

Subspace Learning for Face Verification

Ching-Ting Tu, Meng-Ying Lin, and Shih-Hsun Hsiao

Department of Computer Science and Information Engineering, Tamkang University, Taiwan
*cttu@mail.tku.edu.tw, virds1101@gmail.com, sssoo01@gmail.com

Abstract

Face verification has been widely studied due to its importance in surveillance and forensics applications. In this paper, our goal is to verify identity of a pairwise sample, where each sample contains two heterogeneous facial images captured under different scenarios (e.g., low-resolution vs. high-resolution, or occluded facial image vs. non-occluded one). In order to address the appearance difference between these two images, an example-based verification framework is proposed. We adopt principal component analysis (PCA) approach to learn subspaces in order to capture the appearance correlation between pairwise images. Furthermore, a feature selection process is included to select eigenvectors that maximize the identity discrimination between two images. Experiments on variety face verification tasks demonstrate the effectiveness of our proposed learning framework.

Keywords: Eigenface, Face verification, Subspace learning.

Introduction

Face verification task is widely studied for surveillance and forensics applications. However, the verification performance is decreased when only low-quality images are available. For instance, police officers would like to identify suspect shown in store surveillance videos, where the qualities of probe images are limited, e.g. with low-resolution, pose or lighting variations. Furthermore, suspects usually wear sunglasses or mask to hide his or her face.

Face verification (or recognition) system for images taken under low-quality cases is still a challenge issue in computer vision. Those unfavorable conditions make the system performance unstable, because some significant facial features are missing. Generally speaking, there are two types of approach for solving the above problems. One is the local patch-based approaches and the other is synthesis-based approach. Example approaches of local patch-based approaches are methods proposed in studies [1, 7]. In these studies, the feature representation of each image is extracted from local regions. Subsequently, the low-quality images and the high-quality images are matched by the similarity of their local features. However, the local regions are with less personal characteristics. Examples of synthesis-based approaches are [2, 3, 4, 5, 8], where the appearance correlation between different local image regions are learned by training examples. Thus, more reliable image appearances are provided for verification or recognition tasks.

In this study, we proposed a framework to derive combined subspaces for modeling cross-domain data, i.e. the low-resolution vs. high-resolution and the occluded region vs. non-occluded region of image. Following, we proposed a feature selection framework to extract significant combined

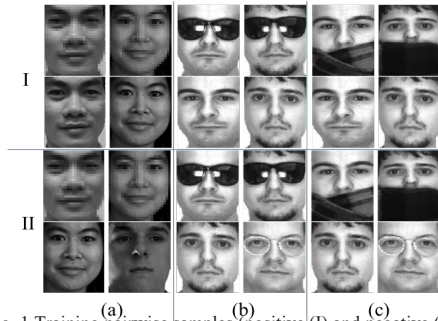


Fig. 1 Training pairwise samples (positive (I) and negative (II)) for three different low-quality scenarios. (a) low-resolution case; (b) sunglass occlusion case; (c) scarf occlusion case.

eigenfaces, where subject identity discrimination is maximized.

Face Verification System

A. Combined Subspace Construction

Assume that the training dataset consists of N different individuals, where each individual has two images from two different but related domains. Denote the dataset $X = [x_1, \dots, x_N]$ as the source domain (X -domain), and the dataset $Y = [y_1, \dots, y_N]$ as the target domain (Y -domain). In the proposed framework, images belong to X -domain are low-quality images, i.e. low-resolution image or occluded facial images. On the other hand, Y -domain images are high-quality images.

The training dataset consists of two sets of pairwise examples, namely a positive set and a negative set. Accordingly, we have $N_p = N$ positive samples, i.e. each pairwise sample of the same individual; and there are $N_n = N \times (N - 1)$ negative samples, i.e. the pairwise images of each sample are from different individuals. In order to preserve the strong correlation between these two domains, two images of each sample, i.e. x and y , are directly connected as: $A = [x^T \ y^T]^T$, where each image (with size $w \times h$) is represented as a random vector consists of wh random variables. Accordingly, the size of A is $2wh \times 1$. Figure 1 presents some typical positive and negative training samples for three different low-quality scenarios.

