

Analysis and Synthesis of Pose Variations of Human Faces by a Linear PCMAP Model and its Application for Pose-Invariant Face Recognition System

Kazunori Okada^{1,3}

¹ Laboratory of Computational and Biological Vision
University of Southern California
HNB228 Los Angeles, CA 90089-2520 U.S.A.
kazunori@selforg.usc.edu

Shigeru Akamatsu³

³ Human Information Processing Research Laboratories, ATR
2-2 Hikaridai Seika-cho Soraku-gun, Kyoto 619-0288 Japan
akamatsu@hip.atr.co.jp

Christoph von der Malsburg^{1,2}

² Institut für Neuroinformatik
Ruhr-Universität Bochum
D-44870 Bochum Germany
malsburg@neuroinformatik.ruhr-uni-bochum.de

Abstract

A method of manifold representation for human faces with pose variations is proposed. Our model consists of mappings between 3D head angles and facial images separately represented in shape and texture, via sub-space models spanned by principal components (PCs). Explicit mappings to and from 3D head angles are used as processes of pose estimation and transformation, respectively. Generalization capability to unknown head poses enables our model to continuously cover pose parameter space, providing high approximation accuracy. The feasibility of this model is evaluated in a number of experiments. We also propose a novel pose-invariant face recognition system using our model as the entry format for a gallery of known persons. Experimental results with 3D facial models recorded by a Cyberware scanner show that our model provides a superior recognition performance against pose variations, and that texture synthesis process is carried out correctly.

1 Introduction

Images of objects vary in appearance due to changes in image projection settings, background, object properties, and other sources of variation. The fact that these varia-

tions are entangled with each other and encoded implicitly in the image makes the task of object recognition difficult. With faces, this problem becomes more complex because the innate characteristics of faces, which distinguish one face from another, do not vary greatly across different individuals: the magnitudes of the variations of the innate characteristics in images are often much smaller than the magnitudes of the common variations. For the task of facial identification, analyses of innate facial characteristics are possible only after making these implicitly encoded variations *explicit* in order to process *only* the innate characteristics. This general problem still remains unresolved [15, 13].

Another important question in this realm is the nature of facial representation. Even for a single face, the number of all possible views becomes prohibitively large because of a combinatory explosion due to the number of dimensions of variation. Since it is simply impossible to store all the possible views, the representation needs to possess a generalization capability to continuously cover the face space from a limited number of observable samples. As a solution to this problem, subspace methods based on Principal Component Analysis (PCA), eigenface systems [18, 19], have been widely used as a compact and continuous parameterized model of faces. However, their model parameters are often hard to interpret in relation to the observable physical parameters. Another approach for solving this prob-

lem is to treat it as a problem of learning nonlinear mappings between input representation and physical parameter spaces. A number of systems based on a Radial Basis Function (RBF) network have been proposed for this approach [5, 1, 8]. They showed successful learning capability, and made these variations explicit. However, the generalization capabilities of such systems have not been fully investigated yet.

In this study, we present a representation and processing model of human faces with head pose variations. This model attempts to find *mappings* between facial images and physical parameters, in our case 3D head angles, via *parameterized manifold representations* of faces using the PC-subspace method. For a better generalization capability, we approximated these mappings by using a combination of linear systems: 1) subspaces of input representation spaces spanned by principal components (PC-subspace), and 2) linear transfer matrices between these subspaces and a pose parameter space. We call this model the *linear PCMAP model* [14]. When learned for an individual, the mappings account for various poses of the individual's face (manifold representation) and provide an explicit interface of the model with physical pose parameters, enabling processes of pose estimation and transformation. Since the mappings are bidirectional, pose estimation and transformation can be realized by face-to-pose and pose-to-face mappings, respectively. We call these respective face-to-pose and pose-to-face mappings: the *analysis* and *synthesis* processes.

It was Ullman and Basri [20] who first showed that arbitrary objects in line drawings can be expressed with linear combinations of a small number of 2D templates of the same objects, forming a foundation to a study of manifold representation of objects. *Linear class theory* proposed by Poggio [17] extended this idea by connecting multiple classes of objects. This framework was successfully applied by Beymer [2] and Vetter [21] to facial images. A limitation in these studies is the use of discrete samplings of parameter spaces of physical variations. This not only prohibits smooth coverage of the (pose) parameter spaces with a limited number of templates but also demands a collection of many templates with pre-determined (pose) variations, which is often difficult in practice. Our model tries to solve this problem by providing continuous coverage of the pose parameter space from a limited number of training samples.

