 23 January 2006

Abstract

Thermal infrared (IR) imagery offers a promising alternative to visible imagery for face recognition due to its relative insensitivity to variations in face appearance caused by illumination changes. Despite its advantages, however, thermal IR has several limitations including that it is opaque to glass. The focus of this study is on the sensitivity of thermal IR imagery to facial occlusions caused by eyeglasses. Specifically, our experimental results illustrate that 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. 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 study. The first one is *pixel-based* and operates in the wavelet domain, while the second one is *feature-based* and operates in the eigenspace domain. In both cases, we employ a simple and general framework based on Genetic Algorithms (GAs) to find an optimum fusion strategy. We have evaluated our approaches through extensive experiments using the Equinox face database and the eigenface recognition methodology. Our results illustrate significant performance improvements in recognition, suggesting that IR and visible fusion is a viable approach that deserves further consideration.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Face recognition; Infrared; Visible; Fusion; Principal component analysis; Wavelets

1. Introduction

Face recognition technology has a wide range of potential applications related to security and safety including surveillance, information security, access control, and identity fraud. Considerable progress has been made in face recognition over the last decade especially with the development of powerful models of face appearance (i.e. eigenfaces) [1]. Despite the variety of approaches and tools studied, however, face recognition is not accurate or robust enough to be deployed in uncontrolled environments. Several factors affect face recognition performance including pose variations, facial expression changes, occlusions, and most importantly, illumination changes.

Recently, a number of studies have demonstrated that thermal IR offers a promising alternative to visible imagery for handling variations in face appearance due to illumination

changes [2], facial expression [3,4], and face pose [4] more successfully. In particular, thermal IR imagery is nearly invariant to changes in ambient illumination [2,3], and provides a capability for identification under all lighting conditions including total darkness [5]. Thus, while visible-based algorithms opt for pure algorithmic solutions into inherent phenomenology problems, IR-based algorithms have the potential to offer simpler and more robust solutions, improving performance in uncontrolled environments and deliberate attempts to obscure identity [6].

Despite its robustness to illumination changes, IR has several drawbacks. First, it is sensitive to temperature changes in the surrounding environment. Currents of cold or warm air could influence the performance of systems using IR imagery. As a result, IR images should be captured in a controlled environment. Second, it is sensitive to variations in the heat patterns of the face. Factors that could contribute to these variations include facial expressions (e.g. open mouth), physical conditions (e.g. lack of sleep), and psychological conditions (e.g. fear, stress, excitement). Finally, IR is opaque to glass. As a result, a large part of the face might be occluded (e.g. by wearing eyeglasses). In contrast to IR, visible imagery is more robust to the above factors but more sensitive to illumination changes.

* Corresponding author.

E-mail addresses: bebis@cse.unr.edu (G. Bebis), aglika@cse.unr.edu (A. Gyaourova), saurabh@cse.unr.edu (S. Singh), pavlidis@cs.uh.edu (I. Pavlidis).

The focus of this paper is on the sensitivity of thermal IR to facial occlusions caused by eyeglasses. Objects made of glass act as a temperature screen, completely hiding the parts located behind them. This may affect recognition performance seriously. Our experimental results confirm this claim, showing 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. Although visible imagery can suffer from highlights on the glasses under certain illumination conditions, the problems are considerably less severe than with IR. Therefore, effective algorithms to fuse information from both spectra have the potential to improve face recognition performance. It is worth mentioning that, there exist several approaches for removing eyeglasses in face images [7], however, it is not clear whether these methods would be useful in the IR domain since the eyes are occluded completely due to the fact that eyeglasses block thermal energy (i.e. see Fig. 5).

In this study, we have investigated two different fusion schemes, the first performing *pixel-based fusion* in the wavelet domain, and the second performing *feature-based fusion* in the eigenspace domain. The key objective is computing a fused image from the IR and visible images, capturing the most salient features from each spectrum. In both cases, we employ GAs to find an optimum fusion strategy to combine information from the two spectra. The proposed fusion schemes are different from recent, decision-based, fusion schemes (i.e. fusing the outputs of different classifiers) proposed in the literature [3,8–10]. They are also different from the fusion scheme reported in [10] where eyeglass regions are replaced by an eye-template fusion is performed in the image domain by weighted pixel averaging.

We have evaluated the proposed fusion schemes through extensive recognition experiments using the Equinox face data set [11] and the eigenface recognition methodology [12]. It should be noted that, we have chosen the eigenface approach only for the purpose of demonstrating the benefits of fusion (i.e. any other recognition technique could have been used for the same purpose). Our experimental results illustrate that fusion offers substantial improvements in recognition performance. Moreover, the evolutionary-based fusion methodology provides a simple and general framework that can be used to improve recognition performance in several cases where IR information becomes less reliable due to various environmental (e.g. temperature changes), physical (e.g. lack of sleep) and physiological conditions (e.g. fear, stress). An earlier version of this work has appeared in [13].

The rest of this paper is organized as follows: In Section 2, we review the problem of face recognition in the IR spectrum. Section 3 reviews fusion of visible with IR imagery in general. The proposed fusion schemes are detailed in Sections 4 and 5 correspondingly. The use of GAs to find optimum fusion strategies is described in Section 6. The face data set and experimental procedure used to evaluate and compare the proposed fusion schemes are given in Sections 7 and 8. Our experimental results and discussion of results are presented in

Sections 9 and 10. Finally, Section 11 contains our conclusions and directions for future research.

2. Review of face recognition in the IR spectrum

Face recognition in the IR spectrum has received relatively little attention compared to visible spectrum, mostly because of the high cost of IR sensors and lack of IR data sets. A number of recent studies have shown that face recognition in the IR offers many benefits. Prokoski [14] presents an overview of identification in the IR spectrum while results in an operational scenario are reported in [15]. A recent review on face recognition in the visible and IR domains can be found in [16].

The effectiveness of visible versus IR spectrum was compared using several recognition algorithms by Wilder et al. [8]. Using a database of 101 subjects without eyeglasses, 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 [3,11,17], several appearance-based face recognition methodologies 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. Additional performance improvements were achieved using decision-based fusion [3]. In a later study [18], the impact to recognition of eye detection in the IR spectrum was investigated.

In [19], a statistical hypothesis pruning methodology was introduced for face recognition in IR. First, each IR face image was decomposed using Gabor filters. Then, each image was represented by a few parameters by modelling the marginal density of the Gabor filter coefficients using Bessel functions. Recognition was performed in the space of parameters of the Bessel functions. This approach has been improved by Buddharaju et al. [20].

The effect of lighting, facial expression, and passage of time between the gallery and probe images were examined by Chen et al. [9] 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 at the decision level based on ranking and scoring, they were able to develop schemes that outperformed either modality alone. In a related study, IR-based recognition was shown to be less sensitive to changes in 3D head pose and facial expression [4]. The potential of using near-IR hyperspectral imaging for face recognition was investigated in [21]. The results of this study showed that the local spectral properties of human tissue are nearly invariant to face orientation and expression, improving recognition performance.

Heo et al. [10] considered two types of visible and IR fusion, the first at the data level and the second at the decision level. Data fusion was implemented by applying pixel-based

weighted averaging of registered visual and thermal images. Decision fusion was implemented by combining the matching scores of individual recognition modules. To deal with occlusions caused by eyeglasses, they used ellipse fitting to detect the eyeglass regions in the IR image and replaced them with an eye template. Using a commercial face recognition system (i.e. FaceIt), they demonstrated improvements in recognition accuracy. There are two main issues with this approach. First, choosing an optimum set of weights, both for data-based and decision-based fusion, was not addressed satisfactorily. Second, replacing eyeglass regions with an eye template might not be very appropriate in the context of face recognition since the eyes contain important information for identification purposes. In fact, several studies have demonstrated that face recognition is possible using eyes alone [22].

