

Face recognition using optimal linear components of range images[☆]

Anuj Srivastava^{a,*}, Xiuwen Liu^b, Curt Heshel^b

^aDepartment of Statistics, Florida State University, Tallahassee, FL 32306, USA

^bDepartment of Computer Science, Florida State University, Tallahassee, FL 32306, USA

Received 19 March 2005; accepted 29 July 2005

Abstract

This paper investigates the use of range images of faces for recognizing people. 3D scans of faces lead to range images that are linearly projected to low-dimensional subspaces for use in a classifier, say a nearest neighbor classifier or a support vector machine, to label people. Learning of subspaces is performed using an optimal component analysis, i.e. a stochastic optimization algorithm (on a Grassmann manifold) to find a subspace that maximizes classifier performance on the training image set. Results are presented for face recognition using FSU face database, and are compared with standard component analyses such as PCA and ICA. This provides an efficient tool for analyzing certain aspects of facial shapes while avoiding a difficult task of geometric surface modeling.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Face recognition; Range imaging; Optimal component analysis; Nearest neighbor classifier; Grassmann manifold

1. Introduction

The problem of recognizing humans using images of their faces has become increasingly relevant in recent times. A large number of papers and efforts have been dedicated towards solving this problem. Depending on the operational nature of imaging devices, they can capture several aspects of humans faces, such as landmark geometry, texture variations on faces, skin reflectance, skin emissivity and facial shapes. An important issue is to find an aspect that: (i) differentiates people consistently, (ii) manifests itself well even in low-quality images, and (iii) leads to efficient algorithms. A long-term strategy should be to model all the physical factors that lead to variability among face images, factors such as shapes, textures (or reflectance functions), illumination models, clutter, obscuration and motion relative to the camera. This will address many of the challenges that arise in current analysis when only a subset of these factors is included in the model. In the meantime, one needs to develop tools for representation, analysis and inferences

for individual factors leading up to derivation of efficient algorithms. Although majority of current research is focused on appearance-based recognition of faces, using visible spectrum images ([1–3]) and infrared images ([4, 5]); the shapes of the facial surfaces have not received a similar attention. A statistical analysis of shapes of facial surfaces, including specifications of shape spaces, metrics, distributions, and testing, can prove to be an important tool in facial recognition. The use of 3D shapes of faces in recognition is rather recent and limited: [6] models the geometry of facial shapes and skin reflectance to capture face variation, [7,8] use features from 2D images and 3D face scans to characterize a face, and [9] uses stereo data to analyze 3D shape of faces. Current algorithms involving 3D shape analysis lack efficiency and speed for real-time applications, and there is a need to focus on efficient, albeit approximate, methods.

An important question is: How to capture, represent, and analyze shapes of facial surfaces? To capture or measure facial surfaces, 3D scanning has proven to be a widely acceptable tool. However, the choice of representation and analysis is much more difficult. If one can represent facial surfaces as discrete meshes, then shape analysis involves comparing discrete meshes modulo shape preserving transformations such as rotation, translation, and scaling. Although this approach seems fundamental and appropriate, it remains to be investigated

[☆] This paper is an extension of work first presented at ECCV 2002.

* Corresponding author. Tel.: +1 850 644 8832; fax: +1 850 644 5271.
E-mail address: anuj@stat.fsu.edu (A. Srivastava).

carefully. It is proving difficult to establish a framework, where any two surfaces can be compared, matched, and analyzed geometrically. One solution is to use landmark-based approach for representing and analyzing shapes ([10]) although it is difficult to detect and register these landmarks automatically. Another idea is to embed the surfaces in \mathbb{R}^3 and to compare them using diffeomorphisms of \mathbb{R}^3 ([11]); a major limitation here is the computational cost associated with diffeomorphisms. Seeking efficient algorithms for real-time applications, we form 2D *range images* of 3D facial surfaces, as viewed from a specified direction and distance, and compare the resulting images. Range images capture the depths of facial surfaces in the direction perpendicular to the image plane. Although range pixels relate to the shapes of facial surfaces, they do not contain sufficient information to completely characterize shapes. Despite this lack of completeness in characterizing shapes, range images allow us to use tools from image analysis in comparing 3D objects. In this paper, we investigate the use of range images in face recognition.

Using output of a 3D scanner, one can generate range images of faces using range mapping and orthographic projections. Once the range images are generated, they are registered and pre-processed. Next question is how to compare and analyze them in a statistical fashion. Linear projections of images have frequently been used to reduce image dimensions and provide efficient algorithms for face recognition; examples include principal component analysis (PCA) ([12,13]), independent component analysis (ICA) ([14]), Fisher's discriminant analysis (FDA) ([15]), etc. The use of principal components in analyzing 3D scans of faces has also been studied previously ([6]). Although such linear representations satisfy certain optimality criteria, they may not necessarily be optimal for a specific application at hand (for empirical evidence, see e.g. [15,16]). A recent paper ([17]) presents a technique for finding *linear representations of images that are optimal for specific tasks and specific datasets*, i.e. instead of choosing standard projections, one chooses a projection that is matched to the given problem. The search for optimal linear representation, or an optimal subspace, is based on a stochastic optimization process that maximizes a pre-specified performance function over the set of all subspaces. Since this set of all subspaces (known as a Grassmann manifold) is not a vector space, the optimization process has to account for the curved geometry of Grassmannian. In this paper, we investigate the use of optimal component analysis in face recognition using range images.

