SET-TO-SET FACE RECOGNITION UNDER VARIATIONS IN POSE AND ILLUMINATION

Jen-Mei Chang, Michael Kirby, and Chris Peterson

Department of Mathematics
Colorado State University
Fort Collins, CO 80523-1874 U.S.A.
{chang, kirby, peterson}@math.colostate.edu

ABSTRACT

We present a face recognition method using multiple images where pose and illumination are uncontrolled. The set-to-set framework can be utilized whenever multiple images are available for both gallery and probe subjects. We can then transform the set-to-set classification problem as a geometric one by realizing the linear span of the images in a given resolution as a point on the Grassmann manifold where various metrics can be used to quantify the closeness of the identities. Contrary to a common practice, we will not normalize for variations in pose and illumination, hence showing the effectiveness of the set-to-set method when the classification is done on the Grassmann manifold. This algorithm exploits the geometry of the data set such that no training phase is required and may be executed in parallel across large data sets. We present empirical results of this algorithm on the CMU-PIE Database and the Extended Yale Face Database B, each consisting of 67 and 28 subjects, respectively.

1. INTRODUCTION

Face recognition under variations in illumination and pose has long been recognized as a difficult problem with pose appearing somewhat more challenging to handle than variations in illumination [1]. A direct approach to deal with such images has been to develop algorithms that normalize for variations in illumination and then to focus on a solution for pose [2], [3]. In contrast, as is shown in [4] and [5], it is an appealing and plausible idea that sets of images acquired under varying or non-uniform illumination conditions possess valuable discriminatory information. Furthermore, both theoretical and empirical evidence have demonstrated that there exist low-dimensional representations for a set of images of a fixed object under variations in illumination conditions [6, 4]. This suggests that a wider range of discriminatory information can be captured in a low dimensional model as opposed to discarding a portion of the data as noise. This observation is not new and examples of algorithms that attempt to solve the face recognition problem under variations of illumination and pose without factoring out the illumination variations are [7, 8, 9, 10].

Algorithms that are successful in recognizing subjects in uncontrolled environments rely on good models for both illumination and pose variations. In the typical representation of image data, variations in illumination is inherently linear. More precisely, images collected under a convex set of illumination conditions themselves form a convex set [9] and a vast majority of the energy of such data can be captured with a relatively low-dimensional linear space [4]. In contrast, images collected under variations in pose is not inherently linear. As a consequence, linear methods such as those based on the SVD perform poorly when pose variations are included. One natural non-linear approach for addressing pose variations is with a 3D Morphable Model as described in [10]. Such non-linear approaches often come with the expense of a training phase and manual feature extraction at the recognition stage.

Much work on face recognition has focused on defining the classification problem as comparing distances from a single incidence of the probe class to a single or multiple incidence(s) of the gallery classes. Such methods have a major drawback. For example, if the input image is noisy or the face is occluded, then the recognition result will not be reliable. Therefore, it is natural to consider the paradigm where multiple images for both the gallery and the probe subjects are available. The set-to-set framework used in this work is an example of the model-based approach where a subspace is formed for each set of images. Distances between subspaces are calculated on the Grassmann manifold and identification is accomplished based on the nearest neighbor criterion. Algorithms that rely on the framework of *set-to-set* comparison have shown significant promise in their ability to classify both still-images and video sequences [11, 12, 13, 5].

When a collection of images are available for a subject, we view the data as sampling (with noise) an underlying man-

This study was partially supported by the National Science Foundation under award DMS-0434351 and the DOD-USAF-Office of Scientific Research under contract FA9550-04-1-0094. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the DOD-USAF-Office of Scientific Research.

1-4244-1549-7/07/$25.00

ifold. In this paper, the underlying data manifold, M, captures pose and illumination variations of a fixed subject. If we fix an illumination condition, then the underlying data manifold, X, captures pose variation. There is a natural map $\phi : M \to X$ (under the fixed illumination condition). The fibers of the map (i.e. the inverse images of points on X) capture variations in illumination for a fixed pose. For each pose we can capture, with a low-dimensional linear space, the variations in illumination. Fixing the dimension of the linear space used to capture the illumination data to be k, we obtain a map of X into the parameter space of k-dimensional linear spaces inside the ambient space used to represent the images. We proceed with this model in the background. As we shall see, the set-to-set paradigm affords algorithms that require no training. Although we present our results in the context of other work, given the paradigm shift, direct comparisons are difficult at best. We will attempt to demonstrate the robustness of the proposed algorithm by purposely omitting the preprocessing stage: *we employ the Yale and CMU-PIE data sets with none of the images geometrically normalized and with registration essentially ignored*. We present a suite of experiments that illustrate the potential of the set-to-set framework.