Recently, IR was also used for face detection by Dowdall et al. [23–25] and face tracking by Eveland et al. [26]. In [23–25], multi-band feature extraction was used by capitalizing on the unique reflectance characteristics of the human skin in the near-IR spectrum. In particular, it was found that human skin exhibits an abrupt change in reflectance around 1.4 μm . Exploiting this phenomenology, a highly accurate skin mapping was built 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. In [26], a method for modelling thermal emission from human skin was used to segment faces from IR image. Tracking was performed by combining the segmentation model with the condensation algorithm.

3. Fusion of IR and visible imagery

Information obtained from a given source or sensor generally contains redundancy, which does not contribute to the knowledge about the scene we want to analyze. On the other hand, the information obtained by different means and tools, can provide varying and complementary details of the object under analysis. Finding the complimentary parts and removing the redundancy is the ultimate goal of information fusion approaches.

In the past, IR and visible image fusion has been successfully used for visualization purposes [27], especially in the remote sensing area. In this case, aerial imaging sensors capture information in several spectral ranges. Developing algorithms to visualize this information is thus very important. Sanjeevi et al. [28] provide a review and comparisons among existing image fusion techniques for visualization purposes. Fusing data from IR and visible sensors has also led to improved target recognition rates [29].

Choosing an appropriate fusion scheme is both application and data dependent. In the context of our application, infrared cameras measure the heat energy emitted by the face while optical cameras capture the light reflectance of the face surface. As the surface of the face and its temperature do not have anything in common, the information in the IR and visible images is independent and complimentary. In general, pixel by

pixel fusion does not preserve the spatial information in the image. In contrast, fusion at multiple-resolution levels allows features with different spatial extent to be fused at the resolution at which they are most salient. In this case, important features appearing at lower resolutions can be preserved in the fusion process.

The basic idea behind multi-resolution fusion is performing a multiscale transform on each source image then constructing a composite multiscale representation from these according to some specific fusion rules. The fused image is obtained by taking an inverse multiscale transform [30]. Multiscale face representations have been used in several systems [31]. Some of the most popular multiscale techniques include the Laplacian pyramid [32], Fourier transform, and wavelets [33]. High frequencies are relatively independent of global changes in the illumination, while the low frequencies take into account the spatial relationships among the pixels and are less sensitive to noise and small changes (e.g. facial expression).

Next, we describe the fusion schemes studied in this work. The slow heat transfer through the human body causes natural low resolution IR images of human face. Therefore, our first fusion strategy operates in the wavelet domain, taking into consideration the differences in the resolution of the IR and visible-light images and exploiting the benefits of multi-resolution representations. Our second fusion strategy operates in the eigenspace domain and fuses together global image features. In each case, we assume that faces are represented by a pair of images, one in the IR spectrum and one in the visible spectrum. Also, it is assumed that both images have been normalized prior to fusion to ensure similar ranges of values (Section 7).

4. Pixel-based fusion in the wavelet domain

Wavelets are a type of multi-resolution function approximation that allow for the hierarchical decomposition of a signal or an image [33]. 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:

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

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:

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

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:

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

Different scaling functions $\phi_i^j(x)$ and wavelets $\psi_i^j(x)$ determine various wavelet transforms. In this study, we have employed 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.

4.1. Methodology

Fusion in the wavelet domain involves combining the wavelet coefficients of the visible and IR images. First, we compute a multi-resolution representation for each of the images using Haar wavelets [33]. To fuse the visible and IR images, we select a subset of coefficients from the IR image and the rest from the visible image. The fused image is obtained by applying the inverse wavelet transform on the selected coefficients. The key question in implementing this idea is which wavelet coefficients to choose from each spectrum.

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 image fusion experiments in the wavelet domain have been reported in [34]. The most intuitive fusion approach in the wavelet domain is choosing the coefficients with maximum absolute value [35]. 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.

In this study, we employ GAs to decide which wavelet coefficients to select from each spectrum. Fig. 1 illustrates the main idea of this approach, which is further explained in

Section 6. Recognition is performed using the eigenface approach on the fused images.

5. Feature-based fusion in the eigenspace domain

The eigenface approach uses Principal Components Analysis (PCA), a classical multivariate statistics method, to linearly project face images in a low-dimensional space [12]. 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 v_i , the eigenvectors of $A^T A$, an $R \times R$ matrix. Then μ_i can be computed from v_i as follows [12]:

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

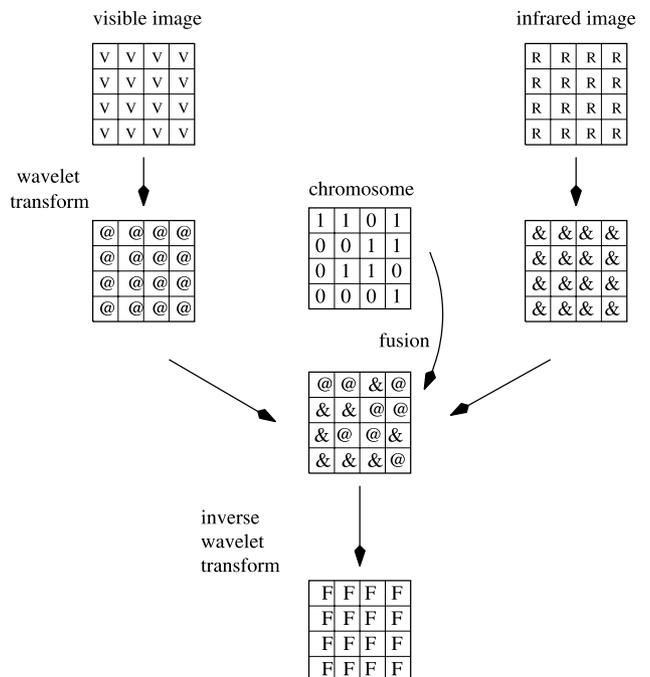


Fig. 1. Fusion scheme in the wavelet domain.

Usually, we only need to keep a much 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:

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

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

5.1. Methodology

Fusion in the eigenspace domain involves combining the eigenfeatures from the visible and IR images. Specifically, first we compute two eigenspaces, one using the visible face images and the other using the IR 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 by projecting the visible face image in the visible-eigenspace. Fusion is performed by selecting some eigenfeatures from the IR-eigenspace and some from the visible-eigenspace. GAs are employed again to decide which eigenfeatures to select from each eigenspace. Fig. 2 illustrates the main steps of this approach. Recognition is performed in the fused eigenspace.

6. Fusion using GAs

Deciding which wavelet coefficients or eigenfeatures to select from each spectrum is essentially a search problem. In this work, we propose using 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 [36]. They were designed to efficiently search large, non-linear, poorly understood search spaces. In the past, GAs have been used in target recognition [37], object recognition [38], face detection/verification [39,40], and feature selection [41–44]. Goldberg [45] provides a nice introduction to GAs and the reader is referred to this source for further information.

Several different methods have been reported in the literature for finding an optimum fusion strategy, some quite simple based on pixel averaging and some more powerful based on probabilistic schemes [46,47]. Our decision to use GAs was based on several reasons. First, the problem of fusion involves searching very large spaces. GAs have demonstrated good performance in searching large spaces in various application domains including object recognition [38] and feature selection [43]. 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 generations [36]). Third, 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 [48,49]. Fourth, they suitable for parallelization and linear speedups are the norm, not the exception [50]. Finally, we have had considerable success in the past applying GAs in related problems [38,41–44].

