Max–Planck–Institut
für biologische Kybernetik
Arbeitsgruppe Bülthoff

Technical Report No. 15 April 1995

A separated linear shape and texture space for modeling two-dimensional images of human faces

Thomas Vetter and Nikolaus Troje

Abstract

Human faces differ in shape and texture. This paper describes a representation of grey-level images of human faces based on an automated separation of two-dimensional shape and texture. The separations were done using the point correspondence between the different images, which was established through algorithms known from optical flow computation. A linear description of the separated texture and shape spaces allows a smooth modeling of human faces. Images of faces along the principal axes of a small data set of 50 faces are shown. We also reconstruct images of faces using the 49 remaining faces in our data set. These reconstructions are the projections of an image into the space spanned by the textures and shapes of the other faces.

This document is available as `/pub/mpi-memos/TR-15.ps.Z` via anonymous ftp from `ftp.mpik-tueb.mpg.de` or by writing to the Max–Planck–Institut für biologische Kybernetik, Spemannstr. 38, 72076 Tübingen, Germany.

1 Introduction

A natural description of objects which belong to the same object class typically evaluates the differences between corresponding features. This comparison can contain differences in the texture (i.e. the color or intensity) as well as in the shape of the objects. In this paper, we describe a fully automated system for separating texture differences in images of human faces from differences in their 2D-shape. Linear space models applied to both, the texture differences and the 2D-shape deformations allow a smooth and low-dimensional representation and enables a continous modeling of human faces. A representation separated in texture and 2D-shape differences has several applications which are outside the scope of this paper. We only want to mention some of them briefly. The linear model based on such a representation can be used for automatic face recognition in the same manner, but presumably with more efficiency, than using principal component analysis on pixel based representations. The representation further serves as a universal feature detector. If the location of an arbitrary feature is known in some reference face, it can be determined in all other faces, too. The representation also enables an estimation about the complexity (i.e. the dimensionality) of the object class "human faces". This knowledge can be used to establish a "linear object class" [10] of faces which can further be used to apply transformations corresponding to other orientations or expressions of a face or even to change its perceived age. Low dimensional representations will also play a role in data transmission, e.g. for teleconferencing.

The main scope of this paper is to develop a method which allows a continuous modeling of human faces. In image based face recognition and image representation, linear space models and especially principal component analysis became very important [12, 13, 9]. Principal component analysis is appropriate for normally distributed data sets which form a dense, convex body (in the best case an ellipsoid) in the underlying linear space. The principal components then yield the direction of the axis of the ellipsoid and the eigenvalues of the corresponding covariance matrix provide a measure for the variances of the data in these directions. In face recognition principal component analysis is usually applied to pixel based representations. It is obvious, that the resulting topology is not at all convex. In general, the average of two faces (corresponding to the point in the face space located just in the middle between the two faces) is not a normal looking face. It is clear, that this is due to the lack of a correct alignment of the faces. The average of a mouth and a nose can never become something sensefull. A proper alignment, however, is not possible applying only linear image transformations since faces differ in the proportions of distances between eyes, nose, mouth etc. To perform nonlinear image transformations, which would align eyes, nose and mouth, it is necessary to establish correspondence between the images of two face, which is not a trivial problem.

Human perception is very good in finding this correspondence. A common approach for establishing correspondence between two images thus uses our built-in system and requires the user to hand-select corresponding feature points, like the tip of the nose or the corners of the mouth. The areas between the distinct feature points are usually triangulated and linearly matched [3, 6]. Although very time consuming this method is widely used for morphing purposes. In contrast to this feature based approach are methods based on the image intensities or their gradients, mainly known from the optical flow literature [2]. Such algorithms which compute for every pixel in one image the corresponding pixels in the other image, were already employed to compute the correspondence between images of faces [5, 7]. For our purposes we will use a gradient based optical flow algorithm which has been adapted from [4].

After solving the correspondence problem, we can represent an image as follows: We code its 2D-shape as the deformation field from a reference image which serves as origin of our space. The texture is coded as the intensity map of the image which results from mapping the face onto the reference face. Now 2D-shape and texture can be treated separately. Both spaces can be expected to be continuous: Intermediate stages between two faces are always naturally looking faces again, and a linear approach seems to be justified.

