Fusion of visible and infrared imagery for face recognition

Xuerong Chen (陈雪荣), Zhongliang Jing (敬忠良), Shaoyuan Sun (孙韶媛), and Gang Xiao (肖 刚)

Institute of Aerospace Information and Control, School of Electrical and Information Engineering, Shanghai Jiaotong University, Shanghai 200030

Received June 17, 2004

In recent years face recognition has received substantial attention, but still remained very challenging in real applications. Despite the variety of approaches and tools studied, face recognition is not accurate or robust enough to be used in uncontrolled environments. Infrared (IR) imagery of human faces offers a promising alternative to visible imagery, however, IR has its own limitations. In this paper, a scheme to fuse information from the two modalities is proposed. The scheme is based on eigenfaces and probabilistic neural network (PNN), using fuzzy integral to fuse the objective evidence supplied by each modality. Recognition rate is used to evaluate the fusion scheme. Experimental results show that the scheme improves recognition performance substantially.

OCIS codes: 100.5010, 350.2660, 110.3080, 110.6820.

Considerable progress has been made in face recognition research over the last decade[1], because of the commercial interest and the development of feasible technologies. Unlike other forms of identification such as fingerprint analysis and iris scans, face recognition is user-friendly and non-intrusive. Current face recognition systems have advanced to be fairly accurate under constrained scenarios, but extrinsic imaging parameters such as pose variation, facial expression and, especially illumination (Fig. 1), still cause much difficulty in correct recognition.

Recently, a number of studies have shown that infrared (IR) imagery of human faces offers a promising alternative to visible imagery due to its relative insensitive to illumination changes. IR imagery is nearly invariant under all lighting conditions including total darkness. An overview of identification in the IR spectrum can be found in Ref. [2]. In Ref. [3], it was claimed that IR imagery of human faces is not only a valid biometric, but almost surely a superior one to comparable visible imagery.

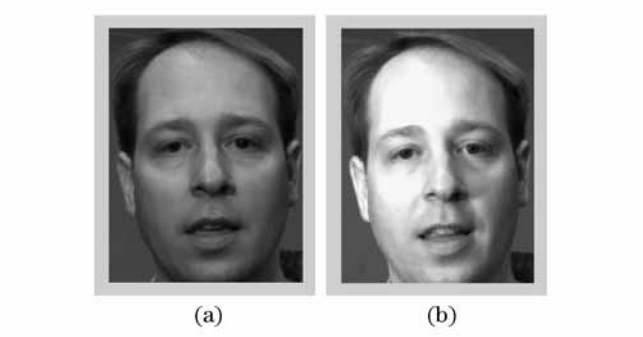

Fig. 1. The same visible face. (a) Frontal illumination and (b) lateral illumination.

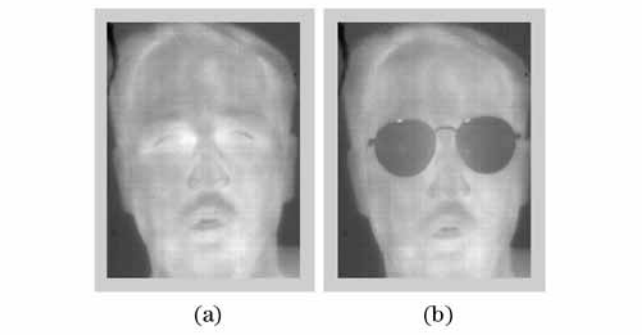

Fig. 2. The same IR face. (a) Without eyeglasses and (b) with eyeglasses.

However, IR has its own limitations[2,4]. One limitation is the higher cost of an IR sensor, compared with a visible sensor. During the last twenty years, IR camera technology has been improved significantly and price has been reduced dramatically. So the price will not be a problem in the next few years. Another limitation of IR imagery is that IR is opaque to glass (Fig. 2), which makes a large portion of the face wearing eyeglasses hid. Other limitations include its sensitivity to ambient temperature, wind, and metabolic process in the subject.