The remainder of this paper is organized as follows. Formation and pre-processing of facial range images is described in Section 2, and an algorithm for computing optimal linear components is presented in Section 3. Some experimental results using FSU face database are presented in Section 4.

2. Range images of facial shapes

We start by describing the process of generating range images from a 3D scanner data, their registration and other pre-processing.

2.1. Generation of range images

There exist range scanners that can record and provide geometries of 3D objects in form of range images directly. However, some 3D scanners, including a Minolta vivid 700 scanner used in experiments presented here, output data only in form of polygonal meshes. Therefore, one needs to pre-process this data into range images before image-based recognition techniques can be applied.

The dataset used in these experiments was collected in a controlled imaging environment as shown in the top panel of Fig. 1. Subjects were imaged in a closed room with fixed uniform illumination and using markers to align their faces. They were asked to stay in a predetermined position and orientation with respect to the camera resulting in a rough, global registration of facial surfaces. Each subject was asked to form six different facial expressions: neutral, smile, frown, angry, squint, and scared, and faces were scanned for each case, as shown in the bottom panels of Fig. 1. To avoid missing data, the subjects were asked to close their eyes during scanning. The scanner outputs a discrete mesh, i.e. a set of vertices in \mathbb{R}^3 and edges connecting the neighboring vertices to form triangles. Next we describe the formation of range images from these triangulated meshes.

A range image is a rectangular array of pixels in an image plane, with pixels values being proportional to the distance (or depth) of the nearest surface in direction perpendicular to the image plane. Although these depth values, or Z-coordinates, are given for vertices present in the mesh, they may not project uniformly to pixel locations in the image plane, thus leaving holes in a range image. Our approach to forming image is essentially range mapping: for each pixel in the image, we traverse along a ray orthogonal to the image plane and seek the triangle (in the mesh) nearest to the image. Then, the Z-coordinate of the point of intersection, between the ray and this triangle, provides the pixel value. Since the Z-coordinate of this point may not be explicitly available, we linearly interpolate between the Z-coordinates of the vertices of that triangle.

For simplification, the actual implementation is reversed, i.e. for each triangle in the mesh, we find pixels in the image plane that are interior to the projection of that triangle using the line crossing algorithm ([18]). Then, for each pixel inside the projected triangle, we compute the Z-coordinate using a linear interpolation of the Z-coordinates of three vertices of that triangle. Care is taken to ensure that the triangle (in the mesh) does not project to a straight line in the image plane. In this way, each triangle in the mesh is traversed and orthographically projected onto the image plane. Finally, if multiple triangles assign values to the same



Fig. 1. Data capture. Subjects stay in a predetermined position and orientation with respect to the 3D scanner. Each subject was scanned for six different facial expressions.

pixel, then the smallest pixel value (denoting the nearest triangle) is selected. After all triangles are traversed, most of the triangles are assigned range values although some pixels may be still left unassigned.

2.2. Registration of range images

Even though the data are captured in a controlled environment, there are still some pose variations in the scanned faces. Assuming that this variability is small, we try to correct it using registration techniques in the image plane. All registrations are performed on the 2D range images, avoiding the computational complexity associated with 3D registration. In other words, we ignore the mismatches, if any, in the direction orthogonal to the image plane. In the image plane, we perform landmark-based registration for rotation, translation, and depth variability, for each range image. Since the use of landmarks here is simply for registration of images, and not comparisons, the choice of landmarks is relatively simple. We have chosen to register

images using the tip and the bridge of the nose. In each image, the nose tip is found as the pixel closest to the camera, i.e. smallest range value. We choose the pixels with the smallest intensities in several of the rows above the nose tip, and fit a straight line through these pixels to discover the bridge of the nose.

1. *Translation in \mathbb{R}^2 .* Each range image is translated in the image plane so that the tip pixel is moved to the image center. This process removes the translational variability.
2. *Rotation.* To force rotational alignment, the line representing the bridge of the nose is made vertical as shown in the left panels of Fig. 2. This removes the rotational variability in the images.
3. *Translation in Z-coordinate.* Finally, the depth adjustment restricts variability in the subject's distance from the camera during data collection. The pixel value at the image center, i.e. the distance between the nose tip and the camera, is forced to be the same value in all images.

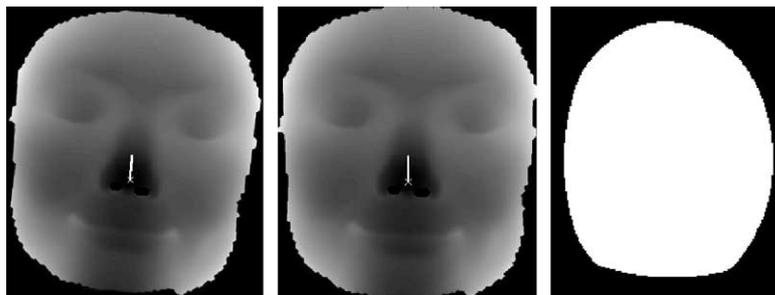


Fig. 2. The asterisk indicates the tip of the nose and the white line indicates the bridge of the nose. Images show range images before (left panel) and after (middle panel) rotational alignment. Right panel shows the mask imposed on range images to crop periphery.

