Image-Set Matching using a Geodesic Distance and Cohort Normalization

Yui Man Lui, J. Ross Beveridge, Bruce A. Draper
Department of Computer Science
Colorado State University
Fort Collins, CO 80523, USA
{lui}{ross}{draper}@cs.colostate.edu

Michael Kirby
Department of Mathematics
Colorado State University
Fort Collins, CO 80523, USA
kirby@math.colostate.edu

Abstract

An image-set based face recognition algorithm is proposed that exploits the full geometrical interpretation of Canonical Correlation Analysis (CCA). CCA maximizes the correlation between two linear subspaces associated with image-sets, where an image-set is assumed to contain multiple images of a person's face. When these linear subspaces are viewed as points on a Grassmann manifold, then geodesic distance on the manifold becomes the natural way to compare image-sets. The proposed method is tested on the ORL data set where it achieves a rank one identification rate of 98.75%. The proposed method is also tested on a subset of the Face Recognition Grand Challenge Experiment 4 *data. Specifically,* 82 *probe and* 230 *gallery subjects with* 32 *images per probe and gallery image-set. Our algorithm achieves a rank one identification rate of* 87% *and a verification rate of* 81% *at a false accept rate of* $\frac{1}{1{,}000}$*. These results on FRGC are significantly better than the well-known image-set matching algorithm, Mutual Subspace Method (MSM), which does not use geodesic distance. Another important finding is that cohort normalization boosts verification performance by* 50% *when used in conjunction with image-set matching. These results suggest that excellent levels of face recognition performance are possible when using image-sets, geodesic distance and cohort normalization. Finally, the proposed approach is generic in the sense that no training is required.*

1. Introduction

Face recognition performance improves when multiple images are exploited [3,15]. While others have championed a variety of approaches to image-set matching [1, 4, 10, 14, 19, 21, 22], no single standard approach has emerged. Perhaps the simplest brute force approach is to match images one by one and take the maximum similarity as a match score. More sophisticated approaches utilize properties of the image sets themselves. One such technique is Canonical Correlation Analysis [9] (CCA).

CCA is an established technique in multivariate statistics and its mathematical properties are well-defined. For face recognition, CCA has several attractive properties. First, the number of query images does not need to be the same as the number of target images. Second, CCA is affine transformation invariant. This makes CCA robust relative to global intensity variations. Third, CCA captures the full geometrical relationship between linear subspaces spanned by two face image sets.

Since CCA maximizes the correlation between two linear subspaces, its underlying geometry is related to Grassmann manifolds. Essentially, a set of face images spans a subspace residing on a Grassmann manifold as a point. The corresponding distance between points on a Grassmann manifold is characterized by geodesic distance, i.e. arc-length.

Here we present an image-set face recognition algorithm based upon geodesic distance. For this algorithm the following is important. First, a Gabor filter is used to compensate for illumination variation and emphasize local structures. Second, cohort normalization is used to adjust similarity scores for probes based upon their overall response to the gallery. Third, as mentioned above, image-sets are compared using geodesic distance on a Grassmann manifold.

The prior work closest to ours is the Mutual Subspace Method (MSM) [22]. However, MSM only considers what is called the first principal angle and neglects the larger geometry captured by geodesic distance. A direct comparison between our algorithm and MSM is presented for the ORL [18] and a subset [1] of the Face Recognition Grand Challenge (FRGC), Version 2, Experiment 4 [15] data sets. The ORL data set is widely known, and our algorithm achieves an average rank one identification rate of 98.75% while a comparable MSM algorithm achieves 98%.

The subset of the FRGC is more challenging. Our al-

[1] There are too few images for some subjects for us to include them in an image-set experiment.

978-1-4244-2154-1/08/$25.00

gorithm achieves an average rank one identification rate of 87% and an average verification rate of 81% at a false accept rate (FAR) of $\frac{1}{1,000}$. The comparable MSM algorithm, even with the advantage of the Gabor filtered images and Cohort normalization, only achieves an average rank one identification rate of 76% and an average verification rate of 59% at a false accept rate (FAR) of $\frac{1}{1,000}$. The improvement derived from using geodesic distance is significant.

For both our algorithm and MSM, cohort normalization improves verification performance by roughly 50%. Cohort normalization is interacting with image-set matching and boosting performance far more than is typically observed in other biometric matching tasks.