Following, two subspaces are constructed, namely the subspaces U_p and U_n , which are constructed by positive and negative samples, respectively. In this study, principal component analysis (PCA) approach is used for subspace construction process. Each subspace (Eigenspaces) contains two parts corresponding to the source and target domain, i.e. $U_p = [U_{p,X}^T \ U_{p,Y}^T]^T$ and $U_n = [U_{n,X}^T \ U_{n,Y}^T]^T$ (both $\in \mathbb{R}^{2wh \times k}$). In the Eigenfaces method, the subspaces can be used to

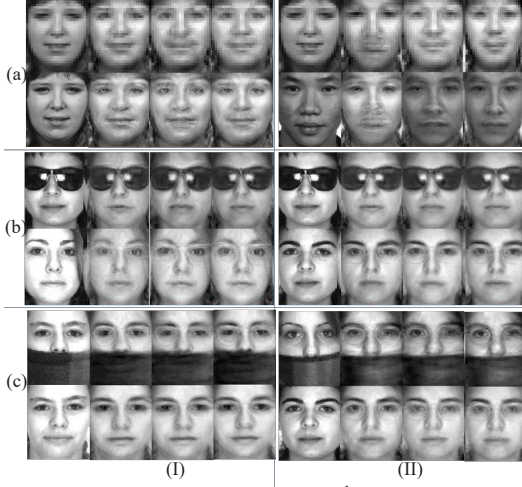


Fig. 2 Examples of pairwise samples $\hat{A}$ (positive (I) and negative (II)), and their three augmented (reconstructed) samples, $\hat{A}_p$ (2nd column), $\hat{A}_N$ (3rd column), and $\hat{A}_F$ (final column). (a) low-resolution case; (b) sunglass occlusion case; (c) scarf occlusion case.

reconstruct a facial image. Similarly, U_p and U_N can be used to reconstruct a pairwise sample, denoted as $\hat{A}_p = [\hat{x}_p^T \hat{y}_p^T]^T$ and $\hat{A}_N = [\hat{x}_N^T \hat{y}_N^T]^T$. Some examples of reconstructed results are shown in Fig. 2; as shown in this figure, these two image appearances of $\hat{A}_p$ can be identified as the same individual; on the other hand, two image appearances of $\hat{A}_N$ are from different.

Here, the similarity between A and the $\hat{A}_p$ provides a way to measure if A well modeled by the subspace U_p (similar for the role of the similarity between A and $\hat{A}_N$). The higher value indicates that the sample A is similar to the training data that constructing such subspace. On the contrast, the lower value means that A has different properties from the data samples in that subspace. In this study, the cosine-similarity is applied, i.e.

$$C(A, \hat{A}) = \frac{A \cdot \hat{A}}{\|A\| \|\hat{A}\|}, \quad (1)$$

where C function is used to measure the similarity between A and $\hat{A}$.

As shown in Fig. 2, the reconstructed appearances of positive and negative samples are different in these two subspaces. Accordingly, we design a feature formulation $f(A) = C(A, \hat{A}_p) - C(A, \hat{A}_N)$ for discriminating the reconstruction properties of A in subspaces U_p and U_N . Base on such feature representation, a simple threshold can be calculated for binary classification problem:

$$\theta = \frac{1}{2} \left(\frac{1}{N_p} \sum_{i \in \mathcal{I}_p} f(A_i) + \frac{1}{N_N} \sum_{i \in \mathcal{I}_N} f(A_i) \right), \quad (2)$$

where $\ell_i \in \{+1, -1\}$ is the sample label. As shown in Fig.3, such feature value is good enough for classifying training positive and negative samples. However, it tends to overfitting when the testing sample varied from the training samples.

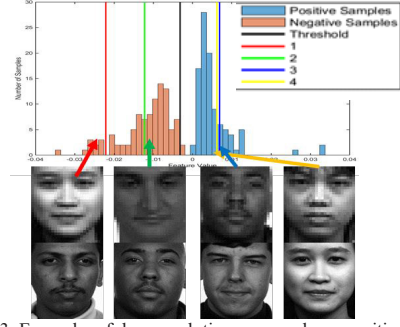


Fig. 3 Example of low-resolution case, where positive and negative data distribution of the training set are correctly classified by the threshold (black line; θ). However, given four examples of testing negative samples, the 3rd and the 4th samples are misclassified.