2 An Implementation

Images: We used the two-dimensional images of human faces that were generated as projections from a database of three-dimensional head models. The head models had originally been collected for psychophysical experiments from male and female volunteers between twenty and forty years old. The volunteers were asked to remove glasses

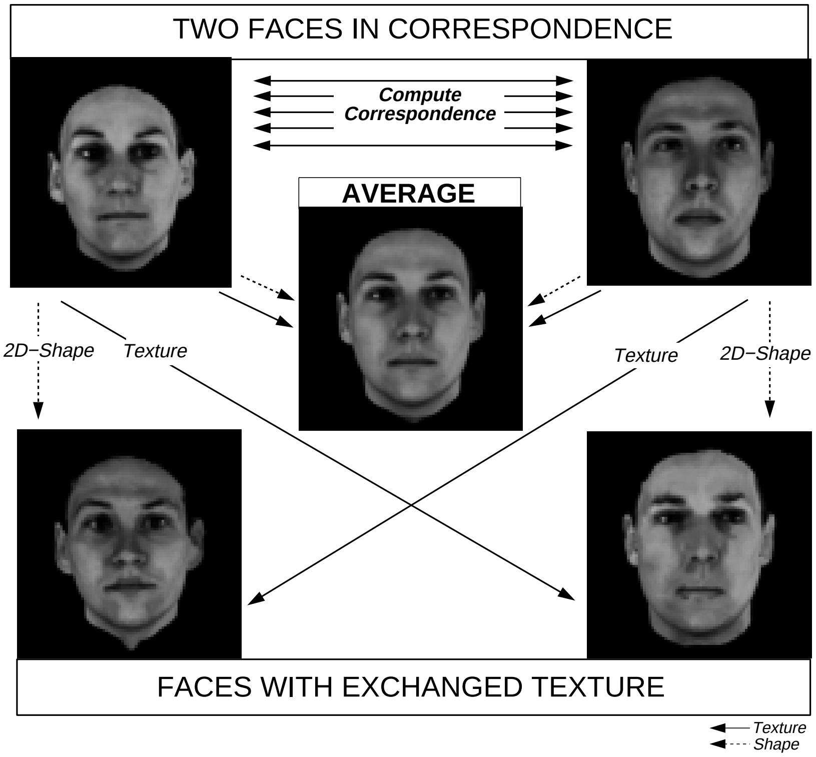

Figure 1: *As in standard image morphing, the automatically computed correspondence of two images of faces (upper row) can be used to generate the average (or any other linear combination) of the two faces (center), by taking the average of the corresponding pixel intensities and the average of the deformation between the two images. It is also possible to exchange the textures of two faces (lower row).*

and earrings. Males were only scanned if they had no beard. The images we used were the same as used in psychophysical experiments but the resolution was reduced to 256-by-256 pixels and the color images were converted to 8-bit grey level images. The images were generated under controlled illumination conditions and the hair of the heads was removed completely from the images.

Preprocessing: In a first step the faces were segmented from the background. This was easy since the background was set to a special intensity value during the image generation. Then the faces were roughly aligned by automatically adjusting them with respect to their two-dimensional center. The center was computed by evaluating separately the average of all x,y coordinates of the pixels related to the face, independent of their intensity values.

Image matching: The essential step in our approach is the computation of the correspondence between two images for every pixel location. That means we have to find for every pixel in the first image, e.g. a pixel located on the nose, the corresponding pixel location on the nose in the second image. This can be a hard problem. However, since we controlled for illumination, and since all faces are compared in the same orientation, a strong similarity of the images can be assumed and problems attributed to occlusions should be negligible. These conditions make an automatic mechanism for comparing the images of the different faces feasible [5]. Such algorithms are known from optical flow computation, in which points have to be tracked from one image to the other. We used a coarse-to-fine gradient-based optical flow algorithm [1] and followed an implementation described in [4]. Begining with the lowest level of a resolution pyramid for every pixel (x, y) in the first image, the error term $E = ((\delta I_1/\delta x)\Delta x + (\delta I_1/\delta y)\Delta y - \Delta I_{1,2})^2$ was minimized for Δx and Δy. I_1 denotes the intensity of the pixel (x, y) in the first image and $\Delta I_{1,2}$ stands for the difference of the intensities in the two images. The resulting vector field $(\Delta x, \Delta y)$ was then smoothed and the procedure was iterated through all levels of the resolution pyramid. The finally resulting vector field was used as the corresponding pattern between the two images.