The rest of this paper is organized as follows. Section 2 reviews related work. The scheme for face image preprocessing is given in Section 3. CCA and the use of geodesic distance are described in Section 4. Cohort normalization is discussed in Section 5. Baseline algorithms are introduced in Section 6, and experimental results are presented in Section 7. Finally, Section 8 summarizes our conclusions.

2. Related Work

Image-set matching is a relatively new paradigm. The applications of this new scheme include both multiple still images and video sequences. A recent survey for image-set face recognition can be found in [25].

Yamaguchi *et al*. [22] first introduce image-set matching for face images taken from video sequences using a mutual subspace method (MSM). This method essentially computes the intersection between two subspaces and takes the first canonical correlation as the similarity measure. The authors show that MSM yields higher performance than the conventional single image subspace method.

Nishiyama *et al*. [14] extend MSM to incorporate a hierarchical framework using multiple cameras. Three matching layers are used. On the first layer face images are matched using a single camera and then fragmentary sets are connected between cameras in the second layer. In the third layer, face matching is performed with the connected set using MSM. The authors report that using multiple cameras and dividing the matching process into sub-processes reduce the effects of pose and illumination variations.

Another school of thought for image-set matching is to compute the Kullback-Leibler divergence (KL divergence) between probability distributions. Shakhnarovich *et al*. [19] develop a closed-form evaluation of the KL divergence between two distributions. This method assumes that samples are drawn from $i.i.d$ (independent and identically distributed) distributions such that k-ary hypothesis tests can be used as likelihood statistics.

Arandjelovic *et al*. [1] notice that the simple KL-divergence approach performs rather poorly because it does not capture non-linear variations. The authors introduce a resistor-average distance (RAD) as a distance measure which is a symmetric version of KL-divergence. This algorithm is applied to face recognition in video sequences and consists of several steps. First, the RANSAC algorithm is used to remove outliers. Second, synthetic affine images are constructed. A projection matrix is then created using RANSAC kernel PCA with the synthetic and outlier removal images. Finally, the data sets are projected onto this subspace and RAD is used to compute the distance on the projected data. This method shows nearly a factor of two gain over the simple KL-divergence [19] and about 11.4% improvement over the MSM [22]. Nevertheless, this method seems complicated and time-consuming.

Wolf and Shashua [21] employ a kernel to map data from the original space to a nonlinear feature space, so that principal angles are implicitly computed in the high dimensional space. This technique is applied to face recognition in video sequences and motion trajectory discrimination. Training is required and the choice of kernel is ad hoc.

Cheng *et al*. [4] apply principal angles to compute subspace distances on Grassmann manifolds for face recognition in illumination space, and demonstrate that the CMU PIE database can be perfectly separated due to the linear characteristic of illumination space. Nonetheless, real-world face recognition does not operate in such well controlled environment.

Kim *et al*. [10] propose an LDA-like optimization method that maximizes the canonical correlations of within-class sets and minimizes the canonical correlations of between-class sets for face recognition in video sequences. Test sequences are then transformed using the trained discriminant function and similarity scores are computed using canonical correlations. Large amounts of training data are required for this method.

3. Face Image Preprocessing

Face images are first aligned using eye coordinates supplied with the data from the FRGC. [2] An elliptical bounding mask is applied to all face chip images. Figure 1 shows a geometrically normalized face chip image.

To compensate for the variations of illumination and obtain localized features, we employ a Gabor filter [11] for image preprocessing. Gabor filters have localization properties in both spatial and frequency domains such that particular orientations and spatial frequencies can be band passed, thus, it is useful for local feature extraction. The formulation of a DC-free Gabor filter is depicted as follows:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{ik_{u,v}z} - e^{-\sigma^2/2} \right] \quad (1)$$

[2] Face chip images are provided for the ORL data set.

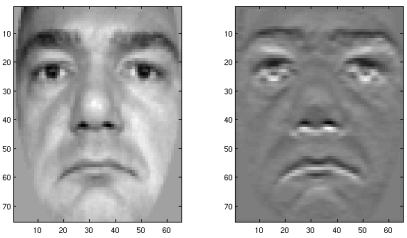

Figure 1. The left figure shows a geometrically normalized face chip and the right figure shows the image after the Gabor transform

where $k_{u,v} = \frac{k_c}{f^v} e^{i\phi_u}$, u and v are the control parameters for orientations and scales, respectively, and z is the position. In our experiments, we employ a single Gabor wavelet as a preprocessing filter in which we set the $k_c = \frac{4\pi}{5}$, $\sigma = \frac{3\pi}{2}$, $v = 0$, and $\phi_u = 0$. An example of a Gabor processed face chip image is given in the right figure of Figure 1.

4. Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) was first introduced by Hotelling [9] in 1936. This technique finds the degree of intersections between two linear subspaces. From a geometric point of view, CCA simultaneously rotates two coordinate frames such that the correlation of the transformed data is maximized. Essentially, finding the rotation matrices is equivalent to seeking linear transformations of variables. A recent tutorial of CCA can be found in [8].

Given two images $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, CCA seeks two projection vectors that yield a maximum correlation in a new coordinate system. Formally, the canonical correlation between two projected vectors can be defined as

$$\rho = \max_{w_x, w_y} \frac{\text{Cov}(w_x^T x,\ w_y^T y)}{\sqrt{\text{var}(w_x^T x)\ \text{var}(w_y^T y)}} \quad (2)$$

$$= \max_{w_x, w_y} \frac{w_x^T C_{xy} w}{\sqrt{(w_x^T C_{xx} w_x)\ (w_y^T C_{yy} w_y)}} \quad (3)$$

where C_{xy} is $x\,y^T$, C_{xx} is $x\,x^T$, and C_{yy} is $y\,y^T$. Equation (3) can be rewritten as

$$\rho = \max_{w_x, w_y} w_x^T C_{xy} w_y \quad (4)$$

subject to

$$w_x^T C_{xx} w_x = 1, \qquad w_y^T C_{yy} w_y = 1$$

Equation (4) can then be reformulated as a Lagrangian formulation shown as

$$\mathcal{L}(w_x, w_y, \lambda_x, \lambda_y) = w_x^T C_{xy} w_y + \lambda_x(1 - w_x^T C_{xx} w_x) + \lambda_y(1 - w_y^T C_{yy} w_y) \quad (5)$$

where λ_x and λ_y are the Lagrangian multipliers. By differentiating $\mathcal{L}$ with respect to w_x and w_y, the solution of Equation (5) becomes a generalized eigen-system described as follows:

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \Lambda \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \quad (6)$$

where Λ is the canonical correlation.

Generalizing for a set of images, let $X = [x^{(1)} \dots x^{(p)}]$ and $Y = [y^{(1)} \dots y^{(q)}]$ be two image sets where the number of samples for a data set X and Y are p and q, respectively. The combined data samples can be represented as $Z = [X|Y]^T$.

The covariance matrix, Cov(Z) = $Z\,Z^T$, can then be written as a block matrix:

$$\text{Cov}(Z) = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \quad (7)$$

where $C_{xy} = {C_{yx}}^T$. Note that C_{xy} represents the covariance between sets whereas C_{xx} and C_{yy} represent the covariance within sets. Applying the eigen-solution from Equation (6), the canonical correlations between data set X and data set Y are embedded in the diagonal elements of Λ.

CCA computes the canonical correlation between the range of two data sets. The next subsection describes a geometric interpretation of CCA in terms of the spatial relationship between the two image-sets X and Y.

4.1. Geodesic Distance on Grassmann Manifolds

It is known that subspaces can be represented as points on Grassmann manifolds [6]. Unlike a vector space, manifolds are usually curved, and the shortest distance between two points on a manifold is geodesic distance. In general, geodesic distance should be used to compare subspaces on the Grassmann manifold, and in particular we adopt arc-length defined as [6]:

$$d(X, Y) = \left(\sum_{i=1}^{k} \theta_i^2 \right)^{\frac{1}{2}} = ||\theta||_2 \quad (8)$$

where θ_i is the ith principal angle and k is the number of dimensions of the Grassmann manifold which is equal to the minimum rank of X and Y. Since canonical correlations are equivalent to cosines of principal angles, we can interchange the geodesic distance with a similarity score described as

$$S_{geodesic}(X, Y) = \left(\sum_{i=1}^{k} \rho_i^2 \right)^{\frac{1}{2}} \quad (9)$$

where ρ_i is $\cos(\theta_i)$.

Although canonical correlations (principal angles) have been exploited for image-set matching, most algorithms do not consider the geometry of the space. Yamaguchi *et al*. [22] only use the first canonical correlation as a similarity measure. Wolf and Shashua [21] adopt the average of the twenty smallest principal angles for face recognition between two video sequences. Unlike these previous methods, we take the full geometry of manifolds into account and use the geodesic distance defined in Equation (9).