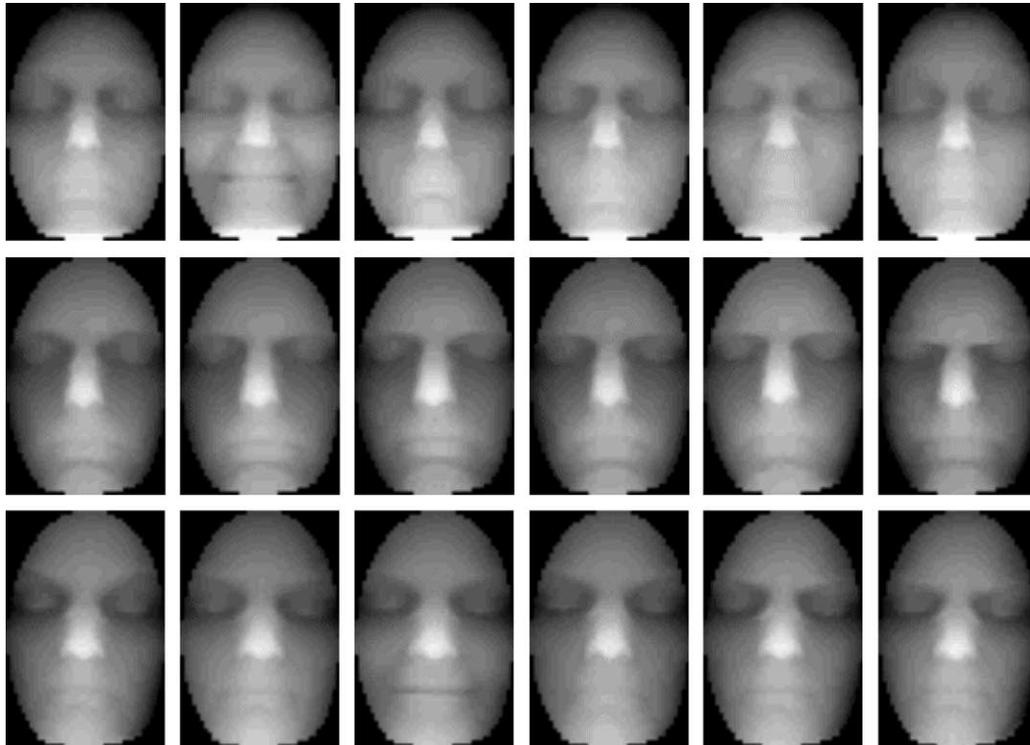


Fig. 3. Range images of same faces under different facial expressions. Top row shows subject 2, middle row shows subject 13 and bottom row shows subject 19.

Through these steps, we obtain registered range images of all subjects under all facial expressions.

2.3. Cropping rough areas

Although range images are now well aligned, there are still sources of variability that we want to avoid. For example, subjects' hair if visible in the image leads to unreliable data, and we want to remove them from the image. Similarly, other parts such as ears, bottom of the chin, etc. should also be removed to provide consistency.

- (1) *Cropping images.* To remove undesired variability in the peripheries, we simply mask the range image with a filter shown in the right panel of Fig. 2. The basic idea is to remove hair, chin, and other parts of faces that are not scanned with high precision by the scanner. It must be noted here that we have used the same filter on all images and, for a population with large variations in facial sizes, this may lead to irregular cropping. In future, we plan to crop range images using filters that are adaptive to facial sizes.
- (2) *Patching.* It is possible to have pixel locations in range images that do not have an assigned value, i.e. there may be holes in the images. We patch a hole using linear interpolation between the neighboring pixel values, as long as they themselves are not holes.

This completes generation and pre-processing of range images, and now they are ready for analysis. For later

computational efficiency, we downsample these images to smaller sizes. Figs. 3 and 4 show some examples of the resulting range images. The three rows in Fig. 3 show examples of range images of people under different facial expressions. The top row shows subject 2, the middle row shows subject 13, and the bottom row shows subject 21, under six facial expressions each. Fig. 4 shows range images of first 24 subjects under the same—neutral—facial expression.

In these range images, lighter pixels indicate parts of the face closer to the camera while darker pixel values indicate parts of the face away from the camera, and completely black pixels indicate no data. Note that the imaging set-up ensures that the tip of the nose is always nearest to the camera.

3. Optimal component analysis

Next, we describe the use of these range images in face recognition using optimal component analysis. Our analysis of range images for face recognition is based on linear projections of images to smaller subspaces, and a statistical analysis of the resulting coefficients. In the past, the choice of projection has proven to be an important factor in the recognition performance. Instead of choosing a standard projection, such as PCA, ICA, or FDA, we tailor the choice of projection to this application by choosing a projection that maximizes the performance face recognition on the training data.