A representation separating texture and 2D shape: We calculated the correspondence pattern with respect to a reference face for all faces in the database. In theory any face could be used as a reference. However, since peculiarities of a face influence the automated matching process, we used a synthetic face which was generated as an average of a small subset of our data base. The computed correspondance between a face and the reference face enables a representation of the face that separates 2D shape and texture information. The 2D shape is coded as the deformation field from the reference image to the face. This deformation field is identical with the calculated correspondance pattern. The texture information is coded in terms of the texture map that results from mapping the image onto the reference image by means of the deformation field. Thus each texture had now the same dimension, which was equal to the number of pixels of the reference face. Since the correspondence was computed towards the reference face, the deformation field also had a common basis on the pixels of the reference image. However, the dimension is twice as large as in the case of the textures, since each deformation vector on a pixel has an x and a y component.

Synthesis of a new image: A new image can be generated combining any texture with any correspondence field. This is possible because both are given in the coordinates of the reference image. That means for every pixel in the reference image the pixel value and the vector, pointing to the new location are given. The new location generally does not coincide with the equally spaced grid of pixels of the destination image. A commonly used solution of this problem is known as forward warping [14]. For every new pixel, we used the nearest three points to linearly approximate the pixel intensity. As in standard image morphing [3] we are able to compute an average of two faces (fig.1) by averaging the corresponding pixel intensities and drawing them at the average location, which is at half the distance from the reference image to the other image. Figure 1 also shows the mapping of the texture of one face onto the 2D shape of the other face by simply exchanging the corresponding pixel intensities of the faces.

Linear Analysis of Texture and Deformation: We fitted a multivariate normal distribution by performing a principal component analysis separately on the texture and the 2D shape data. The principal components can be calculated as the eigenvectors of the covariance matrix of the data. Figure 2 and 3 show variations of the deformations and the textures along the first six principal components. To the average face we added the respective normalized principal component with weights corresponding to two, four and six standard deviations in both directions. Although this widely exceeds the range of naturally occuring faces, it gives a better understanding of the information contained in the different principal components. All textures were presented on the average shape and the deformations along the deformation principal component were shown with the average texture. Thus the center row (dashed box) shows always the same image, the average face consisting of the average texture and the average 2D-shape.

Reconstruction: Assuming the separated linear spaces of texture and deformation as complete, a basis set of faces, split up into the textures and the deformation fields, will be sufficient to reconstruct any other face. We tested this possibility of reconstructing faces on our small data set of 50 faces. In each test we selected one face and used the remaining 49 faces as basis set. Reconstruction of a face in our terms is equivalent with computing the linear projection of the new face onto the texture and deformation space spanned by the basis faces. This can be done directly on textures and deformations given by the basis faces using singular value decomposition [11]. Figure 4 shows reconstruction experiments for four different faces. In each case we computed separately the projection onto the texture space and deformation space. Then the projections were combined as described earlier to generate the reconstructed image.

3 Results

The quality of the synthesized faces, the faces along principal components and the face reconstructions, demonstrates the quality of the chosen representation. Starting with a total of 83 faces we could easily select fifty faces without visible matching problems by applying the plain optical flow algorithm without any special adaptation to faces. In approximately fifteen cases the matching was obviously wrong. The remaining cases showed fairly good correspondence, but also contained small errors.

The principal component analysis demonstrates that our “face-space” is continuous over a connected parameter set. Within this set we are able to generate images which all look like a human face, only with increasing distance from the average they change smoothly to something different.