5. Cohort Normalization

Some people are harder to recognized than others [5]. Traditional face verification systems use a fixed threshold to authenticate whether a subject is a claimant or impostor. However, a fixed threshold does not adapt from person to person. Cohort normalization, which is widely used in speaker verification [17], uses the set of matching scores between a probe and gallery to adaptively adjust scores.

Phillips *et al*. [16] report cohort normalization can improve verification results by about 10%. In this paper, we demonstrate that verification results can be boosted by 50% when cohort normalization is applied in conjunction with image-set matching. Specifically, Z-norm [2] score normalization is defined as:

$$\hat{s}_{ij} = \frac{1}{(1+e^{-t})}, \qquad t = \frac{s_{ij} - \mu_i}{\sigma_i} \tag{10}$$

where $\hat{s}_{ij}$ is a normalized similarity score, s_{ij} is a raw similarity score between image/subject i and image/subject j, and μ_i and σ_i are the mean and standard deviation of the match score between the image/subject i and the gallery. The sigmoid function in Equation (10) maps the similarity score to an interval between 0 and 1, and the similarity score is adjusted according to the mean and standard deviation with respect to the entire gallery. To avoid numerical issues, we set the minimum σ_i to be 0.01. Note that cohort normalization does not change identification rank.

6. Baseline Algorithms

Our proposed algorithm is compared to two baseline algorithms, one using simple image correlation and the other being the MSM algorithm proposed by Yamaguchi *et al*. [22]. The correlation algorithm is a modified one-to-one matching algorithm whereas MSM is designed for image-set matching.

The modified one-to-one correlation method for image-set matching may be described as follows: Let $X = [x^{(1)} \dots x^{(p)}]$ and $Y = [y^{(1)} \dots y^{(q)}]$ be two image-sets, $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$, where images in the same set have the same subject identity.

$$S_{cor}(X,Y) = \max_{1 \le i \le p} \max_{1 \le j \le q} \{\text{Cor}(x^{(i)}, y^{(j)})\} \tag{11}$$

where Cor denotes the standard normalized correlation, x is a probe image, and y is a gallery image. This image-set similarity measure essentially computes the maximum single correlation between pairs of image in X and Y. Therefore, it can be regarded as a brute force approach.

The Mutual Subspace Method (MSM) [22] is our second baseline algorithm. Unlike the modified correlation method, MSM was originally designed for image-set matching but does not consider the full geometry of the space. For the sake of completeness, we briefly describe the MSM algorithm [22]. Let matrix Z be

$$Z = \Psi^T \Phi \tag{12}$$

where Φ and Ψ are the orthogonal bases of X and Y, respectively. Then, Z can be diagonalized as

$$U^T Z V = \Lambda \tag{13}$$

where U and V are the left singular vectors and the right singular vectors of Z, respectively, and Λ is the diagonal singular value matrix of Z. Similarity is defined as:

$$S_{msm}(X,Y) = \lambda_{max} \tag{14}$$

where λ_{max} is the square root of the largest singular value. This is essentially the first canonical correlation from CCA.

7. Experimental Results

We evaluate our algorithm on the well-tested ORL database [18] as well as a subset of FRGC, version 2, Experiment 4 [15]. The subset of the FRGC is a much larger and more challenging data set, and the importance of using geodesic distance and cohort normalization becomes apparent.

7.1. The ORL Data Set

The ORL database contains 40 subjects and each subject has 10 different images. All images are taken from a controlled environment with variations in pose and facial expressions. We will present ORL identification rates for our CCA-Geodesic algorithm, the MSM algorithm [22] and five single-image matching algorithms [7, 12, 20, 23, 24].

Because our CCA-Geodesic and the MSM algorithms match image-sets, the protocol is different from single-image matching. Specifically, the gallery and the probe contain 40 entries, each entry is a set of five face images. As the matching process, the probe is said to be identified at rank one if the top ranked gallery image-set is of the same person as the probe image-set. The selection of specific ORL images placed in the gallery and probe sets is randomized for 10 trials and the average rank one identification rate is reported in Table 1.

	Rank One Identification	
Methods	Single-Image	Image-Set
Max Variance Projection [24]	95.75%	-
Wavelet + Adaptation [12]	96.60%	-
GCCA [20]	97.50%	-
MSM [22]	-	98.00%
Wavelet + Kernel [23]	98.20%	-
CCA-Geodesic	-	**98.75%**
Local Graph Matching [7]	100%	-

Table 1. Rank one identification results on the ORL database

The rank one identification rates shown in Table 1 for the other five algorithms are taken from the papers [7,12,20,23, 24]. All of these algorithms involve training and the authors typically report training on five images per subject. It is also assumed that the training and gallery images are the same, hence their galleries contain 200 images with five replicates per subject. Like every coin has two sides, the single-image matching algorithms do not have the benefit of interacting between probe images but they profit from the presence of multiple matches in the gallery.

