MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Face Recognition by Combining Complementary Matchings of Single Image and Sequential Images

Yea-Shuan Huang   Wei-Cheng Liu   Fang-Hsuan Cheng

Computer Science & Information Engineering, Chung Hua University
No.707, Sec. WuFu Rd., Hsinchu, Taiwan, 300, R.O.C.

yeashuan@chu.edu.tw   m09502017@chu.edu.tw   fhcheng@chu.edu.tw

Abstract

This paper describes a face recognition method based on combining two complementary matching algorithms, one is a single-image matching algorithm and the other is a sequential-image matching algorithm. The sequential-image matching algorithm distinguishes each person by using the derived features from a set of sequential images of the same person, and the derived features meaningfully corresponds to the unique face variation pattern of this person. For computing the matching scores, the sequential-image matching algorithm uses a Constrained Mutual Subspace Method (CMSM), and the single-image matching algorithm uses an Euclidean distance. The two matching scores are further combined together so that highly accurate recognition can be achieved. To obtain more robust features under various illuminations, an Anisotropic Smoothing Transform (AST) is also proposed in this paper which can effectively compensates a non-uniform face image into a uniform face image. Experimental results show that the proposed method can achieve outstanding performance and is considerably robust to lighting variations.

1 Introduction

Biometric identification technology is a very popular research field in the recent years. Various methods have been proposed that use different kinds of biometric data. Among them, face recognition consistently obtains a great expectation since it is contact-free and is user friendly. Therefore, a lot of research efforts have been devoted to this field, and many face recognition approaches based on a variety of machine learning theorem have been developed already. For example, subspace methods such as PCA [1,2] and LDA [3,4] are commonly used which project high dimensional features to low dimensional features and not only faster but also better recognition can be achieved.

It is well known that face images are easy to change in color and in shape when there is variation of environment lighting, facial expressions and poses. Therefore, it is apt to be unreliable if recognition is performed based on just a single input image. To obtain a more robust recognition, Yamaguchi proposed the Mutual Subspace Method (MSM) [5] and the Constrained Mutual Subspace Method (CMSM) [6,7,8], and both methods perform recognition by using multiple sequential images. In MSM, similarity is defined to be the minimum angle between the input subspace and a reference subspace. In CMSM, it further projects each individual subspace including the input and the reference subspaces onto a constrained subspace. According to the projection on the same constrained subspace, it can obtain the features which have good discrimination ability among classes and are insensitive to face poses and expressions.

Basically, the feature derived from a single image denotes the location of this image in a high dimensional feature space. In the feature space, the locations of two highly similar images will be in general close to each other, and the locations of two distinct images then will be quite separated apart. Therefore, recognition based on a single image mainly measures the distance (or similarity) of the features between the input pattern and the reference patterns. However, the feature derived from a set of sequential images of the same person can present the unique variation model of this person. Therefore, recognition based on sequential face images can compare the specific variation pattern of the unknown person and that of each enrolled user. The two kinds of recognition seem to be complementary in nature. With this understanding, it will be very useful if both methods are combined together.

This paper is organized as follows. Section 2 introduces an Anisotropic Smoothing Transform (AST) which compensates the lighting variation on face images. Section 3 describes two recognition models; the first is based on a single image and the second is based on sequential images. A linear mechanism is also proposed to integrate both recognition scores. Section 4 presents the experiment results on the famous Banca face database, and the final conclusion is drawn in Section 5.

2 Anisotropic Smoothing Transform

According to a Lambertian model [9], the value of an image pixel x can be formulated as

$$I(x) = \rho(x)n^T(x)s(x)$$

where I is the intensity value, ρ is the albedo, n is the normal vector and s is the spotlight illumination vector of point x. Basically, ρ denotes the texture information of the image and $n^t s$ represents the integral information related to the object orientation and lighting source. To reduce the illumination effect of various lighting conditions, we consider two basic image attributes, one is the high-frequency texture such as edge and face contour, which is less sensitive to various lighting conditions, and the other is the low-frequency information that is rather sensitive to lighting variations. Based on this understanding, if we get rid of the low-frequency attributes, we are more able to obtain much robust high-frequency texture information in different lighting conditions. To serve this purpose, two functions are defined

