Feature Reconstruction for Face Recognition Based on Sample Image Learning

Hongzhou Zhang, Yongping Li, Lin Wang, Chengbo Wang

Abstract—Pose problem is a big challenge for applying face recognition technology under real world conditions. In this paper, appearance based approach was proposed to recognize face across front and non-frontal view images by reconstructing frontal view features. Statistical learning method based on sample images is applied to find transformation matrix which encapsulated general knowledge of pose transition in feature subspace, therefore, different view feature vectors constituted linear equations and transformation matrix can be solved from the equations by least square (LS) approach. Experimental results on popular FERET and CMU databases showed that the proposed method could cope with the head rotation roughly within half profile view. Compared with model based approaches, this method is not dependent on heavy computation and has merit of easy implementing in live conditions.

***Index Terms*—face recognition, feature reconstruction, statistical learning, subspace transformation, pose problem.**

I. Introduction

Face recognition, an effective biometric method, has diverse applications especially as an identification solution which can meet the high demanding needs in security areas. Considerable achievements of face recognition have been attained in recent years [1]. However, there are main issues that are far from being solved, such as pose variation [1], [2].

Many algorithms were developed to overcome pose effect. They can be divided into two main strategies: model-based technology and appearance-based technology. Model-based works [3]-[8], especially 3D face model [3], is effective for posed face recognition, but fitting face model to an inputted image is time-consuming. Affine transformation [6] can be used to reduce computation in a great extent, but it still relies on sets of feature points to align image with the standard view models which also is a hard task [1], [6]. Because of above reasons, there is a long way to apply model-based approaches under real world condition.

Manuscript received March 9, 2007. This work is supported by Hundreds Talent Program of CAS and Pujiang Program under the contract No.05PJ14111.

Hongzhou Zhang is with Shanghai Institute of Applied Physics, Chinese Academy of Sciences (CAS), Shanghai 201800, P. R. China.(email: hongzhouzhang@sinap.ac.cn);

Yongping Li s with Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, P. R. China.(email: ypli@sinap.ac.cn, Tel: +86-21-59554676, Fax: +86-21-59552037).

Techniques based on statistical properties of face images, or appearance-based works, are successful for frontal view face recognition, such as Eigenface [9] and Fisherface [10]. With no time-consuming model fitting and too many fiducial feature points, these algorithms are more suitable for applications under live conditions [1], [19]. Appearance-based work tries to overcome pose problem by enrolling images at additional views into system. These are multi-view subspace approach [11] and parametric subspace approach [12], [13]. Traditional face recognition system only enrolls front view face image. For example, face recognition is strongly recommended in MRTD application [1] whereas only frontal face image can be stored in a smart card due to very limited storage space; In access control or video surveillance system, training images are mainly taken from mugshot. In above scenarios, pose of probe image is usually uncontrollable in real applications, and sometime the subject only shows posed head to the system deliberately. In other words, enrolled image and probe image always belong to different views and face recognition is an across pose task so that the current appearance-based approaches could not achieve satisfied results these applications.

In this paper, general information on pose transformation was learned from sample images to improve appearance-based face recognition by constructing front view features from non-frontal view features. To focus our attention on feature reconstruction, we suppose that pose of every training image has been given out by one kind of pose estimator that could be found in [14], [15]. Experiments show that re-constructed feature is equal to face recognition across poses within large pose rotations.

The rest of paper is organized as follows. In Section 2, we explain how to learn the transformation matrix from sample images and reconstruct frontal view features. Different feature extractions are then introduced in Section 3. The experimental results are reported in the followed section. Discussions are given in Section 5 finally.

II. Linear Learning form Sample Images

Human being has capability to associate stranger's photos with different poses together. The fact implies that correlation between views is helpful to improving posed face image recognition.