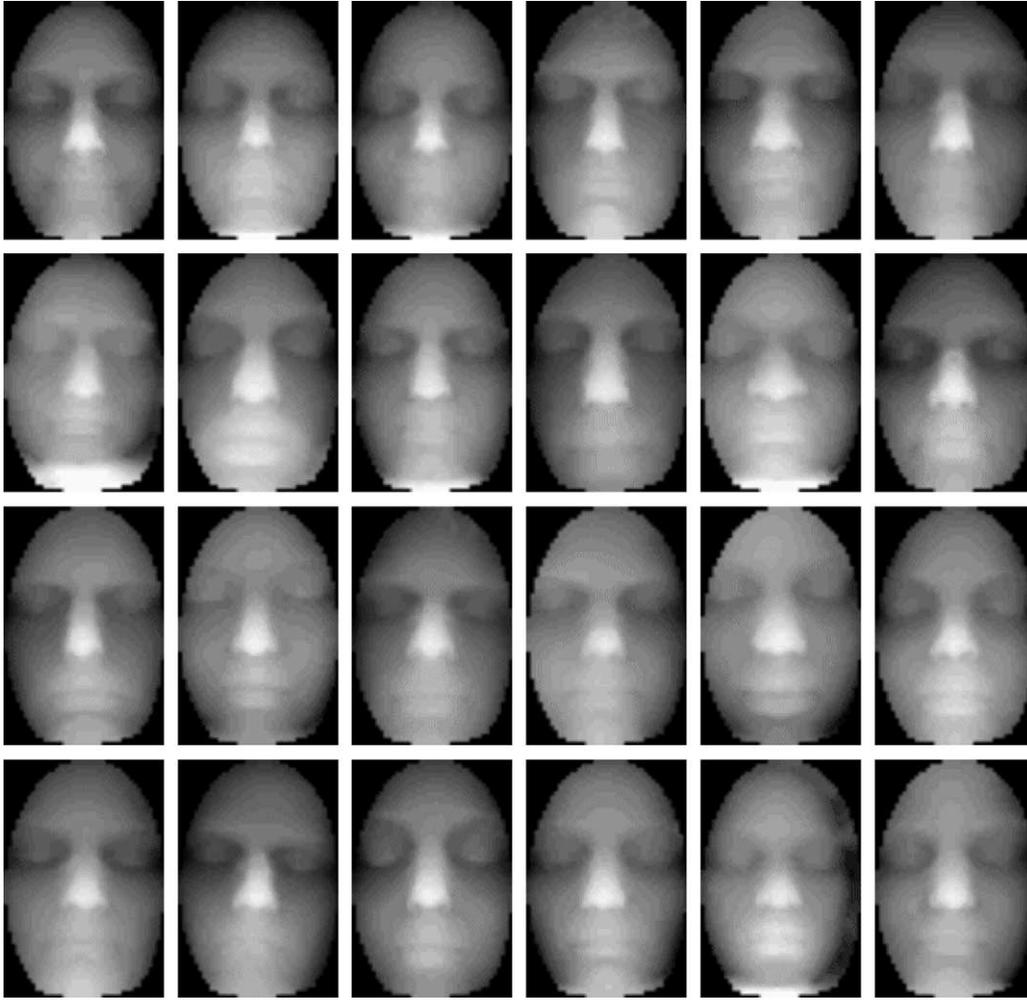


Fig. 4. Range images of first 24 subjects under the neutral facial expression.

We start with a mathematical formulation of the problem. Let $U \in \mathbb{R}^{n \times k}$ be an orthonormal basis of a k -dimensional subspace of \mathbb{R}^n , where n is the size of an image and d is the required dimension of the optimal subspace (generally $n \gg k$). For an image I , considered as a column vector of size n , the vector of coefficients is given by $a(I, U) = U^T I \in \mathbb{R}^k$. Images are compared via their coefficients under the metric $d(I_1, I_2; U) = \|a(I_1, U) - a(I_2, U)\|$, where $\|\cdot\|$ denotes the 2-norm.

For finding optimal U , we need to select a classifier and a recognition performance measure F . In this paper, we choose a nearest neighbor classifier due to its simplicity and popularity. However, other classifiers can easily be substituted instead in this framework. Let there be C classes to be recognized from the images; each class has m_{train} training images (denoted by $I_{c,1}, \dots, I_{c,m_{\text{train}}}$). In order to utilize a gradient-based algorithm, F should have continuous directional derivatives. To ensure that we define $\rho(I_{c,i}, U)$ to be the ratio of the between-class-minimum distance and within-class minimum distance of a training image from class c indexed by i in the leave-one out sense, given by $\rho(I_{c,i}, U) = (\min_{c' \neq c, j} d(I_{c,i}, I_{c',j}; U)) / (\min_{j \neq i} d$

$(I_{c,i}, I_{c,j}; U) + \varepsilon_0)$, where $d(I_1, I_2; U)$ is defined earlier, and $\varepsilon_0 > 0$ is a small number to avoid division by zero. Then, define F according to

$$F(U) = \frac{1}{C m_{\text{train}}} \sum_{c=1}^C \sum_{i=1}^{m_{\text{train}}} h(\rho(I_{c,i}, U) - 1), \quad (1)$$

where $h(\cdot)$ is a monotonically increasing and bounded function. In our experiments, we have used $h(x) = 1/(1 + \exp(-2\beta x))$, where β controls the smoothness of F . Note that $I_{c,i}$ is classified correctly in the leave-one out sense according to the nearest neighbor rule under U if and only if $\rho(I_{c,i}, U) > 1$. It follows that F is precisely the leave-one out recognition performance of the nearest neighbor classifier when $\beta \rightarrow \infty$.

Under this formulation, $F(U) = F(UH)$ for any $k \times k$ orthogonal matrix H as the distance $d(I_1, I_2; U) = d(I_1, I_2; UH)$; the choice of 2-norm in $d(I_1, I_2; U)$ allows for this equality. In other words, F depends on the subspace spanned by U but not on the specific basis chosen to represent that subspace. Therefore, our search for optimal representation(s) is on the space of k -dimensional subspaces rather than on their bases.