The remainder of this paper is organized as follows: in Section 2, we review the set-to-set framework for object classification. In Section 3, we present results of experiments that illustrate the set-to-set comparison paradigm in the face recognition problem with variations across illumination and pose. In addition, we discuss the computational complexity of the proposed algorithm. In Section 4, a few well-known state-of-the-art algorithms are presented to compare and contrast with the proposed method. Section 5 provides a summary of our observations.

2. CLASSIFICATION ON THE GRASSMANNIANS

An $r \times c$ gray scale digital image corresponds to an $r \times c$ matrix where each entry enumerates one of the 256 possible gray levels of the corresponding pixel. The image can then be seen as an element of $\mathbb{R}^n = \mathbb{R}^{r \times c}$. We will group k (generally independent) example images of a subject and consider the k-dimensional *feature subspace* they span in $\mathbb{R}^n$. This is the basis for a many-to-many set comparison paradigm.

If we let $G(k, n)$ denote the Grassmann manifold (Grassmannian) parameterizing k-dimensional real vector subspaces of the n-dimensional vector space $\mathbb{R}^n$, then the set-to-set classification problem can be transformed to a problem on $G(k, n)$ if we realize the linear span of a set of k images as a k-dimensional vector subspace of the space of all possible images at a given resolution. Our objective is to match an unlabeled *set* of images by comparing its associated point with a collection of given points on $G(k, n)$. As a consequence of the encoding of sets of images as points on a Grassmann manifold we may avail ourselves of a variety of well-known distance measures between points on the manifold. For example, if we realize $G(k, n)$ as a submanifold of $\mathbb{R}^{(n^2+n-2)/2}$ (as a Euclidean space) via the embedding described in [14] and if we restrict the Euclidean distance function on $\mathbb{R}^{(n^2+n-2)/2}$ to the Grassmannian, then we obtain the *projection F* norm or *chordal* distance, denoted by d_c. Specifically, the chordal distance between two points $X, Y \in G(k, n)$ is $d_c(X, Y) = ||\sin\theta||_2$, where $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ are the principal angles between the subspaces X and Y. Other example metrics between points on the Grassmannian are the *Fubini-Study* metric and the *arc length* metric[1]. For additional details, see [15].

A recursive definition and an SVD-based algorithm for computing the principal angles between X and Y can be found in [16]. Briefly, if X and Y are two vector subspaces of $\mathbb{R}^n$ such that $p = \dim(X) \geq \dim(Y) = q \geq 1$, then the principal angles $\theta_k \in \left[0, \frac{\pi}{2}\right]$, $1 \leq k \leq q$ between X and Y are defined recursively by

$$\cos(\theta_k) = \max_{u \in X} \max_{v \in Y} u^T v = u_k^T v_k$$

subject to $||u||_2 = ||v||_2 = 1$, $u^T u_i = 0$ and $v^T v_i = 0$ for $i = 1 : k-1$. Note that the standard distance between subspaces that is often presented in linear algebra is determined by the largest angle between the two subspaces. This ignores the geometric information associated with the smaller angles. We have observed that in many instances it is in fact the smallest (not largest) principal angle that carries the most significant information. It is then natural to consider an ℓ-truncated semi-chordal pseudo-metric as the following, $d_c^\ell = ||\sin\theta^\ell||_2$, where $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_\ell$ and $1 \leq \ell \leq \min\{\dim X, \dim Y\}$. It is a pseudo-metric since $d_c^\ell(X, Y) = 0$ whenever $\dim(X \cap Y) \geq \ell$.

If variations in pose are restricted to the equivalent of moving a camera in a horizontal circle at a fixed distance from the subject and with lens pointed at the subject, then by associating to each pose a k-dimensional linear space which captures variations in illuminations we obtain a map of a circle into $G(k, n)$. If variations in pose are restricted to all orientations of the camera a fixed distance from the subject and with lens pointed at the subject, then the variations in illumination provide a map of the lie group $SO(3)$ into $G(k, n)$. The image of these maps are what we think of as illumination invariant pose manifolds. Of course, the coarseness of registration and depth of field yields a more complicated picture. Nevertheless, we find the formulation to be compelling in the design of algorithms.