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. The genetic search process is iterative: evaluating, selecting, and recombining strings in the population during each iteration (generation) until reaching some termination condition (Fig. 3).

Evaluation of each string is based on a fitness function that is problem-dependent. It determines which of the candidate solutions are better. This corresponds to the environmental determination of survivability in natural selection. Selection of a string, which represents a point in the search space, depends on the string’s fitness relative to those of other strings in the population. It probabilistically removes, from the population, those points that have relatively low fitness. Mutation, as in natural systems, is a very low probability operator and just flips

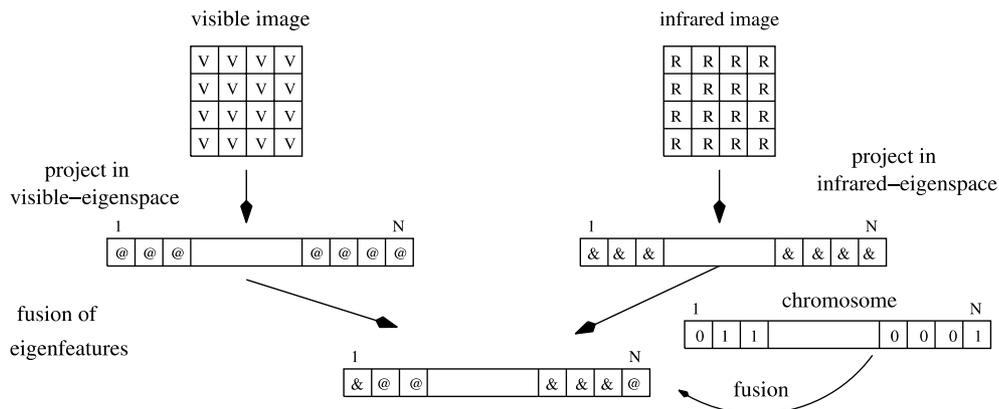


Fig. 2. Fusion scheme in the eigenspace domain.

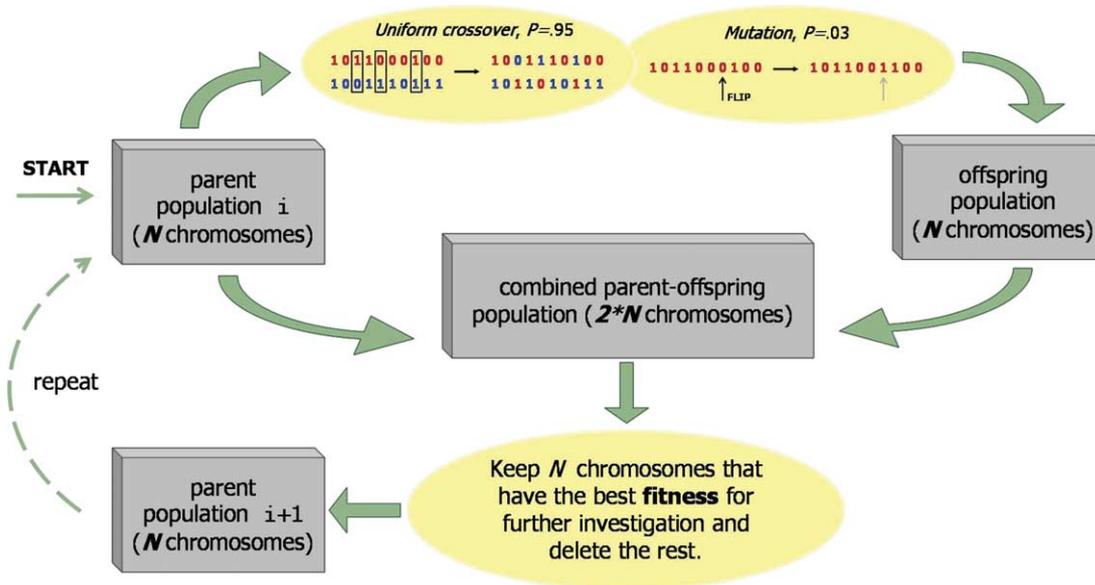


Fig. 3. GA-based framework.

a specific bit. Mutation 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 points. Its goal is to preserve the fittest individuals without introducing any new value.

In summary, selection probabilistically filters out solutions that perform poorly, choosing high performance solutions to concentrate on or exploit. Crossover and mutation, through string operations, generate new solutions for exploration. Given an initial population of elements, GAs use the feedback from the evaluation process to select fitter solutions, eventually converging to a population of high performance solutions. GAs do not guarantee a global optimum solution. However, 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 generations [45]).

6.1. Implementation details

We describe below 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.

6.1.1. 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) (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 (Section 7), 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) (Fig. 2).

6.1.2. 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 (Section 7) 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. Fig. 4 illustrates the evolutionary-based fusion approach in the wavelet domain.

6.1.3. 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.

6.1.4. 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 [51] (Fig. 4).

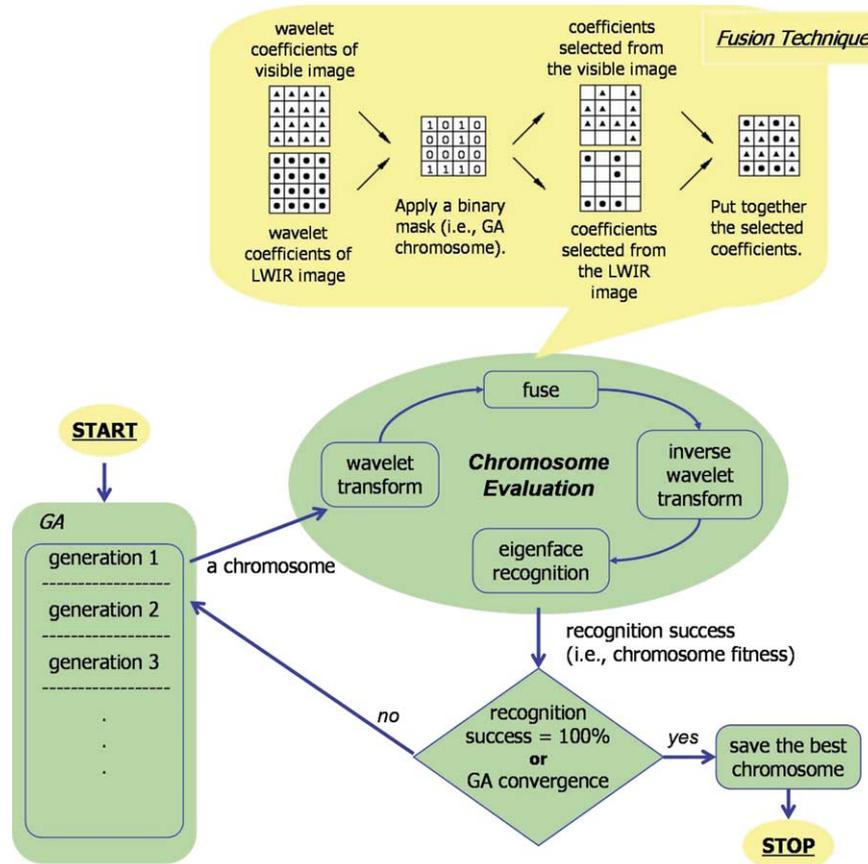


Fig. 4. Illustration of evolutionary-based fusion approach in the wavelet domain.