Traditional subspace recognition, such as Eigenface [13] or Fisherface [15], is actually view-dependent due to images at the single (frontal) view are used to train the representations. When a posed image is represented by this frontal view subspace, wrong image representing will lead to system's failure. By introducing additional view subspaces, Pentland *et al*. avoided the wrong representation and performed recognition within the same or nearly the same view subspace [11]. Compared with face recognition using a unique subspace or parametric subspace, multi-view subspaces are more efficient for image representation and recognition [11], [15]. However, recognition based on this multi-view subspaces is just a duplication of conventional frontal view recognition, it is not suitable to recognize faces across various poses because correlations between views was discarded.

In this work, similar multi-view subspaces are employed for image representation. In order to recognize faces across different poses, learning the correlation between different view subspaces is the key issue. Because only frontal view images are used for enrollment process, we must have the sample images at various views as the separate training data to learn correlation between different views. Such sample images can be collected quite freely by imaging a group of people under the same or nearly the same image acquisition conditions as that for the enrolled frontal-view image.

In such case, subjects' one frontal view and one side view image comprised image pairs; they were used to train two view-dependent subspaces respectively. Images in the same pair were taken as simultaneously as possible. $V_F = \{v_i^F | \text{ i} = 1, 2, \ldots, N\}$ and $V_P = \{v_i^P | \text{ i} = 1,2, \ldots, N\}$ are projections of frontal and side view images in their corresponding view subspaces. Superscript "*i*" denotes the projection of images in the *i*-th pair, *N* is total number of image pairs. Let's suppose that frontal view vectors V_F could be reconstructed from side view vectors V_P by a transformation *T*:

$$V_F = T(V_P) \tag{1}$$

The above equation defined *feature transformation* in subspace. Generally speaking, the transformation denoted by (1) has non-linear portion. A linear transformation is more preferred because linear transformation is convenient for problem description and well researched by scientists in different areas. Fortunately however, many research outputs proved that the linear transformation could ensure satisfying results in image synthesis and recognition. Lanitis et al. [16] have showed that linear model is sufficient to simulate considerable pose variation as long as overlap does not seriously take place. Based on the theory of linear class [17], prior information on view transformation was learned from example images at different views and reconstructed frontal view information achieved satisfying face recognition result in 2D model-based work [5].

For linear feature transformation, matrix *W* connects frontal view and side view feature vectors together:

$$(v_F^1, v_F^2, \ldots, v_F^N) = W \cdot (v_P^1, v_P^2, \ldots, v_P^N) \tag{2}$$

$$V_P{}^T w_k = v_k{}^F \tag{3}$$

where w_K and $v^F{}_K$ are the *k*-th column vector of W^T and $V_F{}^T$ respectively. $(\cdot)^T$ denotes matrix transpose. We call W *transformation matrix*, which can be resolved by means of linear algebra.

The solution of (3) depends on the property of its coefficients matrix, a detail of derivation could be found in [18]. Because V_F and V_P belong to two independent subspaces respectively, generally $\text{rank}(V_P{}^T \mid V_F{}^k) \neq \text{rank}(V_P{}^T)$. $V_P{}^T$ is *N*-by-*m* matrix, *N* is number of image pairs in generic training set and *m* is of dimensionality selected subspace. For face recognition, there is $N > m$. So, (2) is inconsistent and over-determined system of linear equations which are exactly unsolvable. According to theory of matrix, inconsistent system has approximation solutions under 2-norm constraint, which is called least square (LS) solution. LS solution minimizes square error:

$$w_k = \underset{w \in R^m}{\arg\min} \| V_P^T w - V_F^T \| \tag{4}$$

where $\|\cdot\|$ denotes Euclidean norm. Among least square solutions, a particular optimum approximation solution is:

$$W^T = (V_P^T)^+ \cdot (V_F^T) \tag{5}$$

$(V_P{}^T)^+$ is the Moore-Penrose pseudoinverse of $V_P{}^T$ which could be calculated by means of singular value decomposition (SVD).

Once (2) is resolved, a probe feature vector in side view subspace, v_{test}, is mapped into frontal view subspace by the following (6). Recognition could be performed in frontal view subspace using reconstructed feature v'_{test}.

$$v'_{test} = W \cdot v_{test} \tag{6}$$

The scheme of our transformation work can be illustrated in Fig. 1.