$$I = \rho\, n^T s = \rho W, \text{and}$$
$$\widetilde{W} = I * G$$

where $W = n^T s$ is an image which mainly reflects the low-frequency information of I and $\widetilde{W}$ is a convolution image constructed by the original image and a Gaussian filter. Basically, the lighting factor can be implicitly attributed to W. Because both W and $\widetilde{W}$ correspond to the low frequency signal of an image, we can use $W \approx c\widetilde{W}$ to present the relationship between both images and c is a constant value. If I is divided by $\widetilde{W}$, a new image N can be constructed which inherently reveals the high frequency attribute ρ. That is

$$N = \frac{I}{\widetilde{W}} = \frac{\rho W}{\widetilde{W}} \approx c\rho.$$

Although the high-frequency information N is less sensitive to different lighting variations, it is still sensitive to noise. To solve this problem, we design an Anisotropic Smoothing Transform (AST) to reduce the noise influence. Let k be the 4-direction neighbor points (i.e. $k \in \{E, W, S, N\}$), Δ_k be the intensity difference between the current point and its neighbor point k, δ be a scale factor of an exponential function, and W_k be the smoothing weight of direction k. Here, W_k is defined to be

$$W_k = \exp^{-\frac{|\Delta_k|}{\delta}} \quad \text{for } k \in \{\mathrm{E}, \mathrm{W}, \mathrm{S}, \mathrm{N}\}.$$

Basically, AST takes each image pixel and its four neighboring pixels with different smoothing weights to generate a new smoothed image value. That is

$$N_{i,j}^{t} = (1-\lambda)N_{i,j}^{t-1} + \lambda\left(W_E N_{i+1,j} + W_w N_{i-1,j} + W_S N_{i,j+1} + W_N N_{i,j-1}\right)$$

where $N_{i,j}^{t}$ is the image value progressed t times at the coordinate (i, j) and λ is a scaling parameter of the weighted sum function. Figure 1 shows the compensated results of several face samples which clearly demonstrate that the proposed AST indeed can extract essential texture information from images under various lighting conditions.

3 Face Identification Method

In this section, we describe the proposed face recognition framework which integrates a single-image matching module and a sequential-image matching module. For pattern matching, the single-image matching module uses the Euclidean distance metric, and the sequential-image matching module uses a CMSM (Constrained Mutual Subspace Method) metric. For making the final decision, a weighted sum is used to combine the two matching scores. This section consists of five subsections. The first introduces the concept of the canonical angle which can reflect the similarity of two subspaces. The second explains how the constrained subspace is generated. The third describes matching in the constructed constrained subspace. The fourth states the matching of a single image which adopts an Euclidean distance in a LDA-transformed reduced feature space. The fifth describes the weighted-sum combination scheme.

3.1 Concept of Canonical Angle

In linear algebra, the similarity between two subspaces is calculated by the angle between them. Suppose $\{R_1, \ldots, R_r\}$ is a set of r reference patterns, $\{I_1, \ldots, I_s\}$ is a set of s input patterns, and each pattern is represented by an f-dimensional feature vector. With PCA, an r_{no}-dimensional reference subspace Ω can be constructed from $\{R_1, \ldots, R_r\}$, and an s_{no}-dimensional input subspace Λ can be constructed from $\{I_1, \ldots, I_s\}$ respectively. Therefore, Ω is an $r_{no} \times f$ matrix and Λ is an $s_{no} \times f$ matrix. In general, the relations of r, s, r_{no} and s_{no} are chosen to be $r_{no} \leqq r$, $s_{no} \leqq s$ and $r_{no} \leqq s_{no}$. We can further obtain r_{no} canonical angles $\{\theta_1, \ldots, \theta_{r_{no}}\}$ between subspace Ω and subspace Λ by the following equations:

$$XC = \lambda C$$

$$X = (x_{ij}),\ x_{ij} = \sum_{k=1}^{r_{no}} (\psi_i \cdot \phi_k)(\phi_k \cdot \psi_j)$$