Our work is related to a number of previous studies. Maurer and von der Malsburg proposed an algorithm for pose transformation that maps two jets sampled at two different head poses [10]. This algorithm, however, requires a priori knowledge of 3D facial structure and its application has been limited to a small number of discrete poses. Murase and Nayar presented an object representation model using parametric eigenspace [12]. They utilized cu-

bic spline interpolation for computing continuous manifolds in compact PC-based subspaces. Their study addressed the same problem of parameterizing object views by pose parameters, however they applied it to only a few number of generic objects and considered only one degree of freedom from the 3D rotations. Beymer et al. proposed analysis and synthesis systems of pose and expression variations based on RBF networks [1]. Although their framework is similar to ours, they only exploited pixel-value based single view representations and analyzed only one degree of freedom from the 3D rotations of heads. Recently, Lanitis et al. [9] presented a facial processing system using PCA based manifold representations. They also used separate shape and texture representations and proposed a pose estimation system similar to our model. Their texture representation, however, was based on pixel values instead of our Gabor jet based texture representation. Moreover, their pose estimation did not include planar rotations and they did not discuss pose transformation and generalization capability to unknown head poses.

In this paper, the linear PCMAP model is first described. We next analyze the model's performance when used as a representation of single individuals. We also present a novel pose-invariant face recognition system using the linear PCMAP model as an entry to a known person's database. Finally, we conclude this paper by discussing the results of the analyses and our future work.

2 Linear PCMAP Model for Representing Faces with Pose Variations

2.1 Model Description

The learning and matching stages of the linear PCMAP model are described in this section. In the learning stage of this model, pairs of 2D facial images and their corresponding 3D head angles are used as a training data set. We employed separate representations for the shape and texture of human faces. The benefits of separately representing shape and texture information for 2D facial images have been shown in a number of studies [4, 22, 9]. Since the head pose variation can be considered as a geometrical problem, we designed our model such that pose variations are only related to the shape information of faces. Texture information is then related to shape information, exploiting the correlation between the shapes and textures of faces. The effective range of a model depends on the training data set used for the learning stage of the model. In this section, we treat these mappings as manifold representations of *individual* faces with pose variations by using training data sets consisting of facial images of *single* individuals.¹

¹however the model is not innately restricted to single individuals.

We denote a training data set by $(\vec{v}^m, \vec{\theta}^m)_{1,\dots,M}$, where \vec{v}^m and $\vec{\theta}^m$ express the m -th training facial image and its 3D head angles, respectively. In the first step, \vec{v}^m is decomposed to a pair of shape and texture representations, $(\vec{x}^m, \vec{j}^{m,n})$. Shape information is represented by a $2N$ -component vector \vec{x}^m of object-centered image coordinates of N facial landmarks. For each landmark x_n^m , an L -dimensional Gabor jet $\vec{j}^{m,n}$ is recorded from \vec{v}^m as the localized texture representation of the landmark n in the frame m , where $j_l^{m,n}$ is the jet coefficient derived from the l -th Gabor filter.

Next $(\vec{x}^m)_{1,\dots,M}$ and $(\vec{j}^{m,n})_{1,\dots,M;1,\dots,N}$ are independently subjected to PCA resulting in a set of PCs as orthonormal bases of shape and texture representation spaces, $(\vec{y}^p)_{1,\dots,P}$ and $(\vec{b}^{s,n})_{1,\dots,S;1,\dots,N}$, where s and p are the indices of PCs in decreasing order of their corresponding variances. Shape and texture subspaces are defined by selecting P_o and S_o^n as small as possible but still large enough to have the subspaces $(\vec{y}^p)_{1,\dots,P_o}$ and $(\vec{b}^{s,n})_{1,\dots,S_o^n}$ cover a large share of the data variance. We call the shape and texture subspaces *shape and texture models*, respectively. In this study, for simplicity, we used the same S_o for all N landmarks. The shape and texture models have an optimal reconstruction property by a linear combination in the least square sense,

