

Inter-Subspace Distance: A New Method for Face Recognition with Multiple Samples

Jiun-Hung Chen[†], Shih-Liang Yeh[‡] and Chu-Song Chen[‡]

jhchen@cs.washington.edu, {slyeh,song}@iis.sinica.edu.tw

[†]Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA

[‡]Institute of Information Science, Academia Sinica, Taipei, Taiwan

Abstract

In this paper, we develop a systematic method that can cope with multiple images simultaneously for face recognition. The proposed method, referred to as Inter-Subspace Distance, employs the minimal distance between the two subspaces formed by training and test images, respectively. The advantages of our method are that it can use temporal information (image sequences) and multiple sampling (in scales or spatial positions) for face recognition. In addition, our method can ease the burdens of face detection by dealing with inaccuracies of positions and scales of the detected faces.

1. Introduction

To extract a face region with accurate position and scale, so that the detected face aligns well with that pre-collected, is difficult. Hence, in many real situations we need to identify a detected region that is not aligned well with the stored pattern. Since face identification usually strongly relies on successful face detection results, such a misalignment problem causes degradation of the overall performance of a face recognition system.

One possible way to overcome the misalignment problem is to employ multiple samples or observations. In this way, the detected region is treated as a coarse estimation of the pose and scale of a face, and a more appropriate region (“appropriate” in the sense that a better alignment is available) is assumed as existing in a neighborhood of the detected region. Hence, multiple samples (along with their variations such as rotation, size-scaling, and illumination-scaling) generated by re-sampling in a neighborhood of the detected region are employed for face recognition. Given such a large set of samples, how to employ them efficiently and effectively becomes an important issue and is our focus in this paper. For readers who are interested to deal with the misalignment problem in only one sample case, see [10][8] for reference.

In fact, taking multiple samples for face recognition has received much attention in recent years. Currently most ex-

isting methods employ multiple samples by grabbing a sequence of images that provides information about the temporal variability in the appearance of face. To use such a set of sample images, a straightforward idea is to match each single sample with a list of training models by a single-shot-based method, and then accumulate each single recognition result and select the identity receiving the highest voting. Such kind of simple extensions of single-shot methods has been proposed in [5]. Similarly, in [6], a weighted voting method (where each weight of voting is estimated based on the distance between a single test sample to the appearance manifold formed by the training samples) is employed for sequence-based face recognition. Although using a sequence of images increases the robustness of face recognition, this kind of approaches still suffers from the misalignment problem that may occur for each sample. In addition, performing a single-shot method multiple times also diminishes the efficiency of a recognition task.

Instead of performing a single-shot method multiple times, more efficient methods were proposed. In [11], recognition is performed by matching probability distributions formed by the training and the testing samples in the feature space, where the relative entropy or equivalently Kullback-Leibler divergence (KLD) was employed as a similarity measure of the two probability density functions. This method may be good when the testing images have been sampled in an approximately similar way as that for the training images (so that similar probability distributions formed by the multi-sample appearances can be obtained for both the training and the testing data sets). However, since the testing images are usually grabbed with a free-form sampling process, it is hard to expect that the test images form a probability distribution approximate to that formed by the training images in the feature space in practice. In [14], a mutual subspace method (MSM) was proposed, which employs the subspaces spanned by the feature vectors of the training and the test samples, respectively, for face recognition. In this approach, the minimal angle between the two subspaces serves as a similarity measure between the training and the testing data sets. Unlike [11], this approach does not require that the training and the test-

ing data sets have similar distributions in the feature space.

In this paper, we propose a method that can cope with multiple samples simultaneously for face recognition. The set of multiple samples can be obtained by varying and re-sampling in a neighborhood of the detected region, gathering multiple observations in an image sequence, or combining both. Our approach employs the minimal distance between the two subspaces formed by training and test samples, respectively, which has a close relationship with the linear-combination approaches. Note that object recognition via linear combination has received considerable attention in recent years, which was shown to be effective for dealing with illumination and pose changes. In [13], Ulman and Barsi showed that the variety of views depicting the same object under different transformations can often be expressed as the linear combination of a small number of views. In [4], Belhumeur and Kriegman have shown that a new image under all possible illumination conditions can be expressed as a linear combination of some basis images formed by a convex polyhedral cone if the illumination model is Lambertian. In our approach, linear combinations of the test samples are employed as prototypes for recognition, which exploits advantages of the linear-combination approaches introduced above.