where ψ_i and ϕ_i denote respectively the i-th f-dimensional orthonormal basis vector of subspace Ω and Λ, λ is an eigenvalue of X and C is the eigenvectors of X, and X is an $r_{no} \times r_{no}$ matrix. The value $\cos^2 \theta_i$ of the i-th smallest canonical angle equals to the i-th largest eigenvalue of Λ. The largest eigenvalue (i.e. $\cos^2 \theta_1$) is taken to denote the similarity between subspace Ω and Λ. Figure 2 shows a schematic diagram of the canonical angle between two subspaces.

3.2 Generation of Constrained Subspace

In CMSM, it is essentially important to generate a proper constrained subspace C which contains the effective matching components but eliminating the unnecessary ones. By projecting the input subspace and reference subspaces to a constrained subspace, it could extract discriminating features for recognizing pattern classes.

Suppose there are in total N_p reference subspaces. To generate a constrained subspace, we compute the projection matrix Ω_j of the j-th reference subspace using

$$P_j = \sum_{k=1}^{r_{no}} \psi_k^j \left(\psi_k^j\right)^T$$

where r_{no} is the number of eigenvectors of a reference subspace, ψ_k^j is the k-th orthonormal basis vector of the j-th reference subspace, and each P_j is a $f \times f$ matrix. Then, we calculate the eigenvectors of the summation matrix $S = \left(P_1 + P_2 + \cdots + P_{N_P}\right)$, that is $SA = \lambda A$, where λ and A denote the eigenvalues and the eigenvectors of S respectively. Finally, the t eigenvectors $[A_1, \ldots, A_t]$ corresponding to the t smallest eigenvalues are selected to construct the constrained subspace CS (that is $CS = [A_1, \ldots, A_t]_{t\times f}$). For a more detailed description of CMSM, please see [6].

3.3 Matching on Constrained Subspace

Suppose there are in total K recognition classes. Π denotes the input subspace derived from the input sequence samples, and T_i ($1 \leqq i \leqq$ K) denotes the subspace derived from the training sequence samples of class *i*. Five steps need to be performed for pattern matching as follows:

1. Project each T_i onto *CS* and generate an $r_{no}\times t$ projection matrix P_i;
2. Normalize each P_i, and with a Gram-Schmidt algorithm derive a reference subspace Ω_i;
3. Project Π onto *CS* and generate an $s_{no}\times t$ projection matrix Q;
4. Normalize Q, and with a Gram-Schmidt algorithm derive the input subspace Λ;

Compute the similarity between Λ and Ω_i by using the canonical angle computation described in Sec. 3.1.

3.4 Matching in LDA-transformed Space

Besides matching by CMSM with sequential images, there is another matching based on a single image. The image smoothed by the AST algorithm is first normalized to a fixed size (such as 36×36 in this paper), then it further forms a feature vector. To speed up the feature matching and obtain a better recognition accuracy, the Linear Discrimination Analysis (LDA) algorithm is adopted which computes a linear transformation W^* by maximizing the following criterion:

$$W^* = \arg\max_{W} \frac{\det(W^T S_b W)}{\det(W^T S_w W)}$$

where S_b is the between-class scatter matrix and S_w is the within-class scatter matrix. If S_w is non-singular, then in that case the ratio of $S_w^{-1}S_b$ is maximized and W^* is computed. Whenever W^* is decided, the original feature vectors are then multiplied with the transpose of W^* to generate the projection coefficients which are further used to form the transformed feature vectors with a much small feature dimension. Let $\hat{I}$ be the transformed input feature vector and $\hat{R}$ be a transformed reference feature vector, *d* be the transformed feature dimension, then $sim_{distance}$ be the distance between $\hat{I}$ and $\hat{R}$ which is defined as

$$sim_{distance} = \sum_{i=1}^{d} (\hat{I}_i - \hat{R}_i)^2$$

In our proposed recognition method, there are two matching modules, and the final decision is made by combining their matching results by a weighted sum scheme. The similarity of the image-sequence matching module is calculated by the smallest canonical angle $\cos^2\theta_1$, and the similarity of the single-image matching module is calculated by Euclidean distance. Suppose sim_{angle} and $sim_{distance}$ denote the canonical angle metric and the Euclidean distance metric, respectively. The integrated value of similarity is calculated as

$$similarity = \omega_1 \times sim_{angle} + \omega_2 \times \left(1 - \frac{sim_{distance}}{\sigma}\right)$$

where ω_1 and ω_2 are the combining weights of the two matching scores, and σ is a normalized parameter.