6.1.5. 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 0.95.

6.1.6. Mutation

Mutation is a very low probability operator, which flips the values of randomly chosen bit. The mutation probability used here was 0.03.

7. Face dataset

In our experiments, we used the face database collected by Equinox Corporation under DARPA’s HumanID program [52]. Specifically, we used the long-wave infrared (LWIR) (i.e. 8–12 μm) and the corresponding visible spectrum images from this database. The data was collected during a 2-day period. Each pair of LWIR and visible light images was taken simultaneously and co-registered with 1/3 pixel accuracy (Fig. 5). The LWIR images were radiometrically calibrated and stored as gray-scale images with 12 bits per pixels. The visible images are also gray-scale 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 [11,17]. The goal of preprocessing was to align and scale the faces, remove background, and account for some illumination variations (Fig. 1). For comparison purposes, we have

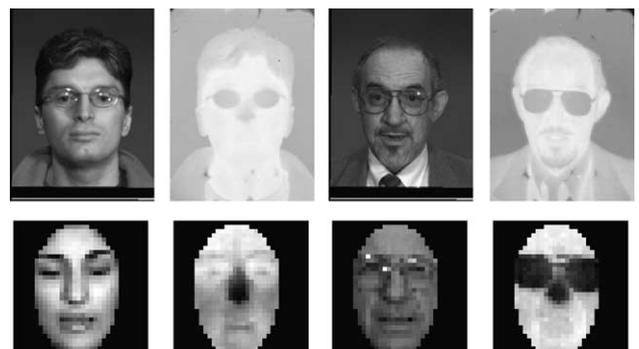


Fig. 5. Examples of visible and IR image pairs (first row) and preprocessed images (second row).

attempted to evaluate our fusion schemes using a similar experimental protocol to that given in [11,17]. 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).

8. Experimental procedure

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 [11,17]. In the eigenface approach, an image is represented as a low-dimensional feature vector, containing the coefficients of the projection of the image in the eigenspace. Recognition is 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) [12]. 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. If N is the number of images in the test set, the recognition ratio r is computed as follows

$$r = \frac{1}{N} \sum_{i=1}^N x_i, \quad (10)$$

where $x_i = 1$ if the top match from the gallery set belongs to the same subject and $x_i = 0$ otherwise. 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 data set 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 nine runs for each gallery/test set experiment. We used population sizes between 100 and 200 and 100 generations. For convergence, we used the ratio between the best fitness and the average fitness in each generation.

8.1. Eyeglasses experiments

Measuring the effect of eyeglasses is done by using the expression frames. There are 43 subjects wearing glasses in the EA set making a total of 822 images. Following the terminology in [11,17] we created the following test sets (Fig. 6): EG (expression frames with glasses, all illuminations), EnG (expression frames without glasses, all illuminations), EFG (expression frames with glasses, frontal illumination), ELG (expression frames with glasses, lateral illumination), EFnG (expression frames without glasses, frontal illumination), ELnG (expression frames without glasses, lateral illumination). The inclusion

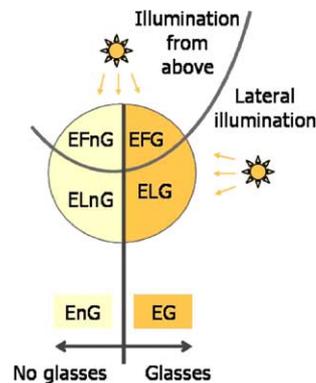


Fig. 6. Eyeglasses data.

relations among these sets are as follows:

$$\begin{aligned} EG &= ELG \cup EFG, \quad EnG = ELnG \cup EFnG \quad \text{and} \\ EG \cap EnG &= \emptyset \end{aligned} \quad (11)$$

8.2. Facial expression experiments

The test sets for the facial expression experiments include the images containing the three expression frames and three vocal pronunciation frames. There are 90 subjects with a total of 1266 pairs of images for the expression frames and 1299 for the vocal frames. Some of the subjects in these tests sets wear glasses while others not. Following the terminology in [11,17] we have created the following test sets (Fig. 7): EA (expression frames, all illuminations), EL (expression frames, lateral illuminations), EF (expression frames, frontal illumination), VA (vocal frames, all illumination), VL (vocal frames, lateral illumination), VF (vocal frames, frontal illumination). The inclusion relations among these sets are as follows:

$$\begin{aligned} EA &= EL \cup EF, \quad VA = VL \cup VF \quad \text{and} \\ VA \cap EA &= \emptyset \end{aligned} \quad (12)$$

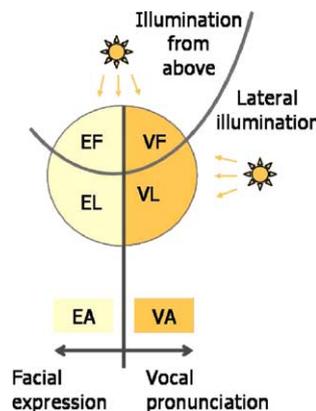


Fig. 7. Facial expression data.

Table 1

Averages and standard deviations for the eyeglasses experiments using fusion in the wavelet domain

	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

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.

9. Experimental results

9.1. Eyeglasses

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 (Table 1 and Fig. 8) and eigenspace (Table 2 and Fig. 9) 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 (Table 1). Between the two fusion schemes tested, fusion in the wavelet domain yielded higher recognition performance overall. Note that we did not perform experiments when the intersection between gallery and test sets is not empty.

9.2. Facial expression

The facial expression tests had varying success as shown in Tables 3 and 4 and Figs. 10 and 11. In general, fusion led to improved recognition performance compared to recognition in the visible spectrum. Comparing IR with fusion, sometimes IR performed better than fusion and vice versa. However, the observed differences were not statistically significant and are mainly accounted to the presence of undesired illumination effects in the fused images (Section 10). As before, we did not perform experiments when the intersection between gallery and test sets is not empty.

10. Discussion

The horizontal and vertical double lines divide each of Tables 1 and 2 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*