2. Inter-Subspace Distance

Similar to the idea of evaluating the minimal angle between two subspaces [14], we try to define a distance referred to as Inter-Subspace Distance (ISD) between two subspaces formed by training and test images, respectively.

Based on this idea, given two subspaces, we estimate a pair of vectors from these subspaces respectively, such that the distance between these two estimated vectors is minimal. The details of the proposed inter-subspace distance (ISD) are in the following.

First, let $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$, and $\bar{\mathbf{x}}$ be the first M principal components, and mean vector of the subspace D , respectively. Similarly, let $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{M'}$ and $\bar{\mathbf{x}}_G$ be the first M' principal components and mean vector of the subspace G , respectively.

Given a set of M PCA coefficients $\{z_1, z_2 \dots z_M\}$, we can construct $\mathbf{d} = P\mathbf{z} + \bar{\mathbf{x}}$ where $P = [\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_M]$ and $\mathbf{z} = [z_1 z_2 \dots z_M]^t$. Similarly, given a set of M' PCA coefficients $\{s_1, s_2 \dots s_{M'}\}$, we can construct $\mathbf{g} = E\mathbf{s} + \bar{\mathbf{x}}_G$ where $E = [\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_{M'}]$ and $\mathbf{s} = [s_1 s_2 \dots s_{M'}]^t$.

Then, we define the inter-subspace distance between D and G to be $ISD(D, G)$ as shown below.

$$ISD(D, G) \equiv \min_{\mathbf{s}, \mathbf{z}} \|(E\mathbf{s} + \bar{\mathbf{x}}_G) - (P\mathbf{z} + \bar{\mathbf{x}})\|^2 \quad (1)$$

Mainly, the idea of ISD applied for two subspaces is to make the distance between a pair of constructed images re-

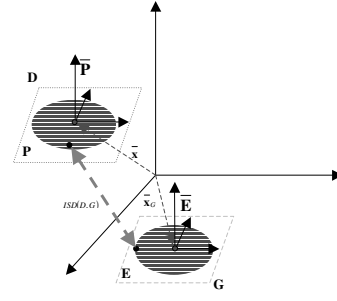


Figure 1. The idea of the Inter-Subspace Distance.

spectively from each subspace minimal. Figure 1 shows the idea of ISD.

Since image dimension is much larger than the sum of the number of eigenvectors (i.e., $d \gg (M + M')$) in common cases, we can obtain \mathbf{s}^* and \mathbf{z}^* such that $ISD(D, G) = \|(E\mathbf{s}^* + \bar{\mathbf{x}}_G) - (P\mathbf{z}^* + \bar{\mathbf{x}})\|^2$ by using the following least square equation.

$$\mathbf{v}^* = (\mathbf{K}^t \mathbf{K})^{-1} (-\mathbf{K}^t \mathbf{h}) \quad (2)$$

where

$$\mathbf{K} = [E|P], \mathbf{v}^* = [\mathbf{s}^{*t} | -\mathbf{z}^{*t}]^t, \mathbf{h} = (\bar{\mathbf{x}}_G - \bar{\mathbf{x}}) \quad (3)$$

From the viewpoint of system implementation, we need a fast algorithm for calculating eigenvectors. If the number of images is much smaller than the number of pixels in each image and we are only interested in the first several eigenvectors, a rapid algorithm which uses singular value decomposition (SVD) can be applied to calculate eigenvalues and eigenvectors [9].

In summary, the proposed ISD algorithm can directly measure the dissimilarity between two subspaces. The measurement of ISD is origin invariant, that is, invariant with the origins of the feature space. Since the performance of MSM may be affected by different origins. In addition, compared to KLD, in common cases our method can be computed efficiently without taking care of singular cases. Furthermore, ISD is symmetric but KLD is asymmetric.