So, a future line of research is to develop face recognition algorithms that fuse information from the two sensors. Several fusion schemes were proposed[5], which employed genetic algorithms (GAs) to find an optimum strategy to perform the fusion. But these schemes may not be easily realized due to the huge computation of GAs and existence of many optimal solutions found by the GAs. In this paper, we propose a new fusion scheme. The scheme is based on eigenfaces and probabilistic neural network (PNN), using fuzzy integral to fuse the objective evidence supplied by each modality. We use recognition rate to evaluate the fusion scheme, and the experimental results show recognition performance is improved substantially.

Eigenface approach was developed by M. Turk and A. Pentland in 1991[6]. This method utilizes the idea of the principal component analysis and decomposes face images into a small set of characteristic feature images called eigenfaces.

Let a face image $\Gamma(x, y)$ be a two-dimensional N by N array of intensity values. It can also be represented as a

N^2 vector Γ_n. Let the training set of face images be Γ_1, Γ_2, Γ_3,$\cdots$, Γ_M. The average face of the set is defined as

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n, \tag{1}$$

where M is the number of faces in the training set.

Each face differs from the average by the vector $\Phi_n = \Gamma_n - \Psi$. The covariance matrix is estimated by

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T, \tag{2}$$

where the matrix $A = [\Phi_1, \Phi_2, \cdots, \Phi_M]$. The eigenspace can then be defined by computing the eigenvectors u_n of C. Since C is very large, u_n can be computed as

$$u_n = \sum_{k=1}^{M} v_{nk}\Phi_k, \quad n = 1, 2, \cdots, M, \tag{3}$$

where v_n are the eigenvectors of A^TA.

We pick up only the most meaningful eigenvectors corresponding to the largest eigenvalues, and ignore the rest. Thus, the number of basis functions is further reduced from M to M'. A new face image Γ is transformed into its eigenface components by

$$\omega_n = u_n\left(\Gamma - \Psi\right), \quad n = 1, 2, \cdots, M'. \tag{4}$$

The eigenface components form a vector $\Omega^T = [\omega_1, \omega_2, \cdots, \omega_{M'}]$, used to find which one the new face belongs to in a number of predefined face classes.

The simplest classification method is top-match. In this paper, we employ PNN to classify the eigenface components. PNN was developed by D. F. Specht[7]. This network offers high classification performance, rapid training speed, guaranteed convergence and robustness to noisy examples. Figure 3 illustrates the architecture of a PNN network. The PNN consists of nodes allocated in three layers after the inputs, they are the pattern layer, summation layer, and output layer.

Pattern layer: the number of the nodes in the pattern layer is identical to the counts of the training samples, and the weight w_{ij}^P in the input-to-pattern connection is

$$w_{ij}^P = x_i^j, \tag{5}$$

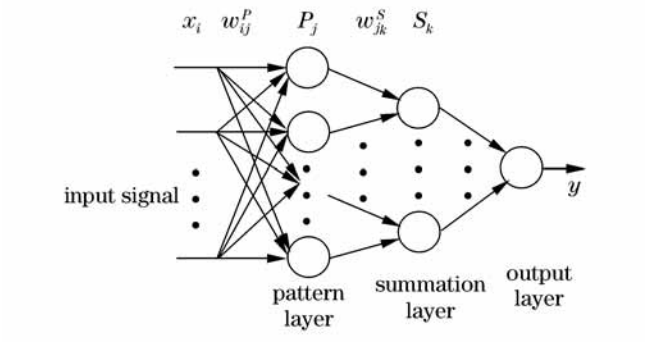

Fig. 3. Architecture of a probabilistic neural network.

where x_i^j denotes the ith node input of the jth training sample at the input layer. The transfer function of the jth pattern node P_j is

$$P_j = \exp\left(-\frac{n_j^2}{2\sigma^2}\right), \tag{6}$$

$$n_j = \sqrt{\sum_i \left(w_{ij}^P - x_i\right)^2}, \tag{7}$$

where x_i denotes the ith node input of a test sample, and σ is the smoothing factor.