3. EXPERIMENTS

The data sets we used to empirically test our algorithm are the Extended Yale Face Database B (E-YDB) [9] and the "illum" subset of the CMU-PIE Database [17]. For the E-YDB, there are 28 different subjects each recorded under 9 poses

[1]Classification results depend on the geometry of the Grassmann manifold, which is determined by the metric selected.

1-4244-1549-7/07/$25.00

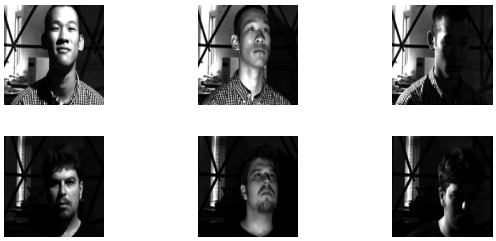

Fig. 1. Example images in the E-YDB that are used in the experiments.

and 65 illumination conditions. For the CMU-PIE Database, there are 67 subjects each recorded under 13 poses and 21 illumination conditions. We denote the image corresponding to subject s, pose p and illumination condition i by $J_{s,p,i}$. See Figure 1 for an illustration of images with variations in pose and illumination from the E-YDB. As you can see, the images in these databases are coarsely centered and coarsely controlled. The rough nature of the data makes the experiments applicable to a wider variety of real-life applications. We consider three experiments to test the proposed algorithm.

In Experiment I, the poses are treated separately. In Experiments II and III, the poses are pooled for each subject. In Experiments I and II, we include the probe pose in the gallery and the probe and gallery images each use a distinct set of k illuminations taken from each pose. These two experiments are merely a sanity check. Any recognition algorithm that claims to be successful in dealing with variations of illumination and pose should perform very well in these two experiments. In Experiment III, we remove the probe pose from the gallery in addition to using a distinct set of k illuminations. To this end, we test the algorithm's ability to recognize novel viewpoints.

We describe the experiments below for a single probe set. Let the number of distinct subjects in either the E-YDB or the CMU-PIE Database be s_0, the number of distinct poses be p_0, and the number of distinct illuminations be i_0. The distance measure used will be $d = d_c^\ell$ in the following experiment descriptions. We will describe how the set of probe and gallery images are selected, indicate how error statistics are compiled and analyze the results.

3.1. Experiment I

In Experiment I we view each pose as an additional subject in the database while retaining the information that each of the p_0 poses are associated with a given subject. The probe set (resp. gallery set) associated with subject α and pose β is written as $P_{\alpha,\beta}$ (resp. $G_{\alpha,\beta}$). We have

$$P_{\alpha,\beta} = \bigcup_{i \in I_P} J_{\alpha,\beta,i} \text{ and } G_{\alpha,\beta} = \bigcup_{i \in I_G} J_{\alpha,\beta,i},$$

where I_P denotes the set of illuminations associated with the probe and I_G denotes the set of illuminations associated with the gallery. The set of indices defining I_P and I_G is chosen randomly with the restriction that $I_P \cap I_G = \emptyset$. In Experiments I, II, and III, we use $|I_G| = |I_P| = k$ with $k = 16$ for E-YDB and $k = 10$ for CMU-PIE. For a fixed α and β,

$$P_{s^*,p^*} = \arg\min_{s,p} d(P_{\alpha,\beta}, G_{s,p}).$$

If $s^* = \alpha$ and $p^* = \beta$, then we have achieved the correct classification. Thus a single pose of a single subject is compared individually to each pose set of each subject. In essence this requires the algorithm to recognize both pose and identity.

3.2. Experiment II

We now consider the case where, for the gallery, we pool the different poses and illuminations into a single set associated with each subject, i.e.,

$$G_s = \bigcup_{p=1}^{p_0} \bigcup_{i \in I_G} \{J_{s,p,i}\}.$$

Now the gallery set consists of s_0 sets of images where each set has p_0 different poses and $|I_G|$ illuminations. Again, a single probe set $P_{\alpha,\beta}$ is associated with one subject α and one pose β over a set of $|I_P|$ illuminations. For a fixed α and β, we solve $P_{s^*} = \arg\min_s d(G_s, P_{\alpha,\beta})$. If $s^* = \alpha$, then the classification is correct.