$$\vec{x} \approx \vec{x}^0 + \sum_{p=1}^{P_o} q_p \vec{y}^p, \quad (1)$$

where $\vec{x}^0 = 1/M \sum_{m=1}^M \vec{x}^m$, and P_o -component *shape parameters* \vec{q} is defined as $\vec{q} = \langle \vec{x} - \vec{x}^0 | \vec{y}^p \rangle_{1 \leq p \leq P_o}$;

$$\vec{j}^n \approx \vec{j}^{0,n} + \sum_{s=1}^{S_o} r_s^n \vec{b}^{s,n}, \quad (2)$$

where $\vec{j}^{0,n} = 1/M \sum_{m=1}^M \vec{j}^{m,n}$, and S_o -component *texture parameters* at n -th landmark \vec{r}^n is defined as $\vec{r}^n = \langle \vec{j}^n - \vec{j}^{0,n} | \vec{b}^{s,n} \rangle_{1 \leq s \leq S_o}$. Note that (1) and (2) become equations when $P_o = P = 2N$ and $S_o = S = L$.

Next, we linearly relate model parameters \vec{q}^m and $\vec{r}^{m,n}$ and 3D head angles $\vec{\theta}^m$. For face-to-pose mapping (*analysis*), we relate only shape model parameters to 3D head angles because shape parameters showed a higher correlation to head angles than texture parameters in our pilot experiments. For pose-to-face mapping (*synthesis*), we first relate 3D head angles to shape parameters. Texture parameters are then related to shape parameters. In order to compensate for obvious nonlinearity in mappings between shape parameters and 3D head angles, we nonlinearly expand 3-component head angle vectors $\vec{\theta}^m$ to 6-component *pose parameters* $\vec{\varphi}^m$ by using a trigonometric functional transformation K ,

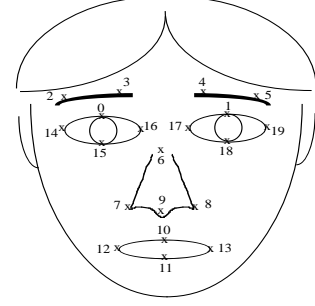


Figure 1. Definition of Facial Landmarks.

$$K : (\alpha, \beta, \gamma) \mapsto (\cos(\alpha), \sin(\alpha), \cos(\beta), \sin(\beta), \cos(\gamma), \sin(\gamma)). \quad (3)$$

Thus, the shape model parameters are related to these pose parameters instead of being directly related to 3D head angles. Now we formulate these relations in matrix notations,

$$\Phi = Q \cdot H, \quad (4)$$

$$Q = \Phi \cdot G, \quad (5)$$

$$R^n = Q \cdot F^n, \quad (6)$$

where $R^n = (\vec{r}^{1,n}, \dots, \vec{r}^{M,n})^t$, $Q = (\vec{q}^1, \dots, \vec{q}^M)^t$, $\Phi = (\vec{\varphi}^1, \dots, \vec{\varphi}^M)^t = (K(\vec{\theta}^1), \dots, K(\vec{\theta}^M))^t$. The transfer matrices H , G , and F^n are computed by solving these equations with the SVD algorithm.

After finding these mappings, we can estimate 3D head angles from a given facial representation with arbitrary pose (analysis) and can synthesize a facial image from given 3D head angles (synthesis) using the learned model. These processes are called the matching stage.

The face-to-pose mapping of the analysis process is written as

$$\vec{v}^a \xrightarrow{L} \vec{x}^a \xrightarrow{Eq.(1)} \vec{q}^a \xrightarrow{Eq.(4)} \vec{\varphi}^a \xrightarrow{arctan} \vec{\theta}^a, \quad (7)$$

and pose-to-face mapping of synthesis process is

$$\begin{array}{ccccccc} \vec{\theta}^a & \xrightarrow{K} & \vec{\varphi}^a & \xrightarrow{Eq.(5)} & \vec{q}^a & \xrightarrow{Eq.(6)} & \vec{r}^{a,1}, \dots, \vec{r}^{a,N} \\ & & & & \downarrow Eq.(1) & & \downarrow Eq.(2) \\ & & & & \vec{x}^a & & \vec{j}^{a,1}, \dots, \vec{j}^{a,N} \\ & & & & \searrow R & & \swarrow R \\ & & & & & & \vec{v}^a \end{array} \quad (8)$$

To separate shape and texture information, we must find facial landmarks in every sample. We used a facial landmark tracking system developed by Maurer et al. [11],



Figure 2. Examples of Training Samples.

which assumes that training and test samples are given by video sequences starting from a frontal view of faces. This decomposition of shape and texture information is denoted by an operator L in formula 7. Figure 1 shows a definition of the 20 facial landmarks used throughout this study.