Summation layer: each summation node receives the outputs from pattern nodes associated with a given class. The weight between the pattern and summation nodes can be expressed as

$$w_{jk}^S = \begin{cases} 1, & T_k^j = 1 \\ 0, & \text{else} \end{cases}, \tag{8}$$

where the value of T_k^j is 1 only when the pattern node j is associated with class k. At the summation layer, each node represents an individual class. The output of each node can be expressed as

$$S_k = \frac{1}{\sum_j w_{jk}^S}\sum_j \left(w_{jk}^S \cdot P_j\right). \tag{9}$$

Output layer: it classifies the input vector into a specific one class if that class has the maximum output value from the corresponding node at the summation layer

$$y = \arg\max_k S_k. \tag{10}$$

In a PNN network, the only factor that needs to be selected for training is the smoothing factor σ. An appropriate σ is chosen by experiment.

Our fusion scheme is described below. We assume that each face is represented by a pair of images taken simultaneously and co-registered, one in the IR spectrum and one in the visible.

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 eigenface components, one by projecting IR face image in the IR eigenspace and the other by projecting visible face image in the visible eigenspace. Each set of eigenface components is treated as a vector, and normalized prior to fusion (Fig. 4).

The goal of fusion is to combine important information from each vector. Figure 5 illustrates the main steps of our fusion scheme. The key of the scheme is to combine two standard PNN networks using a fusion layer to construct a dual probabilistic neural network (D-PNN). The number of the nodes in the fusion layer is the same as that of each summation layer.

The transfer function of the kth fusion node F_k is designed on fuzzy integral[8] which considers the objective evidence supplied by each source (called the h-function) and the expected worth of each source (via a fuzzy measure).

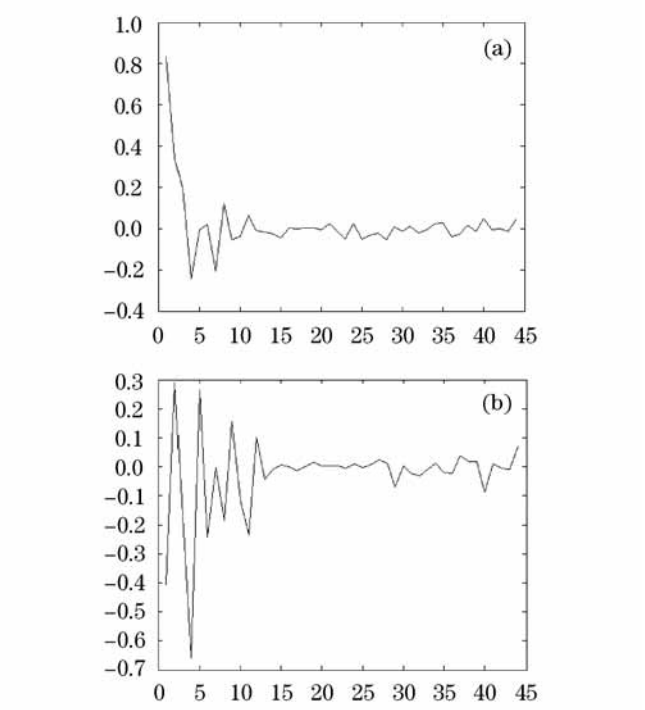

Fig. 4. Eigenface components of a test face. (a) Visible image and (b) IR image.