3.3. Experiment III

In this experiment we remove the pose associated with the probe from all of the sets in the gallery. Hence, for each $\alpha = 1, \ldots, s_0$ and $\beta = 1, \ldots, p_0$ we seek to solve the equation $P_{s^*} = \arg\min_s d(G'_s, P_{\alpha,\beta})$, where

$$G'_s = \bigcup_{\substack{p=1 \\ p \neq \beta}}^{p_0} \bigcup_{i \in I_G} \{J_{s,p,i}\}.$$

If $s^* = \alpha$, then classification is correct.

3.4. Analysis

For each experiment we compute the errors using each pose and each subject, therefore a total of $28 \times 9 = 252$ probe sets for E-YDB and $67 \times 13 = 871$ for CMU-PIE. In addition, we randomly partition i_0 illumination conditions into two disjoint sets of k, one for the gallery and the other for the probe. Table 1 shows the recognition rates for experiments I – III on both databases when using the 1- truncated chordal distance d_c^1.

To visualize the contrast in performance of the algorithm for Experiments I and II versus Experiment III, we examine the first principal vector for a set of probe and gallery set using images in CMU-PIE in each Experiment in Figures 2 and 3.

1-4244-1549-7/07/$25.00

Error Rate (%)	Experiment		
Database	I	II	III
Extended YDB	0	0	6.7
CMU-PIE	0	0	43.2

Table 1. Average recognition error rate for Experiments I – III with d_c^1 on both Extended YDB and CMU-PIE.

When the probe pose for each subject is present in the gallery, the algorithm is able to come up with a good representation in the gallery that models the pose in the probe. However, when the probe pose is not found in the gallery, the algorithm can only use the variations in the gallery to try and come up with a representation that matches the probe pose as closely as possible. This observation is built upon the fact that a subject's pose manifold is non-linear in its standard representation and we are using methods that are linear in nature.

Fig. 2. Top: first principal vector for a sample probe. Bottom: first principal vector for the correct gallery set that the algorithm identifies for experiments I, II , and III from left to right. The first principal angle between the top and bottom vector is 0.066, 0.069, and 0.269 radians, from left to right.

To further understand which viewpoints are difficult to handle, we look specifically at the individual error rates for each viewpoint in CMU-PIE in Table 2. We note that recognition results for probe poses c07, c09, c25, and c31 are not reported in studies [10, 7] and we observed the highest error rates on these poses in our experiments. Without consideration of those 4 poses, our error rate is approximately 26.9% for CMU-PIE in Experiment III. We suspect the reason for the degradation of the algorithm's performance for these 4 poses is due to the variation in depth of field on top of the actual pose variation.

We also conducted experiments where the only variations in the gallery and probe is the viewpoint. For each probe and gallery set in E-YDB and CMU-PIE, we randomly chose 4 and 6 non-overlapping poses and calculated their mirror images to create sets of 8 and 12 distinct pose images, respectively. Then for each $\alpha = 1, \ldots, s_0$, we solve $P_{s^*} = \arg\min_s d(G_s, P_\alpha)$, where

Fig. 3. (c) The first principal vector for a sample probe. (f) The first principal vector for the incorrect gallery set that the algorithm identifies for experiments III. (a),(b) First principal vector for a sample probe set. (d),(e) First principal vector of a sample gallery set of a different subject from the ones in (a) and (b). The first principal angle between the top and bottom vector is 0.731, 0.275, and 0.369 radians, from left to right.

pose	c02	c05	c07	c09	c11	c14	c22
error (%)	13.4	31.3	83.6	73.1	0	1.5	23.9
pose	c25	c27	c29	c31	c34	c37	
error (%)	82.1	22.4	16.4	80.6	76.1	56.7	

Table 2. Average break-down recognition error rate for each pose in Experiment III using d_c^1 on CMU-PIE. We observe that some pose subsets perform much better than others.

$$G_s = \bigcup_{\substack{p \in P_G \\ i=\text{frontal}}} \{J_{s,p,i}\},$$

and P_G denotes the set of poses associated with this gallery set. We repeat this experiment 10 times to create a total of $10 \times s_0$ probe sets. The average error rate is 32.1% and 19.4% for E-YDB and CMU-PIE, respectively. We suspect the reason why the error rate for CMU-PIE is smaller than it is for E-YDB is because there are more pose variations in CMU-PIE, hence creating a better sampled characterization for the subjects. This observation supports the claim that the proposed algorithm captures the characteristics within a family of patterns and can be extended to handle other general object recognition problems.