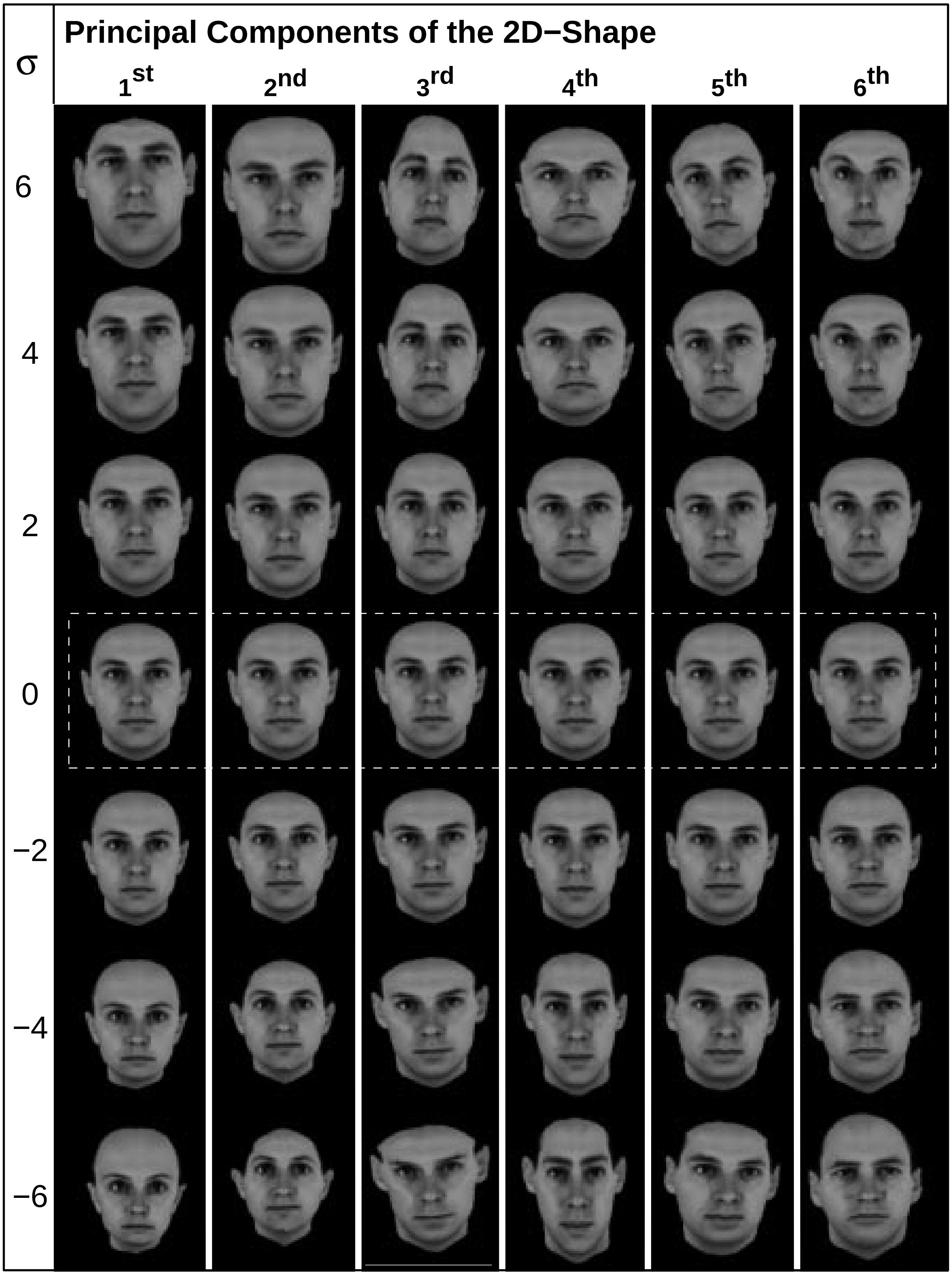

Figure 2: *Examples of a sequence of images along the first six principal components of the shape are shown. The components are computed on a set of 50 faces. The shape variances are visualized by using the average texture. The average face shape is shown in the center row (dashed box). The images in each column show the normalized principal component scaled by σ and added to the average, with σ being the standard deviation of the data set along the selected principal component.*