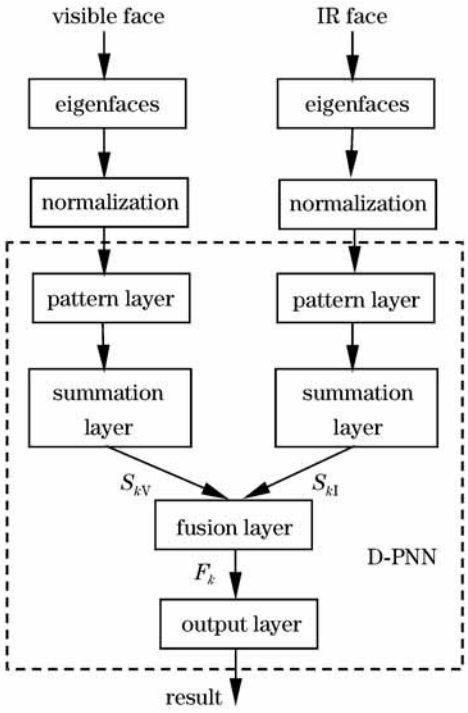

Fig. 5. Flowchart of the fusion scheme.

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of sources, g be a Sugeno fuzzy measure of X, and $h_k(x_i)$ represent the direct evidence that an object satisfies the class hypothesis C_k from the standpoint of information source x_i. The fuzzy integral is defined as

$$\int_A h_k(x) \circ g(\cdot) = \sup_{E \subseteq X} \left[\min \left(\min_{x \in E} h_k(x), g(A \cap E) \right) \right] = \sup_{\alpha \in [0,1]} \left[\min \left(\alpha, g(A \cap F_\alpha) \right) \right], \quad (11)$$

where $F_\alpha = \{x \,|\, h_k(x) \geq \alpha\}$.

In the case with two sources x_1 and x_2, suppose $h_k(x_1) \geq h_k(x_2)$. (If this is not the case for any object instance, then reorder the set of sources X so that this relation holds.) Then the fuzzy integral can be shown to be

$$e_k = \max \left[\min \left(h_k(x_1), g(A_1) \right), \min \left(h_k(x_2), g(A_2) \right) \right], \quad (12)$$

where $A_1 = \{x_1\}$, and $A_2 = \{x_1, x_2\}$.

In our fusion scheme, let x_1 represent the visible source and x_2 represent the IR source. The fuzzy density value $g^i = g\{x_i\}$ is determined via statistical measurements on recognition rate of the single source x_i. Thus the output of the kth fusion node F_k can be expressed as

$$F_k = \begin{cases} \max \left(\min \left(S_{k\mathrm{V}}, g^1 \right), S_{k\mathrm{I}} \right), & S_{k\mathrm{V}} \geq S_{k\mathrm{I}} \\ \max \left(\min \left(S_{k\mathrm{I}}, g^2 \right), S_{k\mathrm{V}} \right), & \text{else} \end{cases}, \quad (13)$$

where $S_{k\mathrm{V}}$ is the output of summation layer for visible image, $S_{k\mathrm{I}}$ is the output of summation layer for IR image.

In our experiments, we used the face database collected by Equinox Corporation. All data were acquired with a newly developed sensor capable of capturing simultaneous co-registered video sequences with a visible charge coupled device (CCD) array and an uncooled long wave infrared (LWIR) micro-bolometer. The format consists of 320×240 pixel image pairs, co-registered to within 1/3 pixel, where the visible image has 8 bits of grayscale resolution and the LWIR has 12 bits. Before experiments, each image was preprocessed to extract the face, and the size of image was reduced to 180×140 (Fig. 1).

In order to compare and evaluate our fusion scheme overall, we created a training set and 5 test sets for each modality. There were 11 subjects with a total of 44 images in the training set, and 99 images in each test set. Our criteria for training and test are as follows. Training set: frontal illumination, no eyeglasses, no expression; test_no set: frontal illumination, no eyeglasses, no expression; test_illumination set: lateral illumination, no eyeglasses, no expression; test_eyeglasses set: frontal illumination, with eyeglasses, no expression; test_expression set: frontal illumination, no eyeglasses, with expression; test_all set: lateral illumination, with eyeglasses, with expression.