3.5. Computational complexity

The algorithm for computing principal angles between a pair of subspaces comprises of two major steps: QR-decomposition of the representation matrices and SVD of the inner product of the Q matrices. The MATLAB qr command is based on Householder reflections. For a general m-by-n matrix, QR-decomposition using Householder reflections costs $2n^2m -$

1-4244-1549-7/07/$25.00

$\frac{2}{3}n^3$ flops. For the same size matrix, the MATLAB svd command costs $4n^2(m - \frac{1}{3}n)$ flops to reduce it to a bidiagonal form using Householder reflections. If only singular values are required, it costs just $O(n^2)$ for the rest of the operations. Therefore, a MATLAB svd routine costs $4n^2(m - \frac{1}{3}n) + O(n^2)$ flops to compute the singular values of a $m \times n$ matrix. Thus, most of the computation is in calculating the QR-decomposition of the gallery sets, which can be computed *a-priori* off-line.

On a 2.8GHz AMD Opteron processor, it takes approximately 0.4 seconds to do a qr on a single probe set of size 147167×10 and 0.25 seconds to do an economy size SVD of a matrix of size 120×10. Therefore, it costs about 0.65 seconds to compare a single pair of probe and gallery sets with the set-to-set paradigm (in Experiment III on CMU-PIE). Table 3 shows the computational speed for a few state-of-the-art algorithms along with the proposed algorithm. Note that our algorithm is significantly faster if the image resolution is smaller. Recall, we purposely omit the preprocessing stage in order to reflect the robustness of the algorithm.

4. RELATED WORK

We review a few start-of-the-art models here to compare and contrast with the experimental results shown in Section 3.4. Works of [18, 13, 19] also used the set-to-set framework to solve a general object recognition problem under varying illumination and viewpoints. However, variations in illumination is normalized away before identification. We will focus on algorithms that model both the illumination and pose variations in the following discussions.

A work which utilizes joint information given by variations in both illumination and pose and that is related to our work is given by Belhumeur et al. [9]. They developed a generative procedure that gives rise to a representation based on pose-specific illumination cones for each face class. A single face representation consists of multiple pose-specific illumination cones that are approximated linearly to capture over 99% of the variability. Recognition of a test image is performed by first normalizing its vector representative to unit length and then computing its distance to each face representation. This way, the pose and identity information of the test image can be revealed. When tested on the Yale Face Database B [9], the average error rate reported in [9] is about 2.9% out of 4050 (45 illuminations $\times$ 9 poses $\times$ 10 subjects) images tested. The algorithm performs the worst on extreme illumination and pose conditions with 12.6% error rate out of 420 (14 illuminations $\times$ 3 poses $\times$ 10 subjects) images tested. Note that the most extreme illumination conditions were not even considered and it is reported in [20] that this is the best result obtained on YDB. Further note that the illumination cone method uses a single-to-many classification scheme, which is fundamentally different from the one afforded by the set-to-set method proposed in this paper.

Another example by Gross et al. [8] uses the concept of light field [3] to handle pose variations and Fisher Discriminant Analysis (FDA) to handle the illumination variations. The concept of light-field is similar to FDA where a set of "ideal" eigen light-fields can be learned from a collection of training images from which the identities are not necessarily found in the testing images. Images in the gallery and probe are then written as linear combinations of the known eigen light-fields with appropriate weights. A probe identity x is then assigned to the identity in the gallery whose eigen light-field representation is the closest to that of x (in the Euclidean metric). The classification is based on many-to-many image comparisons using correlations where any number of gallery and probe images per subject are allowed. Better classification outcomes are achieved as the number of images used in both gallery and probe increases.

When tested on the CMU-PIE Database, the average error rate for the Fisher light-field method is about 53%, while it is 59% for the eigen light-field method when the gallery set consists of simply the frontal pose and frontal illumination. The pose variations in CMU-PIE are significantly more difficult to recognize compared to those in the Extended YDB. The error rate is improved slightly when the variation of illumination is handled by a Lambertian reflectance model [7] with 47% in the most difficult case. In all the cases above, pose and illumination conditions for the probes are different than the ones from the gallery. Moreover, in all of the methods reported above, a significant amount of training is required.