An algorithm for a grey-level image reconstruction of a Gabor jet based graph representation of faces [7, 23], which was developed by Poetzsch et al. [16] performs a reverse operation that reconstructs a facial image from synthesized shape and texture representations. This operation is symbolized by an operator R in formula 8.

Connecting analysis and synthesis stages realizes a process of model matching that allows us to synthesize, from an arbitrary input face, a facial image whose pose is aligned to the input and whose appearance is from one learned in the matched model. We call this combination of processes *analysis-synthesis-chain*. This process will be used in the face recognition system described in the next section.

2.2 Data Set

We first evaluate our model by training it with samples from single individuals. In this case, the model can be considered as a manifold representation of individual faces with an explicit interface of pose variations. Grey-level image sequences of various head poses are recorded for three individuals in this analysis. Each sequence consists of 1200 frames and four different types of pose variations. For first three types shown in figure 2, subjects are asked to rotate their heads along only one axis at a time. 300 frames are captured for each rotation: horizontal, vertical, and planar rotation (denoted in the figure 2 and 3 as **1,2,3**, respectively). The total of 900 frames of these three types are used as training samples for our model. For the rest of the 300 frames, subjects are asked to move their heads freely. These frames are used for test samples. During the acquisition of these images, lighting conditions and the background are unchanged and subjects are asked to not change their facial

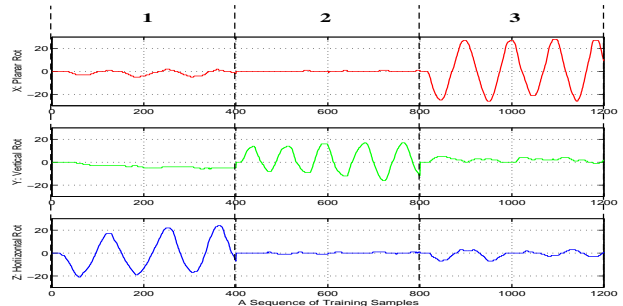


Figure 3. Variation of Head Pose in The Training Samples.

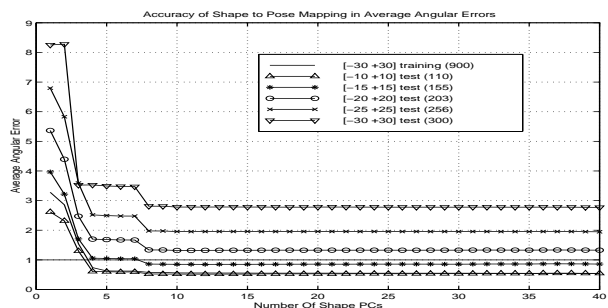


Figure 4. Accuracy of Shape to Pose Mapping.

expressions.

We also measure the physical head angles for each frame by a magnetic sensor synchronized to a frame grabber for the image acquisition. Horizontal, vertical, and planar rotations of head are independently measured as a continuous angular deviation from the frontal pose of the head, as shown in figure 3. The range of rotation along each axis is between -30 and $+30$ degrees.

2.3 Experiments

In order to evaluate the accuracy of the linear PCMAP model, we analyzed errors between test and synthesized samples. For each individual, a linear PCMAP model is trained with 900 training samples, as described in the previous section. Each learned linear PCMAP model is tested by using a number of test sample sets: 1) 900 training samples, 2) test samples whose pose range is between ± 10 , 3) ± 15 , 4) ± 20 , 5) ± 25 , and 6) ± 30 .

Figure 4 shows the average accuracy of a shape analysis with shape-to-pose mapping. We compared the 3D head angles of each test sample to estimated angles by using the

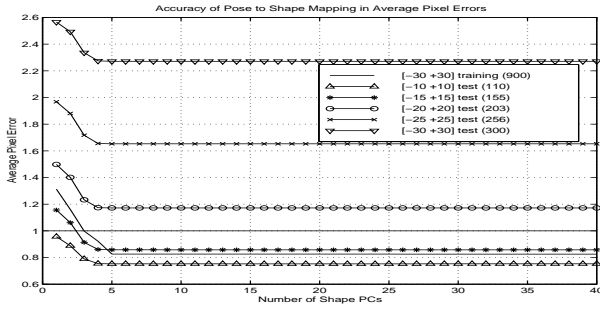


Figure 5. Accuracy of Pose to Shape Mapping.