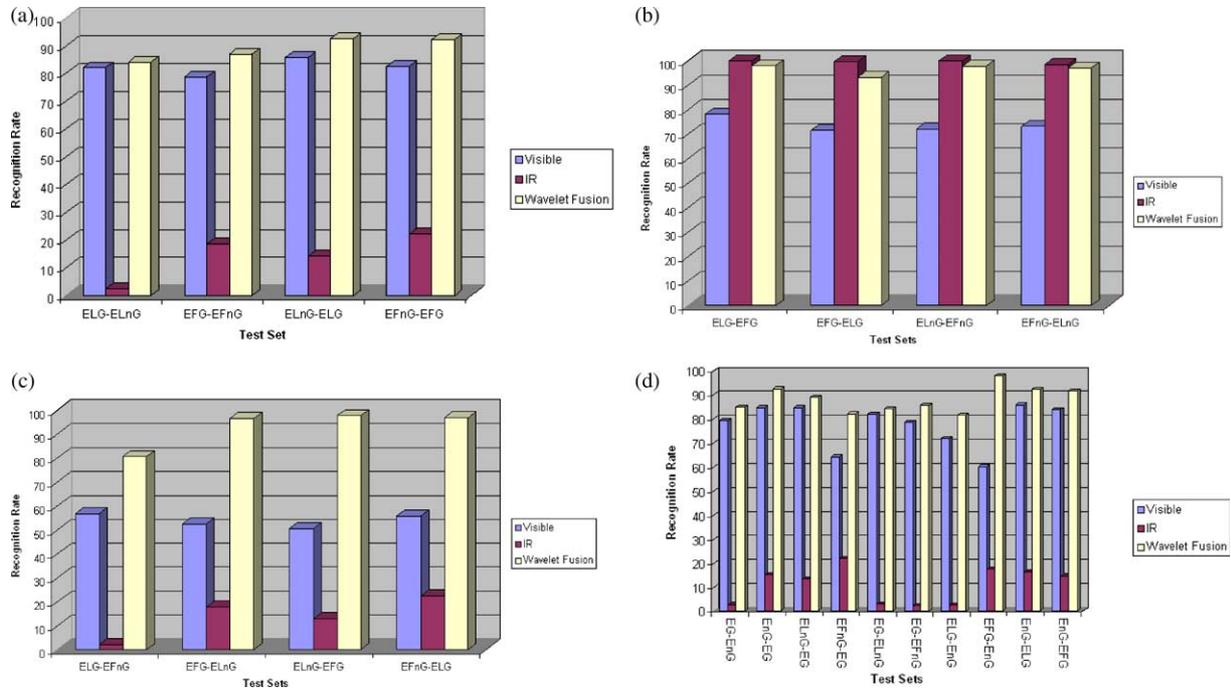


Fig. 8. Eyeglasses results in the wavelet domain: (a) same illumination conditions—eyeglasses are not present both in the gallery and probe sets; (b) eyeglasses are present both in the gallery and probe sets—illumination conditions are different; (c) eyeglasses are not present both in the gallery and probe sets—illumination conditions are different; (d) similar to (c) except that the gallery and probe sets contain multiple illuminations.

Table 2
averages and standard deviations for the eyeglasses experiments with fusion in the eigenspace domain

	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

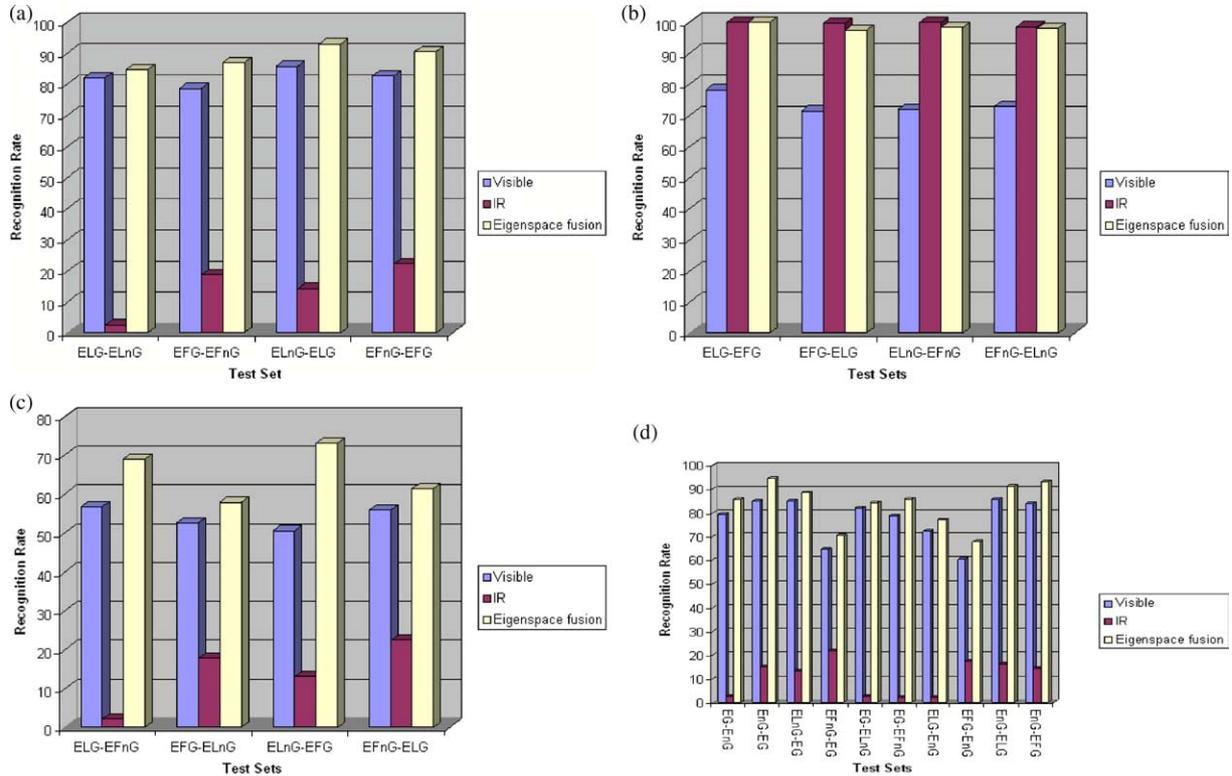


Fig. 9. Eyeglasses results in the eigenspace domain: (a) same illumination conditions—eyeglasses are not present both in the gallery and probe sets; (b) eyeglasses are present both in the gallery and probe sets—illumination conditions are different; (c) eyeglasses are not present both in the gallery and probe sets—illumination conditions are different; (d) similar to (c) except that the gallery and probe sets contain multiple illuminations.

Table 3
Averages and standard deviations for the facial expression experiments with fusion in the wavelet domain

	VA	EA	VF	EF	VL	EL
VA	×	(65.0, 1.0) (92.3, 0.7) (93.6, 0.7) 92.8 – 94.2	×	×	×	×
EA	(65.2, 1.2) (93.7, 0.4) (93.1, 0.4) 92.7 – 93.4	×	×	×	×	×
VF	×	×	×	(65.3, 1.8) (90.5, 0.4) (93.3, 0.9) 92.5 – 94.2	(62.4, 0.4) (99.3, 0) (98.1, 0.5) 97.7 – 98.6	(36.4, 0.7) (87.9, 0.6) (91.4, 1.4) 89.8 – 92.3
EF	×	×	(59.2, 0.9) (91.3, 0.7) (91.2, 1.3) 90.3 – 92.8	×	(30.4, 0.7) (91.2, 0.4) (90.0, 1.4) 88.6 – 91.5	(68.5, 1.1) (99.5, 0.4) (98.2, 0.4) 97.7 – 98.4
VL	×	×	(60.5, 0.7) (98.4, 0.2) (97.6, 0.8) 97.0 – 98.4	(28.0, 1.2) (90.4, 0.1) (86.9, 0.8) 85.9 – 87.4	×	(66.4, 0.4) (91.3, 0.4) (94.5, 1.5) 92.9 – 95.7
EL	×	×	(31.6, 1.1) (85.8, 0.7) (84.4, 0.3) 84.2 – 84.7	(63.0, 0.9) (98.6, 0.2) (96.4, 0.7) 95.8 – 97.2	(68.9, 1.0) (91.7, 0.2) (93.7, 0.4) 93.3 – 94.1	×
number of images	1299	1266	435	429	864	837

Table 4
Averages and standard deviations for the facial expression experiments with fusion in the eigenspace domain