In experiments, we first compared classification performance of PNN with top-match. Then we compared recognition performance of fusion with visible and IR modality, all based on eigenfaces and PNN.

We experimented with multiple values for the smoothing factor σ, and found that best performance was obtained with $\sigma = 0.6$ for both visible imagery and LWIR. We set $g^1 = 0.8$ and $g^2 = 0.9$ in Eq. (13) after statistically analyzing recognition rate of the two modalities.

Table 1 shows that the classification performance of PNN is better than top-match. Visible results are

Table 1. Eigenfaces Performance Using Different Classification Methods

	Top-Match	PNN
Test_no	0.98	1
	0.95	0.97
Test_illumination	0.25	0.31
	0.90	0.95
Test_eyeglasses	0.90	0.91
	0.81	0.85
Test_expression	0.99	1
	0.95	0.97
Test_all	0.24	0.27
	0.70	0.79

Table 2. Recognition Performance of Visible Imagery (Using Eigenfaces and PNN), Infrared Imagery (Using Eigenfaces and PNN) and Fusion (Using the Proposed Scheme)

	Visible	Infrared	Fusion
Test_no	1	0.97	1
Test_illumination	0.31	0.95	0.93
Test_eyeglasses	0.91	0.85	1
Test_expression	1	0.97	1
Test_all	0.27	0.79	0.91

reported above the corresponding LWIR results. Recognition performance of two modalities (using eigenfaces and PNN) and fusion (using the proposed scheme) are listed in Table 2. From Table 2, we can see that considerable improvements in recognition performance were achieved by fusing visible and IR imagery. Comparing row 2, 3, 4 with row 1 respectively, we get the influence of illumination, eyeglasses and expression. As expected, variation in illumination conditions between training set and test set resulted in markedly reduced performance for visible imagery, while it had no significant effect on performance of LWIR imagery. The eyeglasses had more influence for LWIR than visible imagery, since glass was completely opaque in the LWIR. Expression had almost no effect on both visible imagery and LWIR in our experiments. Comparing row 5 with row 1, we get the mixed influence of the three extrinsic imaging parameters. In all the test experiments, recognition performance of fusion is above 90%. The improvements are even greater in the test experiment with mixed effect of the three parameters, where recognition performance is improved by 64% compared with that using visible images and by 12% compared with that using LWIR images.

Experimental results show that the fusion scheme improves recognition performance substantially. Future work includes designing more effective schemes to deal with rotated face. As we know, recognition rate decreases sharply when face rotated over 30 degrees. For some security systems used in public places like airports and stations, it is difficult to keep taking face images frontally. Further consideration should also be given to video fusion, which combines information not only from different sensors, but also from different frames.

The research was jointly supported by the National Natural Science Foundation of China (No. 60375008), Aviation Science Foundation (No. 02D57003), and China Ph.D Discipline Special Foundation (No. 20020248029). X. Chen's e-mail address is xrchen@sjtu.edu.cn.

References

1. W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, ACM Computing Survey **35,** 399 (2003).
2. F. Prokoski, in *Proceedings of IEEE Workshop on Computer Vision Beyond the Visible Spectrum* 5 (2000).
3. D. A. Socolinsky and A. Selinger, in *Proceedings of International Conference on Pattern Recognition* **4,** 217 (2002).
4. J. Wilder, P. J. Phillips, C. H. Jiang, and S. Wiener, in *Proceedings of 2nd International Conference on Automatic Face & Gesture Recognition* 182 (1996).
5. S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, Proc. SPIE **5404,** 585 (2004).
6. M. A. Turk and A. P. Pentland, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 586 (1991).
7. D. F. Specht, IEEE Trans. Neural Networks **1,** 111 (1990).
8. J. M. Keller and J. Osborn, International Journal of Approximate Reasoning **15,** 1 (1996).