It is reported in a recent survey paper [20] that the best recognition result on the "lights" subset of PIE is achieved by a 3D Morphable Model [10] when using only front, side, and profile views in both gallery and probe set. However, there are two limitations to the practical version of the 3D morphable model. Training of the faces are required in order to build a 3D model and it is necessary to manually select 7 landmark points on probe images to provide a good estimate of 3D pose.

5. DISCUSSION

We have presented an approach to extract geometric characteristics of families of patterns that are viewed under varying conditions, such as pose and illumination. This is an extension to the previous work in [5], where only variations in illumination were considered. We observe that both the CMU-PIE and Extended YDB may be classified with high accuracy if samples of the pose are available in the gallery. This suggests that sufficient sampling of the variations is important and that this is more critical for pose than illumination (likely due to the non-linearity of the pose manifold).

In summary, the algorithm works under the assumption that multiple images of a class under each variation are available. By realizing sets of images as points on a Grassmannian, we employ a geometric perspective for computing metrics that compare subspaces and extract neighborhood informa-

1-4244-1549-7/07/$25.00

	3DMM	Illumination Cone	Set-to-Set
Data set used	"lights" of PIE	YDB	"illum" of PIE
Image resolution	200×200	42×36	367×401
Identification time for one probe	2.5 min. Pentium IV 2.0GHz	2.5 sec./gal. ind. Pentium II 300MHz	0.65 sec./gal. ind. AMD Opteron 8220SE 2.8 GHz

Table 3. Computational speed of two state-of-the-art face recognition algorithms and the proposed set-to-set algorithm. Given the disparity in processors and in image resolution, care must be exercised in interpreting CPU time.

tion. The algorithm requires no data preprocessing or training and can be trivially parallelized for large data sets. What we have shown here is a blueprint for set-to-set image comparison that can be extended to any general object recognition problem where families of patterns reside in characteristic subspaces.

6. REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] T. Riklin-Raviv and A. Shashua, "The quotient image: Class based re-rendering and recognition with varying illuminations," *PAMI*, vol. 23, no. 2, pp. 129–139, 2001.

[3] R. Gross, I. Matthews, and S. Baker, "Eigen light-fields and face recognition across pose," in *AFGR*, May 2002, pp. 1–7.

[4] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *PAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[5] J.-M. Chang, J.R. Beveridge, B. Draper, M. Kirby, H. Kley, and C. Peterson, "Illumination face spaces are idiosyncratic," in *IPCV*, June 2006, vol. 2, pp. 390–396.

[6] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible illumination conditions," *IJCV*, vol. 28, no. 3, pp. 245–260, July 1998.

[7] S. Zhou and R. Chellappa, "Image-based face recognition under illumination and pose variations," *J. Opt. Soc. Am. A*, vol. 22, no. 2, pp. 217–229, Feb 2005.

[8] R. Gross, I. Mattews, and S. Baker, "Appearance-based face recognition and light-fields," *PAMI*, vol. 26, no. 4, pp. 449–465, April 2004.

[9] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *PAMI*, vol. 23, no. 6, pp. 643–660, 2001.

[10] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *PAMI*, vol. 25, no. 9, pp. 1063–1074, 2003.

[11] O. Yamaguchi, K. Fukui, and K. Maeda, "Face recognition using temporal image sequence," in *AFGR*, 1998, pp. 318–323.

[12] L. Wolf and A. Shashua, "Learning over sets using kernel principal angles," *JMLR*, vol. 4, no. 10, pp. 913–931, 2003.

[13] T-K. Kim, J. Kittler, and R. Cipolla, "Learning discriminative canonical correlations for object recognition with image sets," in *ECCV*, 2006, pp. 251–262.

[14] J. Conway, R. Hardin, and N. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, pp. 139–159, 1996.

[15] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1999.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, third edition, 1996.

[17] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *PAMI*, vol. 25, no. 12, pp. 1615–1618, 2003.

[18] A. Fitzgibbon and A. Zisserman, "Joint manifold distance: a new approach to appearance based clustering," in *CVPR*, 2003, pp. 26–36.

[19] M. Nishiyama, M. Yussa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi, "Recognizing faces of moving people by hierarchical image-set matching," in *Biometrics07*, 2007, pp. 1–8.

[20] S. Romdhani, J. Ho, T. Vetter, and D. Kriegman, "Face recognition using 3-D models: Pose and illumination," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1977–1999, 2006.

1-4244-1549-7/07/$25.00