Let $\mathcal{G}_{n,k}$ be the set of all k -dimensional subspaces of \mathbb{R}^n ; it is called a Grassmann manifold. It is a compact, connected manifold of dimension $k(n-k)$. An element of this manifold, i.e. a subspace, can be represented either by a basis (non-uniquely) or by a projection matrix (uniquely). Choosing the former, let U be an orthonormal basis in $\mathbb{R}^{n \times k}$ such that $\text{span}(U)$ is the given subspace of \mathbb{R}^n . Let $[U]$ denote the set of all the orthonormal bases that span the same subspace, i.e. $[U] = \{UH | H \in \mathbb{R}^{k \times k}, H^T H = I_k\} \in \mathcal{G}_{n,k}$. The problem of finding optimal linear subspaces for recognition becomes an optimization problem:

$$[\hat{U}] = \underset{[U] \in \mathcal{G}_{n,k}}{\operatorname{argmax}} F([U]).$$

Since the set $\mathcal{G}_{n,k}$ is compact and F is a smooth function, the optimizer $[\hat{U}]$ is well defined. $[\hat{U}]$ may not be unique, i.e. it may be set-valued rather than being point-valued.

A recent paper [17] describes a numerical procedure for approximating $[\hat{U}]$ using a stochastic gradient algorithm. The basic idea is to construct a Markov chain that seeks high F -valued points in $\mathcal{G}_{n,k}$. It does so by using randomly perturbed versions of the gradient directions to find candidates for updating the chain; these candidates are accepted and rejected according to a probability that depends upon F . The algorithm is repeated here for reader convenience.

Algorithm 1. MCMC simulated annealing: Let $X(0) = [U_0] \in \mathcal{G}_{n,k}$ be any initial condition. Set $t=0$.

(1) Calculate the directional derivative of F at X_t according to:

$$A(X_t) = Q_t^T \left(\sum_{i=1}^k \sum_{j=k+1}^n \alpha_{ij}(X_t) E_{ij} \right) \in \mathbb{R}^{n \times n} \text{ and where} \quad (2)$$

$$\alpha_{ij}(X_t) = \lim_{\varepsilon \downarrow 0} \left(\frac{F([Q_t^T e^{\varepsilon E_{ij}} J]) - F(X_t)}{\varepsilon} \right) \in \mathbb{R}.$$

Here J is an $n \times k$ matrix made up of the first k columns of the $n \times n$ identity matrix and Q_t is an $n \times n$ orthogonal matrix such that $Q_t^T X_t = J$. Given X_t , Q_t can be found efficiently in $O(nk^2)$ computations using Householder reflection as described in [19]. E_{ij} is an $n \times n$ skew-symmetric matrix such that: for $1 \leq i \leq k$ and $k < j \leq n$

$$E_{ij}(u, v) = \begin{cases} 1, & \text{if } u = i, v = j \\ -1, & \text{if } u = j, v = i \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

(2) Generate $k(n-k)$ independent realizations, w_{ijs} , from standard normal density. Using the value of X_t ,

calculate a candidate value Y according to:

$$dA_t = A(X_t)\Delta + \sqrt{2\Delta T} \sum_{i=1}^k \sum_{j=k+1}^n w_{ij} E_{ij}, \quad (4)$$

$$Y = Q_t^T \exp(\Delta dA_t) J,$$

The operation $\exp(\Delta dA_t)$ can be performed in $O(nk^2)$ operations by exploiting the structure of sparse skew-symmetric matrix dA_t , as described in [19].

- (3) Compute $F(Y)$, $F(X_t)$, and set $dF = F(Y) - F(X_t)$.
- (4) Set $X_{t+1} = Y$ with probability $\min\{\exp(dF/T_t), 1\}$, else set $X_{t+1} = X_t$.
- (5) Set $T_{t+1} = T_t/\gamma$, $t = t + 1$, and go to Step 1.

Here $\gamma > 1$ is the cooling ratio for simulated annealing with a typical value of 1.0025. This algorithm generates a Markov chain $\{X_t\}$ in $\mathcal{G}_{n,k}$ whose convergence properties are briefly discussed in [17,20]. In this paper, we investigate its application to range image analysis.

4. Experimental results

Although the original database consists of 82 subjects, certain imaging errors have ruled out the use of a subset of these images here. As a result, we have used only 67 subjects, at six different facial expressions each, in the results described here. Two types of results are presented: (i) improvement in recognition performance using Algorithm 1, and (ii) improvement in resulting average posterior entropy using Algorithm 1.

4.1. Recognition performance

As the first result, we demonstrate the search for optimal components by maximizing F on the training data using Algorithm 1. In this experiment, the first three images of each subject were considered as training, while the remaining three were used for testing: $m_{\text{test}}=3$, $m_{\text{train}}=3$, and $C=67$. As given in Eq. (1), in this paper the learning process described in Algorithm 1 uses only the images in the training set and the recognition performance is then evaluated on the disjoint test set.

Fig. 5 shows some sample evolutions of Algorithm 1 to find the optimal linear representations with $n=2501$ and $k=10$. For each case, the top panels plot the evolution of leave-one-out performance $F(X_t)$ on the training data using X_t as the projection. Although, the performance is optimized on the training data, its value on the test data is more important; we have evaluated the recognition performance on the test data and have plotted in the corresponding lower panels. The four cases columns differ in their initial conditions: (a) uses PCA, (b) uses ICA, (c) uses first d coordinate axes, and (d) uses a random subspace to initialize