TABLE I FEATURE SELECTION PROCEDURE

Input: Positive and negative training image pairs $\{A_i, \ell_i, w_i\}_{i=1}^{N_p+N_N}$
$\ell_i \in \{+1, -1\}$: Labels (+1: positive, -1: negative)
w_i : Sample weight
Selected eigenvectors: $U_F = \{\}$
Initialization:
Positive (for $\ell_i = +1$): $w_i = 1 / 2 * N_p$
Negative (for $\ell_i = -1$): $w_i = 1 / 2 * N_N$
For $t = 1 : T$ do:
(i) Normalize all weights: $w_i = w_i / \sum_{i=1}^{N_p+N_N} w_i$
(ii) Weak classifier creation:
- For each feature, train classifier h_j based (3)
- Select the optimal feature with the smallest classifier error based on (4)
- The optimal feature u_j is added into U_F
(iii) Example weight updating: For each sample
$w_i = w_i \beta_i^{1- h_j(A_i)-\ell_i }$, where $\beta_i = \epsilon_i^* / (1 - \epsilon_i^*)$
End

In following sections, the objective of proposed framework is to integrate the properties of these two subspaces U_p and U_N in the face verification framework. Especially, for the case that image qualities of pairwise sample are inconsistent, and also to solve the overfitting problem.

B. Discriminative and Reconstructive Feature Selection

To eliminate the image quality differences between the source and target domains and to maximize the discrimination between different individuals, we use Adaboost algorithm [9] to select significant features from the subspaces U_p and U_N . The algorithm is shown in TABLE I.

The concept of AdaBoost algorithm is that the linear combination of a number of “selected weak classifiers” provides a “strong classifier” for the binary classification problem. Details of the algorithm are: T features are selected over T iterations. For each iteration, the most significant feature of the subspaces U_p and U_N is searched exhaustively. To select discriminative and reconstructive features, the j -th feature value of sample A_i is defined as:

$$f_j(A_i) = \begin{cases} C(A_p, \hat{A}_{i,p(j)}) - C(A_p, \hat{A}_N) & , u_j \in U_p \\ C(A_p, \hat{A}_{i,p}) - C(A_p, \hat{A}_{i,N(j)}) & , u_j \in U_N \end{cases} \quad (3)$$

where $\hat{A}_{i,p(j)}$ and $\hat{A}_{i,N(j)}$ are reconstructed results of sample A_i only based on one eigenvector u_j from U_p and U_N , respectively. To be specified, the objective of (3) is to find a eigenvector (feature of Adaboost selection process), u_j , that the similarity between A_i and the synthesized result of A_i based on u_j , i.e. $\hat{A}_{i,p(j)}$ or $\hat{A}_{i,N(j)}$, is extremely different from the similarity between A_i and its synthesized result based on the opposite subspaces of u_j , i.e. $\hat{A}_N$ or $\hat{A}_p$ in U_p and U_N , respectively. In other word, the feature value defined in (3) is to evaluate if the eigenvector u_j is good for modeling the sample than the ability of opposite subspaces.

Accordingly, for each feature representation, a binary classifier is trained according to the weighted positive and negative samples:

$$\varepsilon_i^* = \min_j \sum_{i=1} (w_i |h_j(A_i) - \ell_i|) \quad (4)$$

where a weak classifier $h_j(A_i)$ is a simple threshold function of A_i consisting one optimal polarity and one optimal threshold based on the feature value defined in (3). Subsequently, the selected feature based on (4) is the eigenvector, where has the modeling abilities between these two subspaces are significantly different for positive and negative samples. That is, the feature with the larger marginal feature values between the positive and negative samples (i.e., the features with minimum classification errors) is chosen. Following, the weights of the misclassified training samples are then increased while the weights of the correctly classified samples are decreased. In this feature selection framework, T eigenvectors are selected that either from U_p or from U_N . Singular value decomposition (SVD) is then applied for orthogonal decomposition. Similar to the properties of U_p and U_N , U_F can be used to reconstruct a pairwise sample A , denoted as $\hat{A}_p = [\hat{x}_p^T \quad \hat{y}_p^T]^T$. Examples of $\hat{A}_p$ are shown in Fig. 2, the last column.