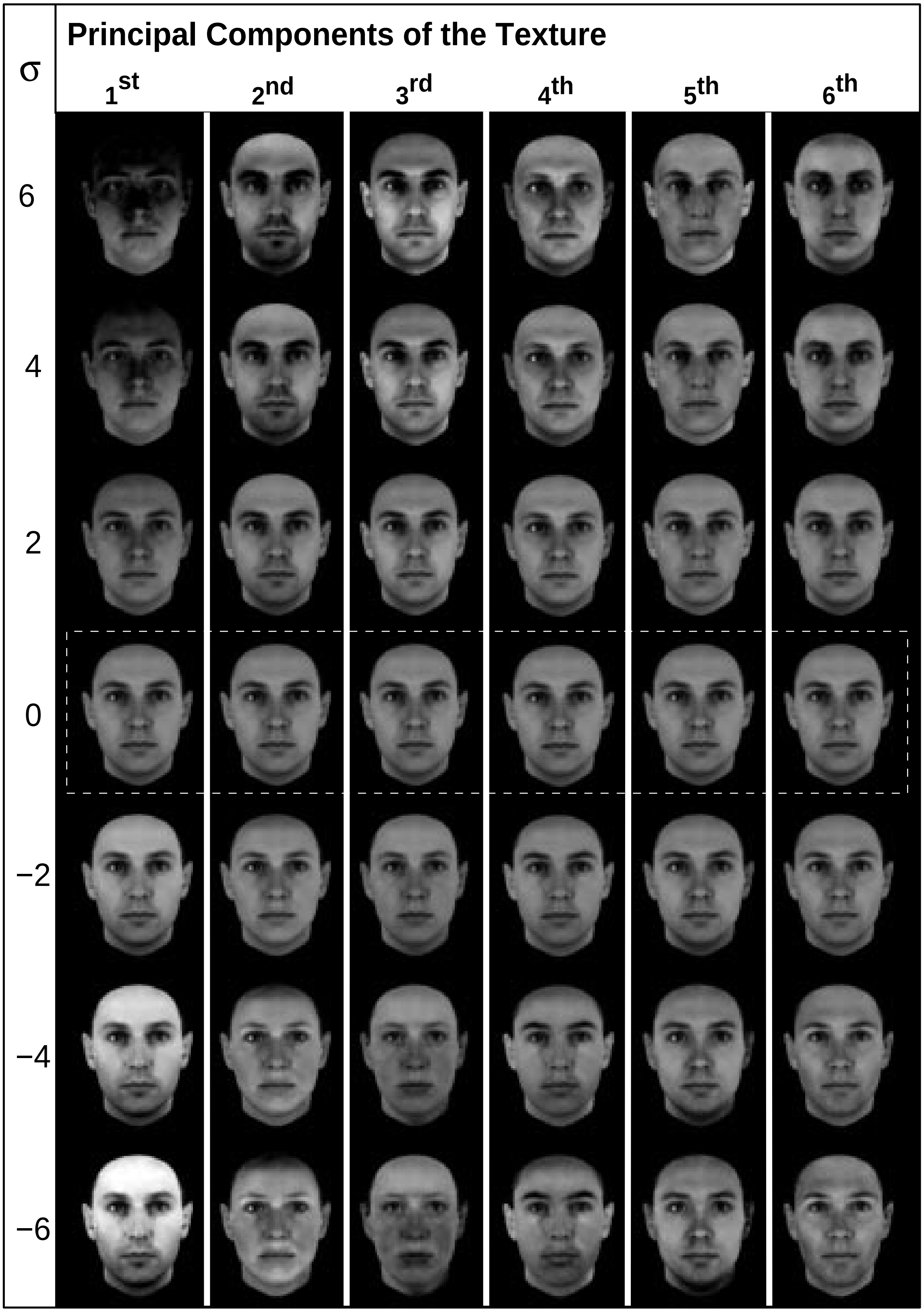

Figure 3: *Examples of a sequence of images along the first six principal components of texture are shown. The components are computed on a set of 50 faces. The texture variances are visualized by mapping them onto the average shape. The average face texture is shown in the center row (dashed box). The images in each column show the normalized principal component scaled by σ and added to the average, with σ being the standard deviation of the data set along the selected principal component.*

The first few principal components clearly mark some characteristics. The first two components of the shape differences show a size effect that goes along with a change of perceived gender, which is also visible in the sixth 'shape-component'. Starting with component three any general size effect disappears and the components characterize only the internal proportions of the faces. Component three and five characterize different head orientations of the faces in our data set. The fourth component shows the transition from a narrow head to a wide head. The first principal component of the textures clearly indicates an illumination change. The light source is moving from above to below. The second texture component shows a change from bright eyebrows to dark ones and also a change in the darkness around the chin, which might be due to the beard of males. The higher components are not so clearly to characterize. They all show various changes in the brightness pattern over the face.