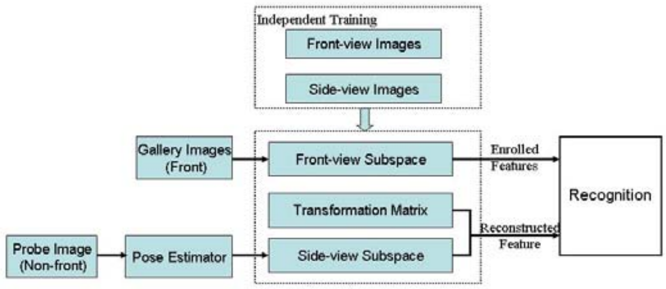

Fig. 1 flow chart of the proposed approach.

III. Feature Extraction

In face recognition, we generally compact the original data from a high-dimensional image space into a considerably low dimensional feature subspace. This procedure extracts features from original image by projecting the face vector *Y* to the subspace's basis vectors, in (7). The projection coefficients X are used as the feature representation of each face image and then perform recognition in the feature subspace.

$$X = U^T Y \tag{7}$$

Different feature extractions use different criterions in finding projection matrix U (comprised of basis vectors).

Accordingly, they are divided into three main categories: reconstruction-based methods, discrimination-based methods and factor-based methods [19]. In this work, three typical feature extractions of those methods, i.e., PCA (Principal Components Analysis), LDA (Linear Discriminant Analysis), and ICA (Independent Components Analysis) were applied.

PCA, also known as Eigenface [9], minimizes reconstruction error of training images, the projection matrix U_{pca} is chosen by:

$$U_{pca} = \arg\max \left| U^T S_t U \right| \tag{8}$$

So PCA is a reconstruction-based method.

Different with PCA, the optimal projection matrix of LDA (or Fisherface) [10] is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix S_b of the projected samples to the determinant of the within-class scatter matrix S_w of the projected samples:

$$U_{pca} = \arg\max \left| \frac{U^T S_b U}{U^T S_w U} \right| \tag{9}$$

However, due to “small sample size (S3)” problem in face recognition, S_w is always singular and recognition score will be deteriorated. Revised LDA algorithm, e.g., Regularized-LDA [20] in contrast with the traditional LDA in present work, is developed to overcome the S3 problem using the following criterion:

$$U_{pca} = \arg\max \left| \frac{U^T S_b U}{U^T (\eta S_b + S_w) U} \right| \tag{10}$$

where η is regularization parameter. Obviously, LDA-based approaches take discrimination into account and are discrimination-based methods.

ICA is a factor-based approach and is derived from blind sources separation. Its components (basis vectors) are designed to be statistically independent. ICA separates the high-order moments of the input in addition to the second-order moments utilized in PCA. We executed ICA using fix-point fast ICA calculation algorithm [21] on Bartlett’s [22] “Architecture One”, where images, not pixels, are treated as independent random variables.

IV. Experimental Results

A. Face databases

Two popular databases were used to evaluate the proposed approach: CMU-P.I.E face database and FERET database. In the pose subset of CMU database [23], 68 subjects were imaged simultaneously under different poses with three expressions, in other words, each subject has three samples at each view. FERET database [24] is another standard database in this area. Each of 200 subjects has one image per view in subset “b-”of FERET. Images in right-side views of FERET were mirrored to their corresponding left-side views (frontal view images were same mirrored), because LDA based subspace requires two simples each class at least for training the representation. Thus there were four side-views remained, but two images for one subject were available under each views. Image under two full profile views in both databases were excluded because the main part of frontal feature is invisible.

Each database was divided into generic training set and testing set according to subjects’ identity. Image representation and transformation matrixes were learned from generic training set. Testing set was further divided: all frontal view images were used as gallery and non-frontal images were used as probe set. The inter-oculars distance was set identical before all images were aligned according to their eyes coordinates. Histogram equalization was applied to reduce illumination effect (see Fig. 2 for a sample). The classification is performed by using the nearest neighbor classifier. Euclidean distance (ED) was the similarity for PCA, ICA feature, Angles or normalized distance (AD) was similarity for two LDA-based features.