4 Experimental Results

We used the famous Banca face database to evaluate the performance of the proposed recognition method. The Banca database contains 52 individuals and each individual has 12 image sequences that were taken in different time, at different locations and by difference cameras. Each image sequence consists of 10 face images with various facial poses and facial expressions. To simplify the problem, only 4 image sequences of each individual taken in different time but at the same locations and by the same camera are used in this experiment. Among the 4 image sequences, only one image sequence is used in the training stage, and the other three are used in the testing stage. Among the 52 individuals, the image samples of the first 12 are used to construct a constrained subspace, and the image samples of the other 40 individuals are used to generate the reference models and to evaluate the recognition performance.

According to the manually marked eye positions, face images are extracted. Each extracted face image is applied first by AST and then resized to 36x36 pixels. In the experiment, the constrained subspace was constructed with 36 training subspaces, r_{no} is set to be 9 and *t* is set to be 1000.

Form the 40 persons, we randomly selected 35 persons for training, and used all 40 persons for testing. In order to obtain unbiased investigation, we performed the face recognition experiment one hundred times. Finally, the average performance of the one hundred experiments was reported.

The experiment results are evaluated by False Rejection Rate (FRR) and False Acceptance Rate (FAR). Figure 3 shows the recognition results of the proposed method and those of CMSM and the used single-image classifier. The recognition rate of the proposed method with no rejection rate is 99.1%, and with a 7.9% false rejection rate it is 90.1% recognition rate. Figure 4 shows the performance of FAR vs. recognition rate. Let C_no denote the number of decisive recognition which is correct and belongs to the enrolled persons, and D_no denote the number of decisive recognition which belongs to the enrolled persons. Then

$$\text{Recognition rate} = \frac{\text{C\_no}}{\text{D\_no}} \times 100\%.$$

The experimental result shows clearly that the proposed method is superior to the other two methods.

5 Conclusions

This paper introduces a face recognition method by integrating both single-image and image-sequences matching modules. To diminish the lighting effect, an Anisotropic Smoothing Transform is proposed. Experiments have shown that the proposed method can achieve a very promising recognition accuracy (99.1%) for the famous Banca face database. In the future, we intend to apply the Generalized Discriminant Analysis (GDA) [11] to the single-image recognition and further investigate the recognition performance on some larger face databases.

Acknowledgement

This research is supported in by Chung Hua University under contract CHU-96-E-12 and by Taiwan NSC under contract NSC972221-E-216-040.

References

[1] Matthew A. Turk and Alex P. Pentland,"Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] Lindsay I Smith, "A tutorial on Principal Components Analysis," http://www.cs.otago.ac.nz/student.tutorials/principal_component.pdf.

[3] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tat Ko, Ja-Chen Lin, and Gwo-Jong Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, pp.1713–1726, 2000.

[4] Hua Yu and Jie Yang, "A direct lda algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.

[5] O. Yamaquchi,K. Fukui and K.maeda, "Face Recognition using Temporal Image Sequence," *Pro, Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 318-323, 1998.

[6] K. Fukui, and O. Yamaquchi, "Face Recognition using Multi-viewpoint Patterns for Robot Vision," *Symp. of Robotics Reseach*, 2003.

[7] Masashi Nishiyama, Osamu Yamaguchi, and Kazuhiro Fukui, "Face Recognition with the Multiple Constrained Mutual Subspace Method," *Audio- and Video-Based Biometric Person Authentication*, 5th International Conference, 2005.