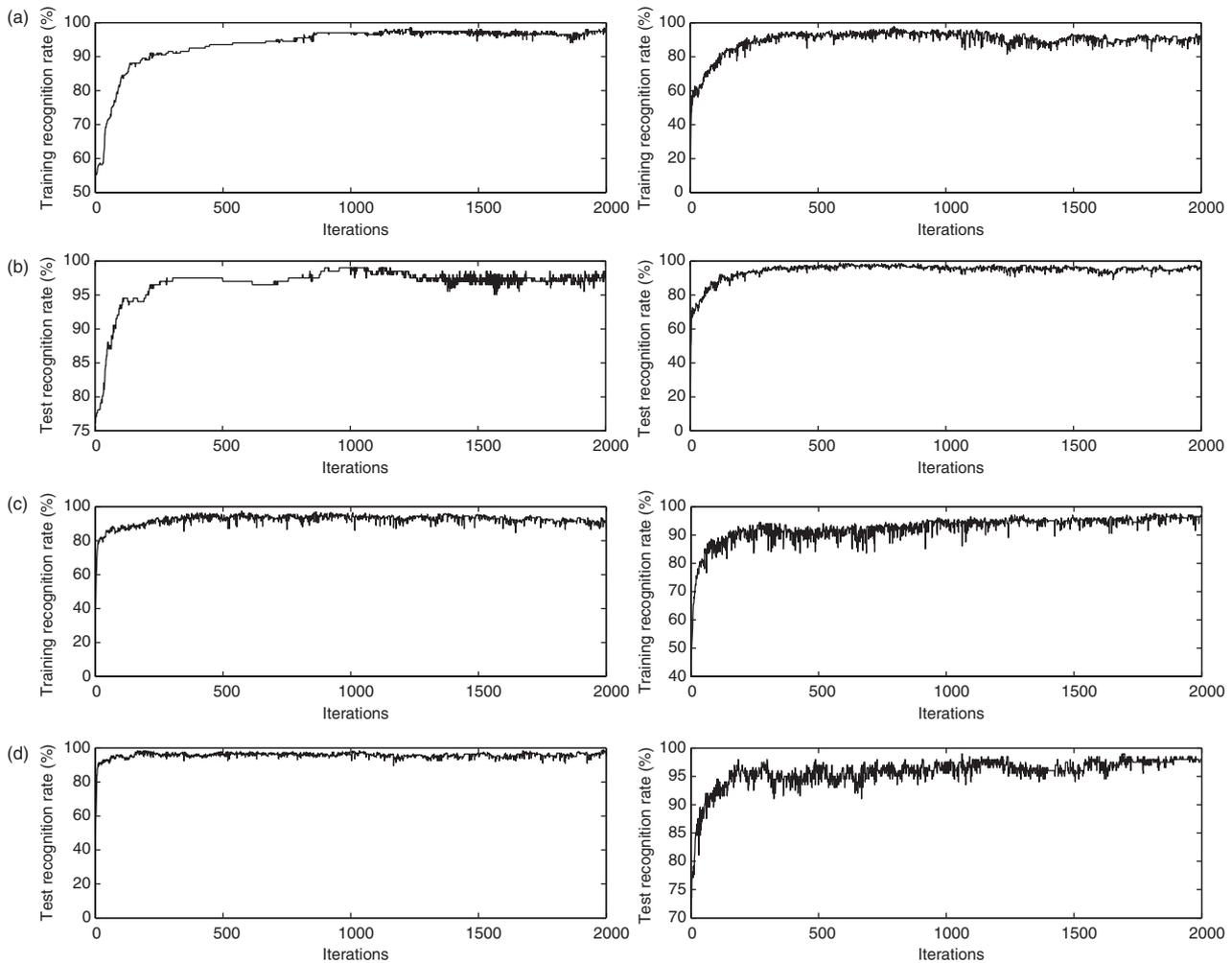


Fig. 5. Evolution of $F(X_t)$ versus t under Algorithm 1 for different initial conditions. For each case, the top panel plots F on the training data while the lower panel plots F on the test data. The initial conditions are as follows: (a) PCA basis, (b) ICA basis, (c) basis specified by first d Euclidean axes, and (d) a random basis.

the algorithm. Here ICA is computed using the FastICA algorithm proposed by [21]. As an example, consider case (a): the PCA basis provides less than 60% recognition performance on the training data and 77% recognition performance on the test data. Under Algorithm 1, the optimal performance for this training data evolves close to 100%, and the corresponding performance of this optimal basis on the test data is found to be 99%. Similar improvement in performances is observed for other starting points. Experiments show that the optimal subspace is not unique and starting from different initial conditions leads to different final points in $\mathcal{G}_{n,k}$. However, the recognition performance for all these points is 99% or more, thus implying multiple global maxima of F . Recognition performances of optimal components on the test data for different initial conditions are listed in Table 1.

Although optimal components are derived assuming a nearest neighbor classifier, their performances on other classifiers also show a consistent improvement. To

demonstrate this, we have computed the recognition performance using a support vector machine after a linear projection. In case the projection is based on standard methods, i.e. PCA, ICA, etc., the performance is much lower than the performance obtained using optimal components. These results are also summarized in Table 1. For different initial conditions, the table lists the recognition performance on the test dataset under the two classifiers: nearest neighbor and SVM. The SVM is implemented based on the SVM-light package by Thorsten Joachims.¹ Here we use polynomial kernels with degree 2 and pairwise classification to handle multiple classes [22].

4.2. Posterior entropy

In addition to the recognition performance, we have also studied the entropy of the posterior distribution on the

¹ Obtained from <http://svmlight.joachims.org>.