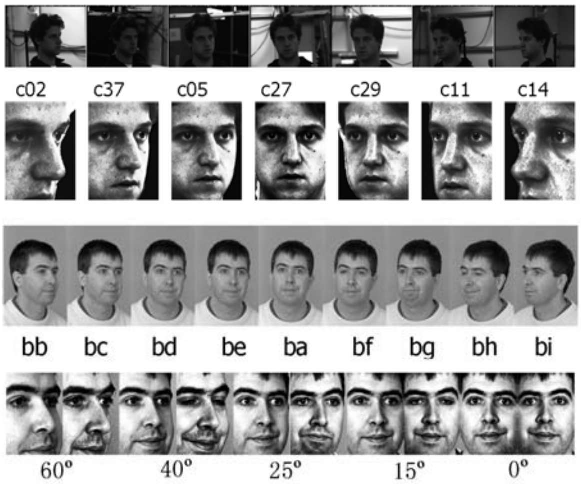

Fig. 2 Samples of CMU-PIE (above) FERET (low) database. The Each first row is snapshots of original images in database; the second row is normalized images in size of 131-by-181.

B. Face recognition across poses

Leave-one-out experiments on CMU database were executed using 34 among all 68 persons’ image as generic training set, and the rest were used as gallery and probe set. The average transformed recognition rate from different side-views to the front view is list in Table 1. Similarly, 200 persons in FERET database were also divided in to two equal size portions, one half was used as training set and the other half is test set. Experiment was also “leave-one-out” test by dividing database into 40 smaller parts and the average performance of transformed recognition were given in Table 2.

It shows that feature transformation improved recognition rate in a great extent under all test views. For example, though there was only 15 degree apart, the direct recognition rate was rather low on “be” view in FERET database. But recognition using reconstructed feature was high especially for probe images belong to near front views. Using PCA and ICA features, transformed recognition rate attained near 100% recognition rate in 22.5 degree view (C05 and C29) on CMU database; the corresponding results were also above 90% on FERET which is larger database.

Recognition rate fells as probe image turns far from front view; half profile view was a tuning-point for feature reconstruction. Recognition rate was high within 45 degree rotation and became low when pose was out of rough 45 degrees. It can be explained as the following. The phenomenon derives from the linear assumption in (1). As probe image belonged to near front views, overlapping was minor and linear transformation worked well; when pose was out of roughly 45 degrees, overlapping could not be ignored and recognition rate dropped for linear assumption was broken.

Table 1 The average recognition rate on CMU-PIE database.

		C02	C37	C05	C29	C11	C14
probe pose		-67.5	-45	-22.5	22.5	25	67.5
PCA	Dir.	2.9	5.7	7.8	21.1	10.7	4.9
	Trans.	55.0	83.9	99.1	97.3	86.7	55.1
ICA	Dir.	3.0	4.9	3.9	2.9	3.9	4.9
	Trans	54.7	84.3	99.2	97.4	86.2	56.2
LDA	Dir.	3.4	3.8	3.6	7.0	5.3	2.9
	Trans	29.6	52.4	85.1	84.6	64.2	34.4
RLDA	Dir.	3.8	3.4	4.6	3.8	3.4	3.9
	Trans	38.6	63.3	92.8	88.6	69.0	40.8

Table 2 Recognition rates on the FERET database. Numbers in brackets denote the dimension of subspace where corresponding score achieved. "Dir." means directly recognition; "trans." denotes the transformed recognition.