[8] Kazuhiro Fukui1, Björn Stenger, and Osamu Yamaguchi, "A framework for 3D object recognition using the kernel constrained mutual subspace method," *Asian Conference on Computer Vision (ACCV)*, 2006.

[9] Michael Oren and Shree K. Nayar, "Generalization of the Lambertian model and implications for machine vision," *International Journal of Computer Vision*, Volume 14 , Issue 3,pp. 227-251,1995

[10] LinLin Shen and Li Bai, "Gabor Wavelets and Kernel Direct Discriminant Analysis for Face Recognition," *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*.

[11] Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach", *Neural Computation*, Vol. 12, No. 10, pp. 2385-2404.

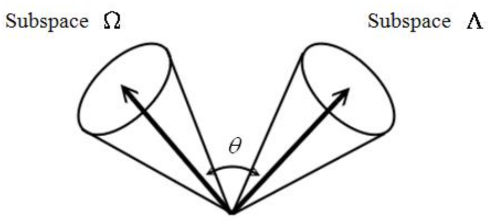

Figure 2. Canonical angle between two subspaces.

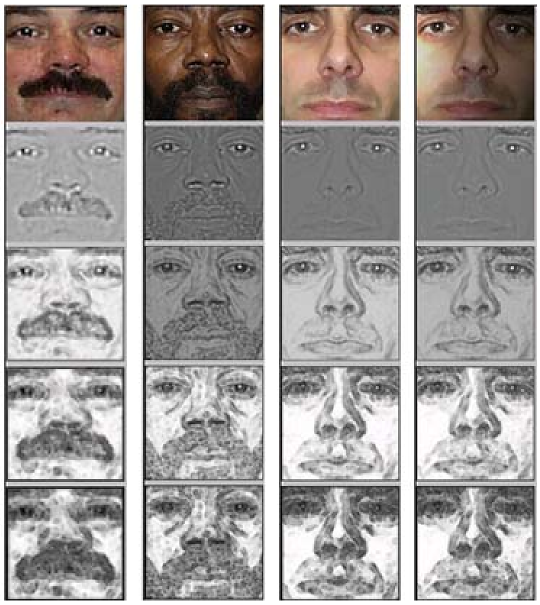

Figure 1. The Anisotropic Smoothing result: the first row is the original face image I, the second row is the ratio images N, the third row is progressed by the AST once N^1, the fourth row is progressed by the AST twice N^2, and the fifth row is progressed by AST thrice N^3.

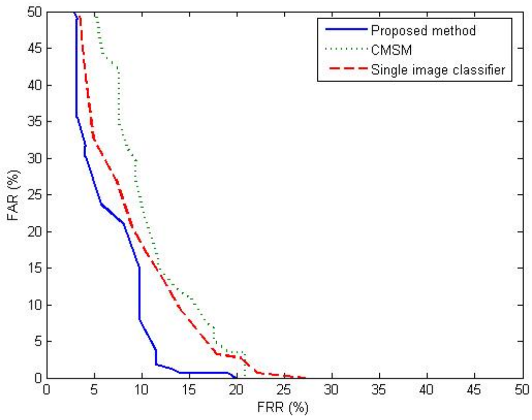

Figure 3. Performance.

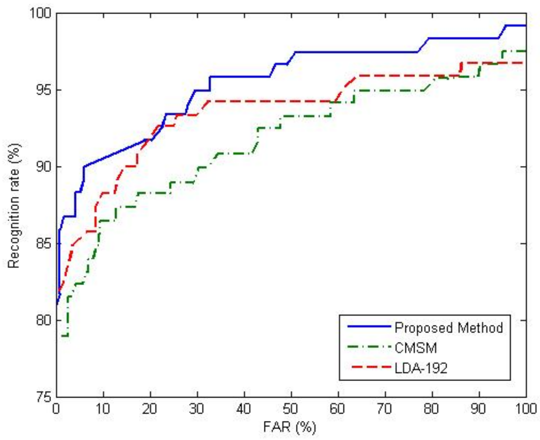

Figure 4. FAR vs. Recognition rate.