Table 1
Recognition performance on the test data before and after using Algorithm 1

Initial condition	Initial F		Optimal F	
	NN class	SVM class	NN class	SVM class
PCA	77.61	75.62	99.00	92.54
FDA	94.53	89.55	99.50	90.55
ICA	3.48	10.45	99.00	88.06
ICA2	72.64	69.15	99.00	92.04
RCA	71.14	62.19	99.00	94.03
AXIS	13.43	12.94	99.00	89.55

subject space. Let $P(c|I')$ denote the posterior probability of a subject c given test image I' is

$$P(c|I') = \frac{1}{Z} \exp\left(\frac{-\{\min_{i=1}^{m_{\text{train}}} d(I', I_{c,i})^2\}}{\bar{d}^2}\right), \tag{5}$$

where $\bar{d} = \min_c \left(\min_{i=1}^{m_{\text{train}}} d(I', I_{c,i})\right)$,

and Z is the normalizer. The entropy associated with this posterior is given by:

$$H(P|I') = - \int_{c=1}^C \log(P(c|I'))P(c|I'), \tag{6}$$

and the average entropy is given by:

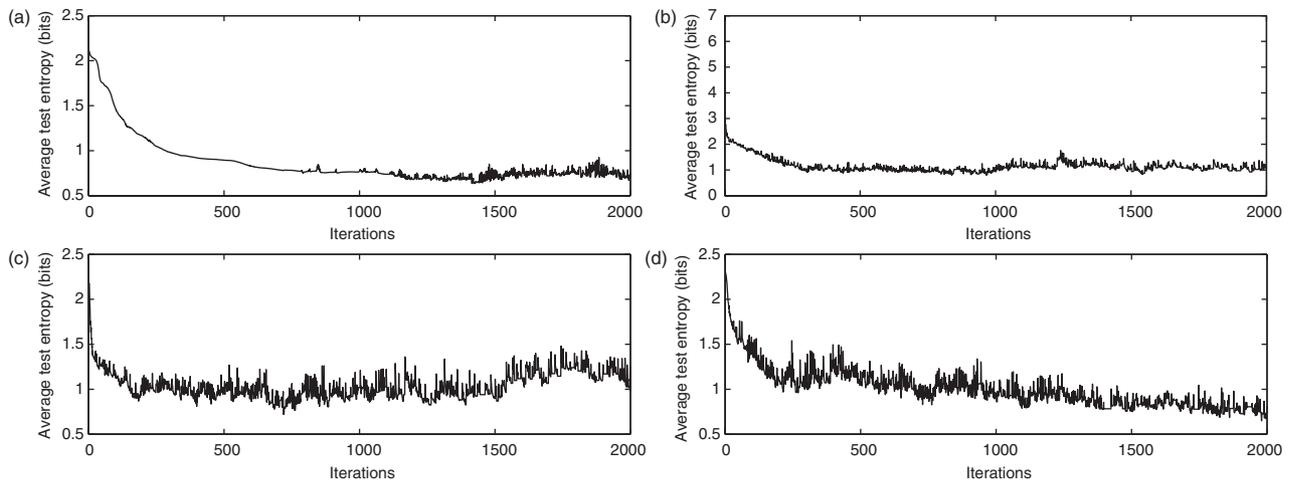


Fig. 6. Evolution of the average entropy $H(P)$ as the subspace changes according to Algorithm 1, under four different initial conditions.

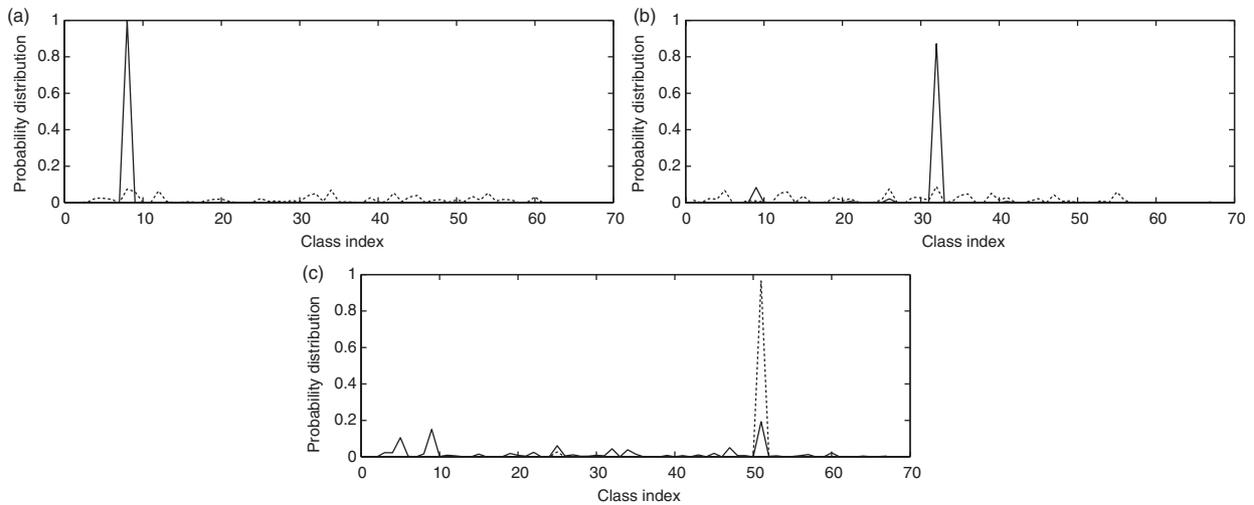


Fig. 7. Examples of probabilities before and after optimization.

$$H(P) = \frac{1}{Cm_{\text{test}}} \sum_{c=1}^C \sum_{i=1}^{m_{\text{test}}} H(P|I'_{c,i}),$$

where $I'_{c,i}$ denotes a test image.