	Probe pose	be(15º)	bd(25º)	bc(40º)	bb(60º)
LDA	Dir.	1.0	1.2	1.3	1.3
(99)	Trans.	38.1	23.2	9.6	5.3
RLDA	Dir.	1.5	1.2	1.3	1.3
(99)	Trans.	87.0	78.1	52.0	29.3
PCA	Dir.	7.6	2.18	1.30	1.25
(199)	Trans.	90.3	83.6	63.2	35.4
ICA	Dir.	1.13	1.13	1,15	1.15
(200)	Trans.	89.1	82.2	60.1	34.1
Ref. 3		99.5	96.9	95.4	94.8
Ref. 6		77.5	55.5	N/A	N/A

Table 3 Recognition rate (RR) varying with model set size (M-Size) on "be" part of FERET. Dimension of subspaces where best score occurred is in row of "dim". AD is used as similarity measurement for LDA and RLDA; ED for PCA and ICA.

M-Size	15	20	25	50	75	100	125	150
PCA	69.1	75.3	78.4	87.2	90.7	91.9	92.7	92.7
ICA	70.15	76.2	80.23	86.6	89.7	91.0	91.7	92.2
LDA	10.7	13.1	17.3	31.3	39.1	47.5	53.8	58.6
RLDA	55.7	61.6	66.6	82.6	88.9	90.6	91.8	92.3

C. Comparison of Different Types Feature

Large numbers of training data brought more satisfactory correct recognition rate due to the statistical inherence of the proposed methods. To further explore performance of different representation trained by varying scales of learning sample, testing set kept a size of 50 persons in the following experiments on FERET database. Experimental results of a "leave-one-out" test on 'be' subset were listed in Table 3. In this experiment, performance of reconstructed features increased while the size of generic training set varied from 25 to 150; for PCA, this increasing stopped when 125 persons were used as training set. In addition, difference of RLDA and LDA or PCA feature became little when larger number of sample image was used. As 150 persons were included into training set, the difference was further reduced to within one percent. This experiment indicated that three kinds of feature are same effective if they were well trained.

V. Conclusion and Discussion

In conclusion, the proposed sample learning work can improve across poses recognition by reconstructing frontal view feature from side-views. Converting ability of Matrix *W*, reflected by correct recognition rate, decreases as face turns away from frontal view.

Comparing our work with others, two typical model based works were listed in last two rows of Table 2. It showed that 3D model [3] is powerful in all tested views because there was only five percents decreased when probe view turned from near frontal to near full profile view. The affine transformation work [6] also attained a satisfying recognition result in "be" view. Within half-profile view, the proposed approach still is comparable with others. It must be pointed out that the proposed approach is real-time one and also do not depend on too many fiducial points compared with [3], [6].

We noticed that a similar linear transformation matrix was used to generate virtual frontal view image in [25]. In present work, face recognition was directly performed on transformed features without image synthesis. Experiments show Maximum Discriminant Feature (MDF), such as R-LDA feature, has the same performance with Maximum Expression Feature (MEF), such as PCA features. It also proved that image synthesis was not necessary. Research has proved that MDF outperforms MEF in face recognition under real conditions [1, 19], so LDA-based features are more robust when the proposed method is used in an applied system.

So far, we took the assumption of linear transformation. Nonlinear effects could be alleviated using modular recognition in [11] or by dividing image in to small sub-images overlapping each other in [8]. Present work focused on performance of three kinds of feature in feature transformation and there was no contribution brought by classifier because the simple nearest neighbor classifier was used. A more sophisticated classifier will reduce noises introduced by transformation and further improve recognition rate. Moreover, we should test the proposed work in live conditions with a pose estimator used prior feature extraction. All the above will be our future work.

ACKNOWLEDGMENT

We acknowledge authors of the databases providing their database for our experiments.

REFERENCE

[1] X. Lu. "Image analysis for face recognition," http:/www.cse.msu.edu/

[2] P.Jonathon Phillips, "Face recognition vendor test 2002: overview and summary", March 2003, http://www.frvt.org.

[3] Blanz, T. Vetter. "Face recognitin based on fitting a 3d morphable model," *IEEE trans. Pattern Analysis and Machine Intelligence*, 2003, Vol. 25, no. 9, pp. 1-12.