	VA	EA	VF	EF	VL	EL
VA	×	(65.0, 1.0) (92.3, 0.7) (94.7, 4.5) 90.8 – 99.6	×	×	×	×
EA	(65.2, 1.2) (93.7, 0.4) (92.7, 6.6) 86.4 – 99.6	×	×	×	×	×
VF	×	×	×	(65.3, 1.8) (90.5, 0.4) (92.5, 1.4) 90.9 – 93.6	(62.4, 0.4) (99.3, 0) (99.3, 0.2) 99.1 – 99.5	(36.4, 0.7) (87.9, 0.6) (86.8, 0.7) 86.1 – 87.5
EF	×	×	(59.2, 0.9) (91.3, 0.7) (90.1, 0.8) 89.2 – 90.1	×	(30.4, 0.7) (91.2, 0.4) (92.6, 1.8) 91.5 – 94.7	(68.5, 1.1) (99.5, 0.4) (99.4, 0.6) 98.7 – 99.8
VL	×	×	(60.5, 0.7) (98.4, 0.2) (97.5, 0.5) 96.9 – 97.9	(28.0, 1.2) (90.4, 0.1) (87.3, 0.4) 86.8 – 87.6	×	(66.4, 0.4) (91.3, 0.4) (92.7, 6.3) 87.3 – 99.6
EL	×	×	(31.6, 1.1) (85.8, 0.7) (84.4, 1.5) 82.6 – 85.6	(63.0, 0.9) (98.6, 0.2) (97.0, 0.2) 96.7 – 97.2	(68.9, 1.0) (91.7, 0.2) (91.7, 1.1) 90.5 – 92.2	×
number of images	1299	1266	435	429	864	837

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.

Several interesting conclusions can be drawn by considering these results. As expected, face recognition in the IR is not influenced by illumination changes. 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 less sensitive to this factor (cases (1) and (2)). Illumination changes had an important influence on the success of face recognition in the visible domain (case (3)). The success of face recognition using fused images was similar in all four quadrants. This implies that fusion was able to become less sensitive both to eyeglasses and illumination changes.

To better illustrate the performance of the fusion approach, we have interpolated the results in Table 1 and used a simple visualization scheme to remove small differences and emphasize major trends in recognition performance (Fig. 12). Our visualization scheme assigns a grayscale value to each average from Table 1) with black implying 0% recognition and white 100% recognition. The empty cells from Table 1 are also shown in black. As it can be observed, Fig. 12(d) is closer to Fig. 12(a) than Fig. 12(b) and (c).

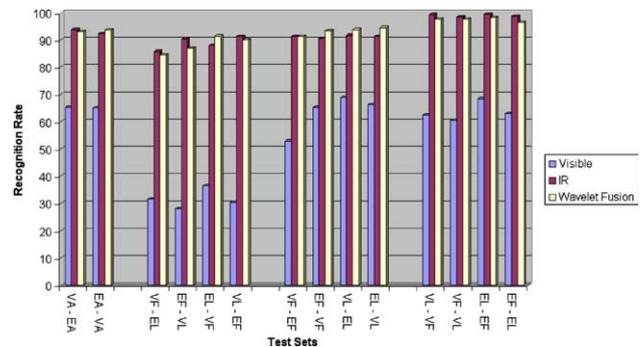


Fig. 10. Facial expression results in the wavelet domain.

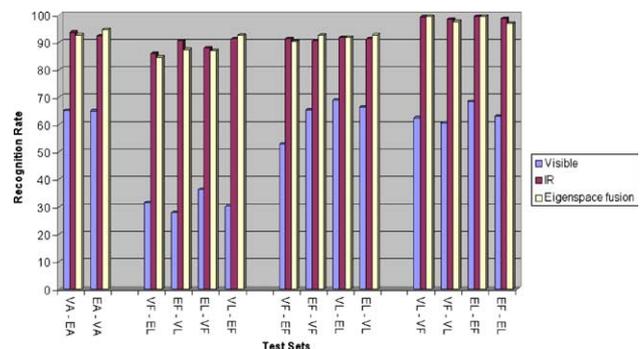


Fig. 11. Facial expression results in the eigenspace domain.

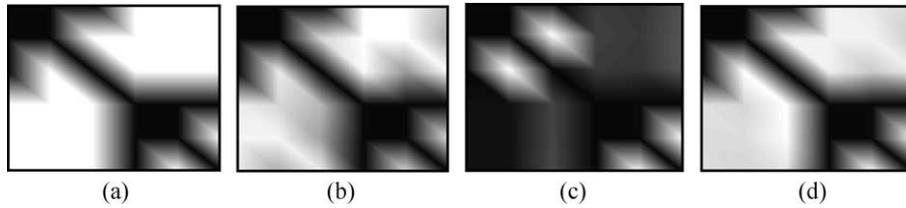


Fig. 12. The average performance values from Table 1, visualized as a grayscale image. See text for details. (a) Ideal case (b) visible images (c) IR images, (d) fused images.

Using fusion in the wavelet domain led to higher recognition performance compared to using visible images only, however, this was not always the case in the IR, especially when recognition in the IR was high (i.e. greater than 97%). Although the observed performance differences were not statistically significant, we noticed that all the cases where IR performed better than fusion were cases where the illumination direction in the gallery set was different from that in the test set and vice versa, assuming no eyeglasses differences (e.g. ELG/EFG, EFG/ELG, ELnG/EFnG, and EFnG/ELnG). This is accounted mainly to the fact that fusion was not able to completely disregard undesired illumination effects present in the visible images.

This assumption was confirmed by observing the reconstructed fused images shown in Fig. 13, as well as their first eigenfaces shown in Fig. 14. The reconstructed images show a blocky structure. Depending on the length of the basis functions the blocks may turn to spots with more round shapes and lower contrast but cannot disappear completely. However, the visual quality of the reconstructed images is not important for our application. Overall, the fused images have higher resolution compared to the IR images, however, they are also affected by undesired illumination artifacts present in the visible images. Clearly, the first few eigenfaces of the fused images still encode illumination direction. Fig. 14(c) shows some fused images using ELnG as gallery while Fig. 14(d) shows some fused images using EG as gallery. Quite interestingly, the reconstructed regions around the eyes are darker in the last case (i.e. to account for the IR blocking effect in the gallery due to eyeglasses).

Fusion in the eigenspace domain always yielded higher recognition performance compared to recognition both in the visible and IR domains with only one exception (i.e. EFG/ELG). This exception could be explained again by the difference in illumination direction between the gallery and test sets. For comparison purposes, we reconstructed several visible and IR images from the ELnG set using only the subset of eigenvectors selected from each spectrum (Fig. 15). 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 was better (Fig. 16(a)) although more computationally intensive. Significant differences in recognition performance can be noticed in the following cases:

EFG/EnG (30% higher recognition), EFG/ELnG (38% higher recognition), EFnG/ELG (23% higher recognition), EFnG/EG (12% higher recognition), and ELG/EFnG (11% higher recognition). We attribute the higher recognition performance using fusion in the wavelet domain to the more powerful eigenfeatures computed using fused images.

In the case of facial expressions (Tables 3 and 4 and Figs. 10 and 11), IR performed quite well as it has been reported in other studies [3,4]. Recognition in the visible spectrum was not satisfactory while recognition using fused images had comparable performance to that in the IR. In particular, we observed some performance differences between fusion and IR with IR being better than fusion in certain cases and vice versa. The differences are not statistically significant and are accounted again to different illumination direction between the gallery and probe sets and the preservation of undesired illumination effects in the fused images. Fig. 16(b) shows the overall accuracy achieved in these experiments.