The rank one identification rates in Table 1 are displayed in separate columns to underscore the different protocols just summarized. This difference notes that comparable results obtained from trained single-image matching algorithms can be achieved with a generic image-set matching method. Regarding the MSM algorithm versus our CCA-Geodesic algorithm, the identification rate difference is only 0.75%. The following FRGC results show where exploiting the underlying geometry becomes important.

7.2. The FRGC Data Set

FRGC, Version 2, Experiment 4 is one of the most recent and challenging tests for face recognition. The experiment has $8,014$ query images collected from an uncontrolled environment and $16,028$ target images collected from a controlled environment. [3] In addition, the number of images per subject varies from 2 to 88. Due to limited numbers of images per subject, we select 32 images for each subject as our primary data set size. This subset of FRGC Experiment 4 includes 82 probe subjects with at least 32 images and 230 gallery subjects that also have at least 32 images. Each image set represents an instance of a subject. In our experiments, we only allow one instance per subject in probe and gallery databases. Furthermore, we randomly select 32 images for each subject and perform 10 trial runs.

Rank one identification rate results are presented in Figure 2. Both the MSM and CCA-Geodesic algorithms outperform the baseline image-set correlation method (Image-Set-Cor). Our CCA-Geodesic algorithm achieves an average 87.2% rank one identification rate. This 15.3% improvement over MSM is due entirely to the switch to geodesic distance, demonstrating the importance of manifold geometry for larger image-sets. To emphasize this point, results for different sized image-sets are shown in Table 2. Note that the values on the last row, size equals 32, match the averages of the rank one identification and verification rates reported in Figure 2 and Figure 3, respectively.

[3] Probe images are taken from the query images and gallery images are taken from the target set.

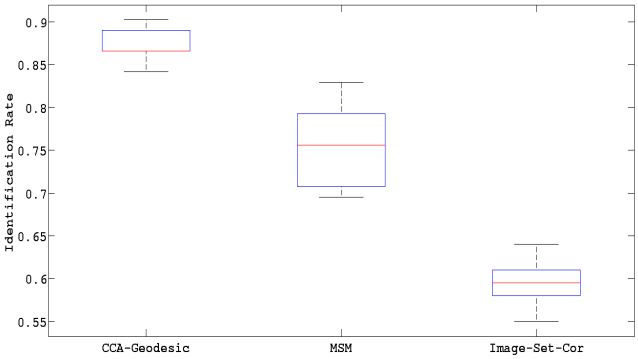

Figure 2. Rank one identification rate using 32 images against 32 images on the subset of FRGC data set (No Cohort Normalization)

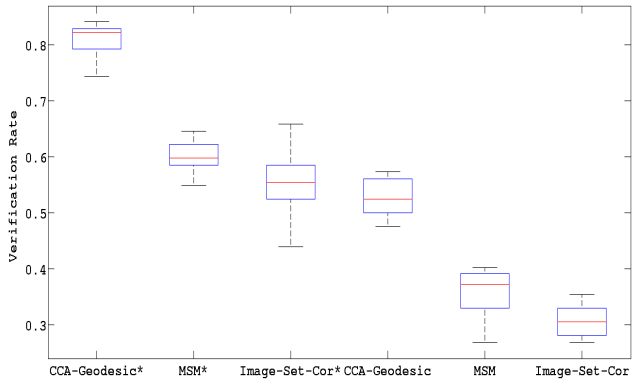

Figure 3. Verification rate at $\frac{1}{1,000}$ FAR using 32 images against 32 images on the subset of FRGC data set (where * indicates the use of cohort normalization)

Set Size	Rank One Identification	Verification
1	14.3%	8.2%
2	24.9%	15.4%
4	40.2%	23.1%
8	59.2%	47.3%
16	79.4%	69.9%
32	87.2%	81.3%

Table 2. The size effect on the FRGC data set.

Verification results at $\frac{1}{1,000}$ FAR are revealed in Figure 3. This figure shows that cohort normalization is particularly preferred with image-set matching. The verification rates are boosted up by 50% when cohort normalization is employed in conjunction with image-set matching. This indicates that cohort normalization adjusts the deci-

sion threshold well and allows image-set matching algorithms to yield better verification. Furthermore, the proposed CCA-Geodesic algorithm (CCA-Geodesic*) outperforms MSM (MSM*) as well as the image-set correlation method (Image-Set-Cor*), and obtains an average 81.3% verification when cohort normalization is applied.