Since the normalizing constant Z is not known, the average entropy can only be computed within a constant. A flatter posterior, with a larger entropy, can imply a larger probability of error, while a posterior concentrated around the correct subject, with a smaller entropy, generally implies smaller error probabilities. Fig. 6 shows the evolutions of average posterior entropy as Algorithm 1 progresses. The four plots correspond to the four initial conditions as earlier.

Fig. 7 shows some examples of the posterior probabilities before and after the optimization process. In each plot, the broken line shows the initial posterior and the solid line shows the final posterior. The concentration of solid lines around a subject denotes both improved recognition and reduced entropy. The first two cases are typical for our experiments, while the last panel shows the worst case result. In this case, the entropy actually goes up although the recognition based on highest probability subject will still be correct.

5. Summary

In this paper, we have described an efficient technique to compare certain aspects of facial shapes using ideas from image analysis. 3D scans of facial surfaces lead to 2D range images, which are projected linearly onto a subspace that maximizes ensuing recognition performance. This search for optimal subspace is based on a stochastic gradient algorithm on a Grassmann manifold, the set of all subspaces. To demonstrate effectiveness of this approach, experimental results from a FSU database of 67 people, scanned under six facial expressions each, are presented. Nearly perfect recognition performance is achieved for several different initializations of Algorithm 1.

Although the recognition performance using these range images is fairly high, it is difficult to predict the performance in larger studies. One application that seems promising is to use these range cameras in conjunction with the regular video cameras to improve the joint recognition performance.

Acknowledgements

This research was supported in part by grants NSF DMS-0101429, ARO DAAD19-99-1-0267, and NSF IIS-0307998. We are grateful to all the people who agreed to be subjects in this study.

References

- [1] A. Pentland, Looking at people: Sensing for ubiquitous and wearable computing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1) (2000) 107–119.
- [2] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of face: A survey, *Proceedings of the IEEE* 83 (5) (1995) 705–740.
- [3] Hallinan, P.W., Gordon, G.G., Yuille, A.L., Giblin, P., Mumford, D., 1999. Two- and Three- Dimensional Patterns of Face. A.K. Peters.
- [4] A. Srivastava, X. Liu, Statistical hypothesis pruning for face recognition from IR images, *Journal of Image and Vision Computing* 21 (7) (2003) 651–660.
- [5] X. Chen, P. Flynn, K. Bowyer, Pca-based face recognition in infrared imagery: baseline and comparative studies *IEEE Workshop on Analysis and Modeling of Faces and Gestures* 2003.
- [6] A.J. O'Toole, T. Vetter, V. Blanz, Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing, *Vision Research* 39 (1999) 3145–3155.
- [7] Y. Wang, C.S. Chua, Y.K. Ho, Y. Ren, Integrated 2d and 3d images for face recognition *11th International Conference on Image Analysis and Processing* 2001 pp. 48–55.
- [8] K. Chang, P. Flynn, K. Bowyer, Multi-modal 2d and 3d biometrics for face recognition *IEEE, Workshop on Analysis and Modeling of Faces and Gestures* 2003.
- [9] S. Malassiotis, F. Tsalakanidou, N. Mavridis, V. Giagourta, N. Gramma, A face and gesture recognition system based on an active stereo sensor *Proc. International Conference on Image Processing* 2001.
- [10] I.L. Dryden, K.V. Mardia, *Statistical Shape Analysis*, John Wiley & Son, 1998.
- [11] U. Grenander, M.I. Miller, Computational anatomy: An emerging discipline, *Quarterly of Applied Mathematics* LV14(1998) 617–694.
- [12] M. Turk, A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuro-science* 3 (1991) 71–86.
- [13] M. Kirby, L. Sirovich, Application of the Karhunen-Loeve procedure for the characterization of human faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1) (1990) 103–108.
- [14] P. Comon, Independent component analysis, a new concept?, *Signal Processing, Special issue on higher-order statistics* 36 (3) (1994) 287–314.
- [15] P.N. Belhumeur, J.P. Hefanpha, D.J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Patterns Analysis and Machine Intelligence* 19 (7) (1997) 711–720.
- [16] A.M. Martinez, A.C. Kak, PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2) (2001) 228–233.
- [17] X. Liu, A. Srivastava, K. Gallivan, Optimal linear representations of images for object recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (5) (2004) 662–666.
- [18] J. Arvo, *Graphics Gems II*, Academic Press, Inc., 1991.
- [19] K. Gallivan, A. Srivastava, X. Liu, P.V. Dooren, Efficient algorithms for inferences on grassmann manifolds *12th IEEE Workshop on Statistical Signal Processing* 2003 pp. 301–304.
- [20] C.P. Robert, G. Casella, *Monte Carlo Statistical Methods*. Springer Text in Stat 1999.
- [21] A. Hyvarinen, Fast and robust fixed-point algorithm for independent component analysis, *IEEE Transactions on Neural Networks* 10 (1999) 626–634.
- [22] U.H.G. Krebel, Pairwise classification and support vector machines in: B. Scholkopf, C.J.C. Burges, A.J.S. (Eds.), *Advances in Kernel Methods: Support Vector Learning*, The MIT Press, 2002 pp. 255–268.