C. Pairwise Sample Verification

Consequently, we have three different subspaces, i.e. U_F , U_p , and U_N , thus for each testing pairwise sample, A , three augmented samples are generated, i.e. $\hat{A}_F$, $\hat{A}_p$, and $\hat{A}_N$. Here, these provide a way to measure if A well modeled by these subspaces, i.e. by using the value of C function in (1) as the measurement metric. To solve the overfitting problem of the prior threshold θ defined in (2), we included three additional adaptive thresholds in the proposed verification framework; Especially, these thresholds are adaptively calculated according to the testing input sample. In order to calculate these three thresholds, the DCM transformation approach [6] is introduced. The DCM approach is a regression method that based on the learnt subspace. In this study, for these three subspace, we respectively train three DCM transformations to predict the target domain image, y , from the image at source-domain x . The formulation of DCM transformation is:

$$\hat{y}^{DCM} = U_y \left(U_x^T U_x + \lambda \times \text{diag} \left(\Sigma^{-\frac{1}{2}} \right) \right)^{-1} U_x^T (x - \bar{x}) + \bar{y} \quad (5)$$

where Σ is the diagonal matrix; where the entries of Σ on the



Fig. 4 Examples of source images x , their predicted target images $\hat{y}_p^{DCM}$, $\hat{y}_N^{DCM}$, and $\hat{y}_F^{DCM}$, and the corresponding ground truth y (the last column).

diagonal are given by the eigenvalues corresponded to subspace U . Subsequently, for the subspaces U_p , U_N , and U_F , we have three predicted versions of y , namely $\hat{y}_p^{DCM}$, $\hat{y}_N^{DCM}$, and $\hat{y}_F^{DCM}$, respectively (as shown in Fig. 4.). Ideally, the $\hat{y}_p^{DCM}$ can be identified as the same individual of input x ; the $\hat{y}_N^{DCM}$ and x are identified as two different subjects; on the other hand, the $\hat{y}_F^{DCM}$ is neutral.

We adopt these synthesized results of the input x as the adaptive thresholds for face verification. To be more specified, given a testing image x , we have three predicted target images, namely, $\hat{y}_p^{DCM}$, $\hat{y}_N^{DCM}$, and $\hat{y}_F^{DCM}$. The input x is concatenated with these synthesized $\hat{y}_p^{DCM}$, $\hat{y}_N^{DCM}$, and $\hat{y}_F^{DCM}$, respectively, to form three pairwise samples. For each pairwise sample, the feature value, which is calculated based on by the formulation, $f(A) = C(A, \hat{A}_p) - C(A, \hat{A}_N)$, is then recorded as the thresholds, θ_p , θ_N , and θ_F . Here, these thresholds are calculated as the formulation in (2).

The feature values of the testing pairwise sample A and its three augmented samples, $\hat{A}_F$, $\hat{A}_p$, and $\hat{A}_N$, i.e. $(f(A), f(\hat{A}_p), f(\hat{A}_N), f(\hat{A}_F))$ are then compared with the four thresholds, $(\theta, \theta_p, \theta_N, \theta_F)$, respectively. The comparison result is then record as either 0 (less than threshold) or 1 (others). Thus, for each sample A , a novel 16-coding feature vector representation is proposed, where four types of threshold and four types of feature representation are embedded. Finally, we use 16-coding vector as sample feature for support vector machines (SVM) classifier creation, which is then applied to make the decision of face verification result.

Experimental Results

The performance of the proposed verification framework was evaluated by performing a series of experimental trials based on two databases, AR database [11] and FERET database [10]. Three conditions are addressed, namely, the low-resolution input image, the input image occluded by either sunglasses or scarf. The FERET images are used for the low-resolution input image case, where there are 130 different subjects, 100 subjects are randomly selected for the training purpose, and 30 for the testing purpose. The image size of target domain is 100×120 , and the source domain image is the down-sampled result by factor 4. For the occlusion cases, images are from AR database, where 111 different subjects are used, where 90 pairwise samples for training purpose, and 21 testing subjects. For two occlusion cases, the source image is the occluded image, while the target domain image is

TABLE II PERFORMANCE COMPARISONS:
(I) low resolution case; (II) sunglass occlusion case; (c) scarf occlusion case. P: the positive sample; N: negative sample.