The reconstruction experiments demonstrate the possibility of reconstructing images of human faces using images of other faces as examples. The reconstruction quality, however, differs. From the 50 face which we reconstructed (using in each case the remaining 49 faces as a basis), in eleven cases the reconstruction looked very similar to the original. Approximately the same amount looked very different. The majority was similar, however, human observer judged them as different persons. This was mostly due to very small differences in the region of the eyes. Humans are very sensitive to changes in that region and often two faces are judged to belong to different persons, although the concrete nature of the differences is hardly detectable. In general the reconstruction of female faces was better than that of male faces.

4 Discussion

The presented automated method offers a tool to provide a natural representation which describes the relation of objects belonging to the same object class. On the image level these relations can be interpreted as 2D-deformations according to a reference face shape and to intensity changes in the matched texture map. The demonstrated separation of shape and texture leads to a natural representations of human faces and objects in general. The resulting "face-space" is continuous over a connected parameter range and allows a smooth modeling of images of faces.

A first step towards a linear description of that space is a principal component analysis, which can be done separately for the deformation and for the texture. Our data set was not very large and conclusions are thus preliminary. However, the fact, that projections of a new face into the space spanned by the 49 remaining faces yields a fairly good reconstruction, shows that the dimensionality of the space might be not much larger. Before being able to fully interpret our analysis, it is necessary to calculate the projections of the data points on the principal components and to investigate the underlying distributions.

A critical point for improving the quality of face reconstructions is an appropriate measure for the similarity of two faces. Such a measure should be based on psychophysical experiments. The sensitivity of human observers to differences in texture or shape depend strongly on the location of these differences in the face. For instance, we are much more sensitive to changes in the eye region than to changes in the region of the ears or the nose. Euklidian distance in the face space does certainly not reflect the perceived difference between to faces. On the other hand, it is likely that the just noticible distance along single principal components stays constant.

A good measure, which evaluates the perceptual difference between two faces could also help, to adapt the optical flow algorithm and to make it more appropriate for finding correspondence between faces. The error function, which the algorithm tries to minimize, could be formulated in a more subtle way. To come back to the above example, changes in the eye region could be made more expensive than changes in other regions of the face. There are several other improvements which could be done to optimize the correspondence within a data set. One possibility is a multi-step procedure. If two faces cannot be set into correspondence correctly, but if there is a third face which is already in good correspondence to both of them, the two deformation fields can be added to get the correct one.

Our goal for the moment is to represent a large data base in the described manner and to provide reliable statements about the statistics of the "face space". This requires also improvements on the data base. The first principal component for the texture seems mainly to account for differences in illumination. The second accounts for the amount