8. Conclusions

We have presented a new approach to image-set matching applied to face recognition. We consider the full geometrical interpretation of CCA and employ geodesic distance accordingly. The importance of using cohort normalization for verification has been demonstrated. Both identification and verification results using image-sets for the FRGC, Version 2, Experiment 4 imagery show that the proposed approach is superior to those achieved by the closest comparable algorithm in the literature. In addition, the proposed algorithm is easily implemented and generic in the sense that it requires no training or tuning.

Both ORL and the subset of FRGC results suggest that substituting additional imagery during recognition for use of additional imagery in a training phase can result in overall recognition performance that is comparable to the best trained single image matching algorithms [7, 13].

Finally, understanding what makes a good image-set, particularly in reference to the geometry of the associated image subspaces, is an important focus for future work.

Acknowledgements

This work is supported in part by the National Science Foundation under Grant No. 0413284 and No. 0434351.

References

[1] O. Arandjelovic and R. Cipolla. An information-theoretic approach to face recognition from face motion manifolds. *Image and Vision Computing*, 24(6):639–647, 2006.

[2] C. Barras and J.-L. Gauvain. Feature and score normalization for speaker verification of cellular data, 2003. IEEE International Conference on ASSP, Hong Kong.

[3] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2d+3d face biometrics. *PAMI*, 27(4):619–624, 2005.

[4] J.-M. Cheng, J. Beveridge, B. Draper, M. Kirby, H. Kley, and C. Peterson. Illumination face spaces are idiosyncratic, 2006. The International Conference on IPCVPR.

[5] G. Doddingyon, W. Liggett, A. Martin, M. Przbocki, and D. Reynolds. Sheep, goats, lambs and wolves - a statistical analysis of speaker performance in the nist 1998 speak recognition evaluation, 1998. The 5th International Conference on Spoken Language Processing, Sidney, Australia.

[6] A. Edelman, R. Arias, and S. Smith. The geometry of algorithms with orthogonal constraints. *SIAM J. Matrix Anal. Appl.*, (2):303–353, 1999.

[7] E. Fazl-Ersi, J. S. Zelek, and J. K. Tsotsos. Robust face recognition through local graph matching. *Journal of Multimedia*, pages 31–37, 2007.

[8] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis; an overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2003.

[9] H. Hotelling. Relations between two sets of variants. *Biometrika*, 28:321–377, 1936.

[10] T.-K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *PAMI*, 29(6):1–14, 2007.

[11] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, March 1993.

[12] J. Lin, J.-P. Lin, and M. Ji. Robust face recognition by wavelet features and model adaptation, 2007. ICWAPR.

[13] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *PAMI*, 28(5):725–737, 2006.

[14] M. Nishiyama, M. Yuasa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi. Recognizing faces of moving people by hierarchical image-set matching, 2007. CVPR, Minneapolis, MN.

[15] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, and W. Worek. Preliminary face recognition grand challenge results, 2006. FGR, UK.

[16] P. Phillips, W. Scruggs, A.J.O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. Frvt 2006 and ice 2006 large-scale results, 2007. http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf.

[17] A. Rosenberg, J. Delong, C. Lee, B.-H. Juang, and F. Soong. The use of cohort normalized scores for speaker recognition, 1992. International Conference on SLP, Canada.

[18] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification, 1994. Proc. IEEE Workshop on Applications of Computer Vision.

[19] G. Shakhnarovich, J. Fisher, and T. Darrell. Face recognition from long-term observations, 2002. ECCV, Denmark.

[20] Q.-S. Sun, P.-A. Heng, Z. Jin, and D.-S. Xia. Face recognition based on generalized canonical correlation analysis, 2005. ICIC (2).

[21] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4:913–931, 2003.

[22] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence, 1998. FGR, Japan.

[23] B. L. Zhang, H. H. Zhang, and S. S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on Neural Networks*, 15:166–177, 2004.

[24] T. Zhang, J. Yang, H. Wang, and C. Du. Maximum variance projections for face recognition. *Optical Engineering*, 46, 2007.

[25] S. Zhou and R. Chellappa. Beyond one still image: Face recognition from multiple still image or video sequence, 2006. in Face Processing Advance Modeling and Methods, Edited by W. Zhao and R. Chellappa, Elsevier.