[4] D. Beymer. "Face recognition under varying pose," Technical Report 1461, MIT AI Laboratory, 1993

[5] D. Beymer and T. Poggio. "Face recognition from one example view." A.I. Memo No. 1536, MIT AI Lab, 1995

[6] X. Chai, S. Shan, W. Gao, X. Liu. "Pose normalization using generic 3D face model as a priori for pose-Insensitive face recognition", *LNCS,* 2004,Vol. 3338, pp. 144-152, [Sino-Biometrics 2004].

[7] T. Cootes, G. Wheeler, K. Walker, and C. Taylor. "View-based active appearance models," *Image and Vision Computing*, 2002, Vol. 20, pp. 657-664.

[8] R. Gross, S. Baker, I. Matthews, Takeo Kanade. "Face recognition across pose and illumination," in *Handbook of Face Recognition*, Ed. Stan Z. Lipage Springer, 2004, pp. 193-216.

[9] Turk, M. and A. Pentland, Eigenface for recognition. *Journal of Congnitive Neuroscience*, 1991, 3(1): pp. 71-86.

[10] P. N. Belhumeur, J. P.Hespanha, D. J. Kriegman, "Eigenface vs. fisherface: recognition using class specific linear projection," *IEEE trans. Pattern Analysis and Machine Intelligence*, 1997, Vol. 19(7), pp. 711-720.

[11] Pentland, B. Moghaddam, and T. Starner. "View-based and modular eigenspaces for face recognition." In *Proc. IEEE 1994 CVPR*, 1994, pp. 84-91.

[12] H. Murase and S. Nayar. "Visual learning and recognition of 3-D objects from appearance." *International Journal of Computer Vision*, 1995,Vol. 14, pp. 5-24.

[13] Frank Y. Shih, Camel Y. Fu, Kai Zhang, "Multi-view face identification and pose estimation using B-spline interpolation", *Information Sciences*, 2005,Vol.169, pp. 189–204.

[14] Z. Zhong, J. Li, D. Zhao, and Y. Hong, Effective pose estimation from point pairs, *Image and Vision Computing*, 2005, Vol. 23, pp. 651-660.

[15] Mohamed Deriche, M.A., Pose-invariant face recognition using subspace techniques, in *Computer-aided Intelligent Recognition Techniques and Applications*, M. Sarfraz, Editor. 2005, John Wiley & Sons Ltd.: England, pp. 169-200.

[16] Lanitis, C. J. Taylor and T. F. Coots, Automatic interpretation and coding of face images using flexible models, *IEEE trans. Pattern Analysis and Machine Intelligence*, 1997, Vol. 19, no. 7, pp.743-976.

[17] Shimon Ullman and Ronen Barsi, Recognition by Linear combinations of models, *IEEE trans. Pattern Analysis and Machine Intelligence*, 1991, Vol. 13(10), pp. 992-1006.

[18] X. D. Zhang, Matrix analysis and applications, Tsinghua and Springer Publishing house, Beijing, 2004, pp. 71-100.

[19] Qiong Yang, Xiaoou Tang, Recent advances in subspace analysis for face recognition, *LNCS,* 2004,Vol. 3338, pp. 275-287, [Sino-Biometrics, 2004].

[20] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos. "Regularization studies of linear discriminate analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letter*, 2005, Vol. 26(2), pp. 181-191

[21] E. Bingham, A. Hyvarinen. "A fast fixed-point algorithm for independent component analysis", *Int. J. of Neural Systems*, 2000, Vol. 10(1), pp.1–8.

[22] M. S. Bartlett，J. R. Moellan, T. J. Sejnnowski. "Face recognition by independent component analysis." *IEEE trans. Neural Networks*, 2002, Vol. 13(6), pp. 1450-1464.

[23] T. Sim, S. Baker, and M. Bsat, The CMU pose, illumination and expression database, *IEEE Transactions on Pattern Recognition and Machine Intelligence*,2003, 25(12), pp.1615-1618.

[24] *http://www.itl.nist.gov/iad/humanid/feret/*.

[25] Hyung-Soo Lee, Daijin Kim, "Generating frontal view face image for pose invariant face recognition," *Pattern Recognition Letters*, 2006, Vol. 27, pp. 747–754.