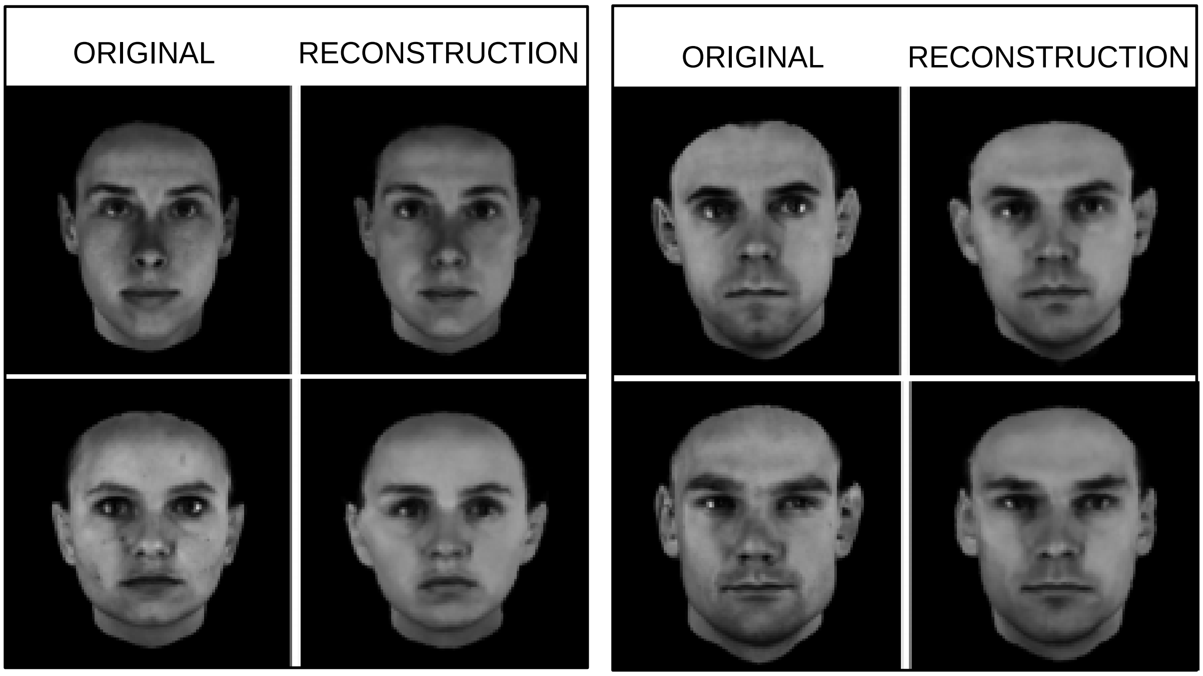

Figure 4: *The reconstruction of two female faces (left block) and of two male faces (right block) is shown. The faces were reconstructed through projecting them on the texture and shape space of the* 49 *images of other faces in our data base.*

of hairs (eyebrows, beard) in the face. Both parameters have not been varied systematically but are rather accidental. Since the data base was originally established for psychophysical purposes, we actually tried to provide constant illumination conditions. The remaining differences, however, were extracted very precisely by the first principal component. Also, male persons were taken up into the data base only if they did not wear a beard. However some of them were obviously not shaved good enough.

Changes in the texture according to changes of illumination form a low-dimensional linear space [8]. So in long terms, changes of illumination and also of other relevant parameters like expression, age and orientation should be included systematically into an enlarged data base. We expect that a principal component analysis will lead to low dimensional subspaces, accounting for these parameters in a way like the first principal component in Figure 3 accounts for the illumination.

A recognition system based on a space spanned by the principal components could solve the problem of being invariant to parameters like illumination, orientation or expression by ignoring the corresponding subspaces. The biological relevance of our model can easily be verified by means of psychophysical experiments. Faces can systematically be varied along the defined directions in the "face space" and the robustness of subjects recognition performance to that component can be measured. These experiments are planed for the future and we hope that they will improve our understanding of human face recognition.

References

[1] E.H. Adelson and J.R. Bergen. The extraction of spatiotemporal energy in human and machine vision. *Proc. IEEE Workshop on Visual Motion, Carlston*, pages 151–156, 1986.

[2] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Int. Journal of Computer Vision*, pages 43–77, 1994.

[3] T. Beier and S. Neely. Feature-based image metamorphosis. In *SIGGRAPH '92 proceedings*, pages 35–42, Chicago, IL, 1992.

[4] J.R. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center Princeton NJ 08540, 1990.

[5] D. Beymer, A. Shashua, and T. Poggio. Example-based image anaysis and synthesis. A.I. Memo No. 1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1993.

[6] Ian Craw and Peter Cameron. Parameterizing images for recognition and reconstruction. In *Proc. British Machine Vision Conference*, 1991.

[7] I.A. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. Technical report 1301, MIT Media Laboratory Perceptual Computing Section, 1991.

[8] P.W. Hallinan. A low-dimensional representation of human faces for arbitrary lightning conditions. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle WA, 1994.

[9] A.J. O'Toole, H.Abdi, K.A. Deffenbacher, and D. Valentine. Low-dimensional representation of faces in in higher dimensions of the face space. *J.Opt. Soc.Am. A*, 10(3):405–411, 1993.

[10] T. Poggio and T. Vetter. Recognition and structure from one 2D model view: observations on prototypes, object classes, and symmetries. A.I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1992.

[11] William H. Press. *Numerical recipes in C : the art of scientific computing.* Cambridge University Press, Cambridge, 1992.

[12] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4:519–554, 1987.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[14] Georg Wolberg. *Image Warping.* IEEE Computer Society Press, Los Alamitos CA, 1990.