Moreover, based on the linear combination concept and following reasons, recognizing a subject according to their subspace is reasonable. First, to deal with illumination changes, Basri and Jacobs proved that the convex Lambertian object can be approximated by a low-dimensional linear subspace [2]. Second, [13] showed that the variety of views depicting the same object under different transformations can often be expressed as the linear combination of a small number of views. Therefore, applying the ISD to measure the dissimilarity between subspaces for face recognition is reasonable and is more robust.



(a)



(b)

Figure 2. Generated virtual images. (a) scaled images. (b) shifted images.

3. Experimental Results and Discussions

In order to evaluate the performance of the proposed method, we compare it with several schemes (MSM [14] and KLD [11]). Because we focus on the problem of face recognition from multiple samples, the testing database which consists of image sequence for each individual is necessary. Therefore, we choose an open database which consists image sequences showing different facial expression for each subject. On the other hand, a performance on a well-known database in single-shot face recognition researches is reported in the second experiment. However, because the advantages of our method are which can cope with multiple samples simultaneously for face recognition, the direct comparison is unfair.

3.1. Preprocessing

To overcome the misaligned situations, the multiple sampling strategy (along with their position and size-scaling variations) is used. For a given test image, we generate 17 virtual images by scaling it with a set of scaling factor (0.8, 0.9, 1, 1.1, 1.2) and shifting each direction at most three pixels. This process is used to synthesize a better aligned test image under the assumption that the best face region is located in the neighbors of detected one. For example, Figure 2 shows the generated virtual images for the test image which is the middle image in Figure 2(a). Hence, for a test image, a test image set which consists of its generated virtual images is formed.

3.2. Face Recognition on CMU Face Expression Database

At the beginning, we test our method, MSM, and KLD on the CMU face expression database [7]¹. This database contains 13 subjects and each subject has 75 images. These face images are collected in the same lighting condition using CCD camera. Face images have been well-registered by



Figure 3. The CMU face expression database.

the eyes locations. Figure 3 shows all images of one subject in this database. Each face image is subsampled to 32×32 pixels and we represent each image by a raster scan vector of the intensity values. The images include variation in expressions. The experiment is performed by using “5-fold like” strategy: Instead of using a single image in a fold as a test image, all 15 images in a fold are used as test images as a whole to generate a test set ($15 \times 17 = 255$) by scaling and shifting them as described in the Section 4.1, other remaining 60 (15×4) images are used for training.

All compared methods are performed on the same kind of situation as described above. ISD and MSM methods can achieve averages 0% error rate. However, the KLD method has worse performance (error rate more than 50%) than MSM and ISD methods in this testing.

3.3. Face Recognition on AT&T Face Database

Similar to experiment 1, we use the AT&T (formerly Olivetti)² database as the benchmark for evaluating the performance of our method. This database contains 400 images of 40 subjects and Figure 4 shows all images of two subjects in this database. We follow the experiment setting as in [15]. Each face image is downsampled to 28×23 pixels. We represent each image by a raster scan vector of the intensity values, and then normalized them to be zero-mean vector. The images include variation in pose as well as scale. The experiment is performed by using “leave-one-out” strategy: To classify an image (a virtual image set (1×17)) in the database, other remaining 9 images in the database are used for training. ISD and MSM method are performed on the same situation. ISD can achieve 2.5% error rate which is better than that of MSM, 3% but the error rate of KLD is still more than 50%.

For completeness, we briefly report the error rates of the state of the art methods which use single-shot for face recognition on this database [15]. For example, error rates of Eigenface [12], Fisherface [3] and Independent Component Analysis (ICA) [1] are 2.5%, 1.5% and 6.25% respectively. In addition, the reduced dimensions are 40, 39 and 80 respectively.

However, it is inappropriate to directly compare our performance with these error rates from single-shot-based face recognition methods. The main reason is that how to ex-

¹ <http://amp.ece.cmu.edu/projects/FaceAuthentication/download.htm>

² <http://www.uk.research.att.com/facedatabase.html>



Figure 4. The AT&T face database.

tend methods which use single-shot for recognition to those which use multiple images so that the performances can be improved is not clear yet.