We have also attempted to analyze the GA solutions in order to understand what parts of the face are encoded by IR features and what parts are encoded by visible features. Our analysis indicated that almost half of the features (i.e. wavelet coefficients or eigenfeatures) came from the IR spectrum while the other half came from the visible spectrum. However, the eye regions were encoded mostly using features from the visible spectrum. We did not observe any other major trends in

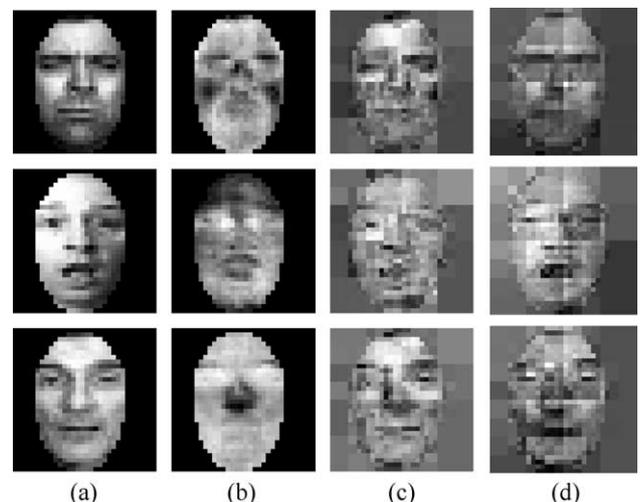


Fig. 13. Fusion in the wavelet domain: (a) visible and (b) IR images, (c) fused images using ELnG as gallery (d) fused images using EG as gallery.

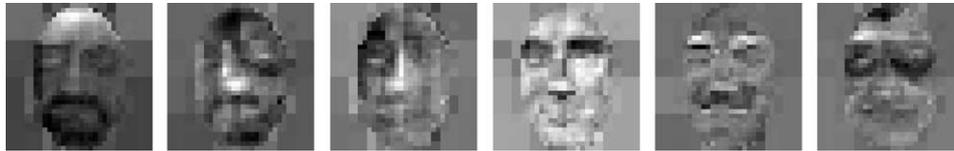


Fig. 14. Fusion in the wavelet domain—the first few eigenfaces of several fused images. The second and third eigenfaces show clear influence of right and left lateral illumination.

encoding other facial features. We also examined whether information from certain resolution levels (i.e. in the wavelet domain) or eigenvector ranges (i.e. in the eigenspace domain) were represented more heavily in the fused images. The solutions obtained indicated that the eye regions were represented mostly using wavelet coefficients from higher resolution levels and eigenfeatures corresponding to smaller eigenvectors. This results indicates a tendency to improve the resolution of the eye regions.

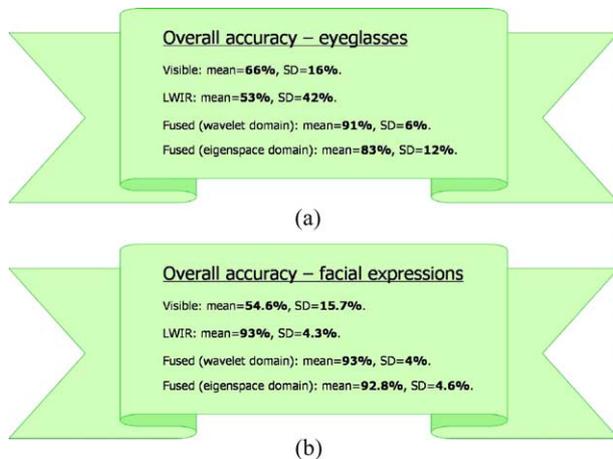


Fig. 15. Overall recognition accuracy: (a) case of eyeglasses; (b) case of facial expressions.

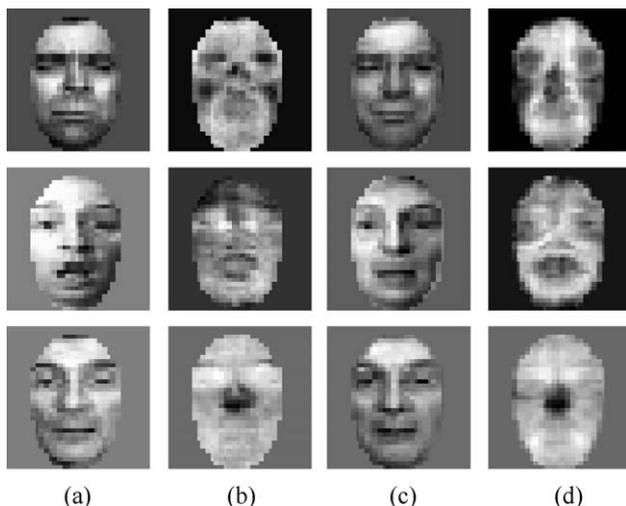


Fig. 16. Fusion in the eigenspace domain: (a) visible and (b) IR images assuming lateral illumination; (c) reconstructed visible images and (d) reconstructed IR images using the eigenvectors selected from each domain only.

11. Conclusions and future work

We presented and compared two different fusion schemes for combining IR and visible imagery for the purpose of face recognition. The first scheme was pixel-based, operating in the wavelet domain. The second scheme was feature-based, operating in the eigenspace domain. Both schemes aimed at improved and robust recognition performance across variable lighting and presence/absence of eyeglasses. Between the two schemes, fusion in the wavelet domain demonstrated better recognition performance overall (Fig. 16) although being more computationally intensive. Fast, Haar-like transforms [53], are worth considering to reduce time requirements.

Further consideration should also be given to the existence of many optimal solutions found by the GA. Although optimal in the training phase (i.e. low validation error), these solutions showed different recognition performances when considered 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. The use of larger validation sets and more selective fitness functions might help to eliminate these issues.

Future work includes considering more effective fusion schemes. We have performed preliminary experiments in the case of eigenspace-based fusion with chromosomes of length 200, using a 0/1 encoding as before. In this case, the first 100 locations of the chromosome correspond to visible eigenfeatures while the last 100 locations correspond to IR eigenfeatures. In this scheme, instead of selecting a single eigenfeature (i.e. visible or IR) at particular location, we can select or reject both eigenfeatures. Repeating the eyeglasses experiments in this case, we noticed a 3% overall improvement in recognition accuracy. A generalization of this scheme would be using non-binary encoding based on a weighted averaging. In this case, each feature from the visible and IR domains is assigned a weight instead of simply selecting/not-selecting it.

We also plan to consider more powerful fitness functions, for example, by adding extra terms to control the number of wavelet coefficients selected from different resolution levels (e.g. we can favor high or low resolution levels) or eigenfeatures from different eigenvector ranges (e.g. we can favor large or small eigenvectors). We also plan to consider fitness approximation schemes [54] to reduce the computational requirements of fitness evaluation,

Additional issues for future research include 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. For example, Chen et al.

[9] demonstrated that IR-based recognition degrades when there is substantial passage of time between the gallery and probe images. The proposed evolutionary-based fusion methodology is general enough and can be applied in these cases as well to improve recognition performance when IR information is not very reliable.

Acknowledgements

This research was supported in part by NSF under CRCRD grant No. 0088086.