(I)				
Dataset Feature	Training		Testing	
	P	N	P	N
F1	1.00	1.00	1.00	0.73
F2	0.71	0.85	0.60	0.40
F3	1.00	0.94	1.00	0.70
F4	1.00	0.94	1.00	0.83
F1~F4	1.00	1.00	0.97	0.87
(II)				
Dataset Feature	Training		Testing	
	P	N	P	N
F1	1.00	0.90	0.95	0.24
F2	0.73	0.70	0.43	0.43
F3	0.99	0.52	0.95	0.33
F4	0.88	0.74	0.62	0.52
F1~F4	0.98	0.97	0.62	0.81
(III)				
Dataset Feature	Training		Testing	
	P	N	P	N
F1	0.97	0.88	0.71	0.48
F2	0.63	0.63	0.48	0.48
F3	0.60	0.80	0.52	0.67
F4	0.74	0.77	0.43	0.57
F1~F4	0.97	0.83	0.67	0.52

non-occluded one.

The SVM verification performances that based on various feature representations are provided in Table II, where the 4-bit binary code vectors of the first (F1: $f(A)$)-value is compared with the proposed four thresholds), second (F2: based on $f(\hat{A}_p)$), third type (F3: based on $f(\hat{A}_n)$), and fourth type (F4: based on $f(\hat{A}_r)$) of feature representation are also included for comparison. To be more specified, for each 4-bit vector representation, one corresponding SVM classifier is trained. The proposed feature, 16-bit binary code vector, improves the performances of other four SVM classifiers in FERET database and AR database for scarf cases.

Conclusions

This study has presented a face verification framework for images with the low-quality case. To be specified, we presented a unified model for jointly eliminating appearance variation elimination and maximizing personal discrimination. Especially, in this study, varied data representations are included to avoid the overfitting situation, and to increase the generalization of the proposed verification system. Overall, our results show that the proposed framework is a promising way to improve the performance of existing automatic face verification system especially for the real cases, that the appearance of input image is varied from those in the database.

References

- [1] L.A. Cament, L.E. Castillo, J.P. Perez, F.J. Galdames, and C.A.

- Perez, "Fusion of Local Normalization and Gabor Entropy Weighted Features for Face Identification," *Pattern Recognition*, pp. 568–577, 2014.
- [2] B.W. Hwang and S. W. Lee, "Reconstruction of Partially Damaged Face Images Based on A Morphable Face Model," *IEEE Trans. Pattern Anal. Mach. Intell.* pp. 365–372, 2003.
- [3] D. Lin, and X. Tang, "Quality-Driven Face Occlusion Detection and Recovery," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–7, 2007.
- [4] J.S. Park, Y.H. Oh, S.C. Ahn, and S.W. Lee, "Glasses Removal from Facial Image Using Recursive Error Compensation," *IEEE Trans. Pattern Anal. Mach. Intell.* pp. 805–811, 2005.
- [5] C.T. Tu and J.J. Lien, "Non-parametric Bayesian Alignment and Recovery of Occluded Face using Direct Combined Model," *International Conference on Computer Vision Theory and Applications*, pp. 495–498, 2010.
- [6] C.T. Tu and J.J. Lien, "Automatic Location of Facial Feature Points and Synthesis of Facial Sketches Using Direct Combined Model," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, pp. 1158–1169, 2010.
- [7] B. Yao, H. Ai, and S. Lao, "Person-Specific Face Recognition in Unconstrained Environments: A Combination of Offline and Online Learning," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 1–8, 2009.
- [8] D. Yu and T. Sim, "Using Targeted Statistics for Face Regeneration," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–8, 2008.
- [9] P. Viola, M. J. Jones, "Robust Real-Time Face Detection," *International Journal of computer Vision*, 57(2), pp. 137–154, 2004.
- [10] P. Philips, H. Moon, P. Pauss, and S. Rivzvi, "The Feret Evaluation Methodology for Face-recognition Algorithms," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 137–143, 1997.
- [11] A. M. Martinez and R. Benavente, "The AR Face Database," *CVC Technical Report #24*, June 1998.