3.4. Discussions

In past researches of face recognition from videos or image sequences, they focus on manipulating image data with temporal information. The experimental results of [5] and [6] support the importance of temporal information. On the other hand, multiple sampling is a common strategy for dealing with the misalignment problem caused by biased face detection. The underlying assumption of multiple sampling is that there exists a better aligned face around the detected region. Since these two strategies have their individual benefits, our method can combine these strategies for face recognition simultaneously. In Section 3.2, we test the performance of the combination of temporal information and multiple sampling. The result shows that our method has high accuracy and robustness even every image has high expression variation.

4. Conclusion

In this paper, we propose a new method, ISD, which compares two image sets for face recognition. Mainly, the idea of ISD applied for two subspaces is to make the distance between a pair of constructed images respectively from each subspace minimal. Compared with the MSM method, the performance of our method is invariant to origins of the feature space. Furthermore, the solution of our method can be obtained by using standard least square method in most common cases.

According to past researches, it has been shown that the pose or illumination variations can be modelled as the combinations of lower linear subspace. Therefore, the proposed method which recognizes a subject according to their subspace is more robust and suitable for on-line face recognition systems. Moreover, another advantage of our method is that it can use multiple sampling for face recognition and thus ease the burdens of face detection.

Since current version of ISD is based on Euclidean distance, one future direction is to extend ISD to Mahalanobis distance. In addition, we will investigate the ISD method on other similar applications such as face detection or object recognition.

Acknowledgments

This work is supported in part by the National Science Council under project NSC 92-2213-E-001-022 and Ministry of Economic Affairs of R.O.C. under project 93-EC-17-A-02-S1-032. This work was done while J. -H. Chen was a research assistant at Institute of Information Science, Academia, Taipei, Taiwan.

References

- [1] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," in *Proceedings of the SPIE Symposium on Electronic Imaging III*, San Jose, California, Jan., 1998.
- [2] R. Basri and D. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE TPAMI*, vol. 25, No. 2, pp. 218-233, 2003.
- [3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE TPAMI*, vol. 19, no. 7, JULY 1997.
- [4] P. N. Belhumeur and D. J. Kriegman "What is the Set of Images of an Object under all Possible Illumination Conditions?," *IJCV*, vol. 28, pp. 245-260, 1998.
- [5] Z. Biuk and S. Loncaric, "Face Recognition from Multi-Pose Image Sequence," in *Proc. of 2nd International Symposium on Image and Signal Processing and Analysis*, pp. 319-324, Pula, Croatia, 2001.
- [6] Y. Li, S. Gong, and H. Liddell, "Video-Based Online Face Recognition Using Identity Surface," in *Proc. of RATFG-RTS*, pp 40-46, Vancouver, Canada, July 2001.
- [7] X. Liu, T. Chen, and B.V.K. V. Kumar, "Face Authentication for Multiple Subjects Using Eigenflow," *Pattern Recognition, special issue on Biometric*, Volume 36, Issue 2, February 2003, pp. 313-328.
- [8] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE TPAMI*, vol. 24, pp. 748-763, 2002.
- [9] S. K. Nayar and T. Poggio, *Early Visual Learning*, chapter 5, Oxford University Press, New York, 1996.
- [10] K. -K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE TPAMI*, vol. 20, pp. 39-51, 1998.
- [11] G. Shakhnarovich, J. W. Fisher, and T. Darrell, "Face recognition from long-term observations", in *Proc. of ECCV*, 2002.
- [12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-96, 1991.
- [13] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models," *IEEE TPAMI*, vol. 13, pp. 992-1006, 1991.
- [14] O. Yamaguchi, K. Fukui, and K. Maeda, "Face Recognition Using Temporal Image Sequence," in *Proceeding of IEEE International Conference on AFG*, pp. 318-323, Nara, Japan, 1998.
- [15] M. -H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," in *Proceedings of the Fifth IEEE International Conference on AFG*, pp. 215-220, 2002.