References

- [1] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, *ACM Computing Surveys* 35 (4) (2003) 399–458.
- [2] L. Wolff, D. Socolinsky, C. Eveland, Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery, in: *IEEE Workshop on Computer Vision Beyond the Visible Spectrum*, Hawaii, 2001.
- [3] D. Socolinsky, A. Selinger, J. Neuheisel, Face recognition with visible and thermal infrared imagery, *Computer Vision and Image Understanding* (2003) 72–114.
- [4] G. Friedrich, Y. Yeshurun, Seeing people in the dark: face recognition in infrared images, in: *Second BMCV*, 2003.
- [5] A. Jain, R. Bolle, S. Pankanti, *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, Dordrecht, 1999.
- [6] I. Pavlidis, P. Symosek, The imaging issue in an automatic face/disguise detection system, in: *IEEE Workshop on Computer Vision Beyond the Visible Spectrum*, 2000, pp. 15–24.
- [7] J. Park, T. Oh, S. Ahn, S. Lee, Glasses removal from facial image using recursive error compensation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (5) (2005) 805–811.
- [8] J. Wilder, J. Phillips, C. Jiang, S. Wiener, Comparison of visible and infrared imagery for face recognition, in: *Second International Conference on Automatic Face and Gesture Recognition*, Killington, 1996, pp. 182–187.
- [9] X. Chen, P. Flynn, K. Bowyer, 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.
- [10] J. Heo, S. Kong, B. Abidi, M. Abidi, Fusion of visual and thermal signatures with eyeglass removal for robust face recognition, in: *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.
- [11] A. Selinger, D. Socolinsky, Appearance-based facial recognition using visible and thermal imagery: a comparative study, Equinox Corporation no. 02–01, Technical Report, 2002.
- [12] M. Turk, A. Pentland, Eigenfaces for recognition, *CogNeuro* 3 (1) (1991) 71–96.
- [13] A. Gyaourova, G. Bebis, I. Pavlidis, Fusion of infrared and visible images for face recognition, in: *European Computer Vision Conference (ECCV)*, 2004.
- [14] F. Prokoski, History, current status, and future of infrared identification, in: *IEEE Workshop on Computer Vision Beyond the Visible Spectrum*, Hilton Head, 2000.
- [15] D. Socolinsky, A. Selinger, Thermal face recognition in an operational scenario, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [16] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition—a review, *Computer Vision and Image Understanding* 97 (2005) 103–135.
- [17] D. Socolinsky, A. Selinger, Comparative study of face recognition performance with visible and thermal infrared imagery, in: *International Conference on Pattern Recognition*, 2002, pp. 217–222.
- [18] A. Selinger, D. Socolinsky, Face recognition in the dark, in: *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.
- [19] A. Srivastana, X. Liu, Statistical hypothesis pruning for recognizing faces from infrared images, *Image and Vision Computing* 21 (2003) 651–661.
- [20] P. Buddharaju, I. Pavlidis, I. Kakadiaris, Face recognition in the thermal infrared spectrum, *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.
- [21] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyperspectral images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (12) (2003) 1552–1560.
- [22] A. Eide, C. Jahren, S. Jørgensen, T. Lindblad, C. Lindsey, K. Osterud, Eye identification for face recognition with neural networks, in: *SPIE Conference*, Orlando, 1996.
- [23] J. Dowdall, I. Pavlidis, G. Bebis, A face detection method based on multi-band feature extraction in the near-ir spectrum, *IEEE Workshop on Computer Vision Beyond the Visible Spectrum*, 2001.
- [24] J. Dowdall, I. Pavlidis, G. Bebis, Face detection in the near-ir spectrum, *Image and Vision Computing* 21 (7) (2001) 565–578.
- [25] J. Dowdall, I. Pavlidis, G. Bebis, Face detection in the near-ir spectrum, *AeroSense SPIE Conference (Infrared Technology and Applications XXIX)*, 2003.
- [26] C. Eveland, D. Socolinsky, L. Wolff, Tracking human faces in infrared video, *Image and Vision Computing* 21 (2003) 579–590.
- [27] P. Scheunders, Local mapping for multispectral image visualization, *Image and Vision Computing* 19 (13) (2001) 971–978.
- [28] S.S.K. Vani, K. Lakshmi, Comparison of conventional and wavelet transform techniques for fusion of irs-1c, kiss-iii, and pan images, in: *Proceedings of the ACRS*, vol. 1, 2001, pp. 140–145.
- [29] Y. Zhou, R. Hecht-Nielsen, Target recognition using multiple sensors, in: *IEEE Workshop on Neural Networks for Signal Processing*, 1993, pp. 411–420.
- [30] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, *Information Fusion* 4 (2003) 259–280.
- [31] B. Manjunath, R. Chellappa, C. von der Malsburg, A feature based approach to face recognition, in: *Computer Vision and Pattern Recognition*, 1992, pp. 373–378.
- [32] E. Adelson, P. Burt, Image data compression with the laplacian pyramid, in: *Pattern Recognition and Image Processing*, Dallas TX, 1981, pp. 218–223.
- [33] C. Chui, *An Introduction to Wavelets*, Academic Press, London, 1992.
- [34] L. Chipman, T. Orr, Wavelets and image fusion, in: *IEEE International Conference on Image Processing*, vol. 3, 1995, pp. 248–251.
- [35] H. Li, B. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, in: *IEEE International Conference on Image Processing*, Austin TX, vol. 1, 1994, pp. 51–55.
- [36] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison–Wesley, Reading, MA, 1989.
- [37] A. Katz, P. Thrift, Generating image filters for target recognition by genetic learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (9) (1994).
- [38] G. Bebis, S. Louis, Y. Varol, A. Yfantis, Genetic object recognition using combinations of views, *IEEE Transactions on Evolutionary Computation* 6 (2) (2002) 132–146.
- [39] D. Swets, B. Punch, Genetic algorithms for object localization in a complex scene, in: *IEEE International Conference on Image Processing*, 1995, pp. 595–598.
- [40] G. Bebis, S. Uthiram, M. Georgiopoulos, Face detection and verification using genetic search, *International Journal of Artificial Intelligence Tools* 9 (2) (2000) 225–246.
- [41] Z. Sun, X. Yuan, G. Bebis, S. Louis, Genetic feature subset selection for gender classification: a comparison study, in: *IEEE Workshop on Applications of Computer Vision*, 2002.
- [42] Z. Sun, G. Bebis, R. Miller, Boosting object detection using feature selection, in: *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2003.
- [43] Z. Sun, G. Bebis, R. Miller, Object detection using feature subset selection, *Pattern Recognition* 37 (2004) 2165–2176.

- [44] Z. Sun, G. Bebis, R. Miller, On-road vehicle detection using evolutionary gabor filter optimization, *IEEE Transactions on Intelligent Transportation Systems* 6 (2) (2005) 125–137.
- [45] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison–Wesley, Reading, MA, 1989.
- [46] R. Sharma, T. Leen, M. Pavel, Probabilistic image sensor fusion, *Advances in Neural Information Processing Systems* 11 (1999).
- [47] R. Blum, J. Yang, Image fusion using the expectation-maximization algorithm and a gaussian mixture model, in: R. Blum, Z. Liu (Eds.), *Multi-Sensor Image Fusion and Its Applications*, Marcel Dekker, New York, 2004.
- [48] D. Fogel, *Evolutionary Computation, Toward a New Philosophy of Machine Intelligence*, IEEE Press, New York, 1992.
- [49] J. Koza, *Genetic Programming*, MIT Press, Cambridge, MA, 1993.
- [50] E. Cantu-Paz, *Efficient and Accurate Parallel Genetic Algorithms and Evolutionary Computation*, Kluwer, Dordrecht, 2000.
- [51] I. Eshelman, The chc adaptive search algorithm: how to have safe search when engaging in non-traditional genetic recombination, in: *The Foundation of Genetic Algorithms Workshop*, 1989, pp. 265–283.
- [52] <http://www.equinoxsensors.com/products/HID.html>
- [53] P. Viola, M. Jones, Robust real-time object detection, in: *IEEE Computer Vision and Pattern Recognition Conference*, 2001.
- [54] Y. Jin, B. Sendhoff, Fitness approximation in evolutionary computation—a survey, in: *Genetic and Evolutionary Computation Conference*, 2002.