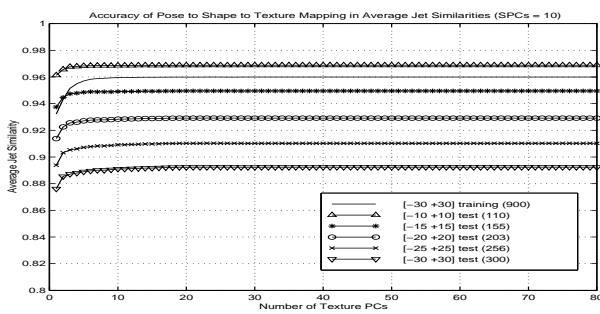


Figure 6. Accuracy of Pose to Shape to Texture Mapping.

analysis process of a learned linear PCMAP model. Angular deviations in degrees between the test and estimated 3D head angles are averaged over three rotational dimensions, a number of test samples, and three individuals. The average angular deviations are plotted against the number of shape PCs used in the linear PCMAP model. The accuracy is less than 1 degree when the pose range of test samples is within ± 15 and more than 10 shape PCs are used.

Figure 5 shows the average accuracy of the shape synthesis with pose-to-shape mapping. We compared the facial landmark locations of each test sample to landmark locations of a synthesized shape representation by using the synthesis process of the learned linear PCMAP model. Pixel deviations of each facial landmark between the test and synthesized shape are averaged over 20 facial landmarks, test samples, and three individuals and plotted against the number of shape PCs used in the linear PCMAP model. The accuracy is less than 1 pixel when the pose range of test samples is within ± 15 and more than 10 shape PCs are used. This condition is the same for the pose-to-shape mapping.

The results of the two above analyses revealed that the

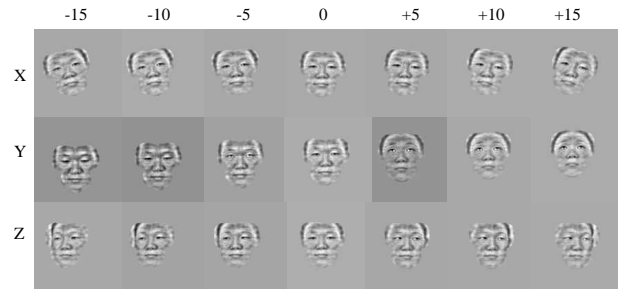


Figure 7. Reconstructed Training Samples.

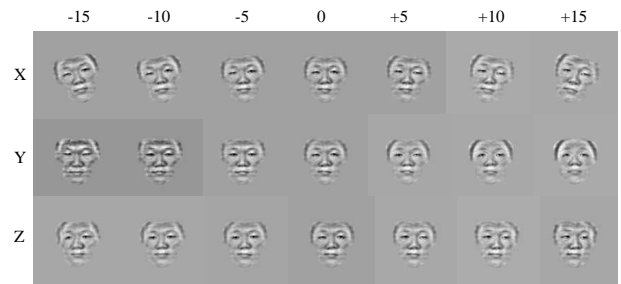


Figure 8. Synthesized Images with 10 Shape PCs and 20 Texture PCs. The Results for Known Poses

top 10 shape PCs in decreasing order of variances gave a satisfactory accuracy for both the shape analysis and the synthesis processes. These 10 PCs also accounted for 98% of the total variances in the training samples.

Figure 6 shows the average accuracy of the texture synthesis with a combination of pose-to-shape and shape-to-texture mappings. A set of 20 jets of each test sample and each synthesized sample are compared using normalized dot-product of a pair of Gabor jet magnitudes averaged over 20 landmarks. These jet-based sample similarities, ranging between 0 and 1, are averaged over the test samples and three individuals and plotted against the number of texture PCs in the linear PCMAP models. The number of shape PCs is fixed to 10 in these analyses. The figure shows that the accuracy reaches a maximum for each testing case when more than 20 texture PCs are included. These analyses also indicate the generalization capability of our model when the pose range of test samples is limited to ± 15 . In these conditions, there are no significant differences between the test results of the test and training samples.

Figure 7 displays examples of reconstructed images that are directly reconstructed from the training samples of various poses. The quality of this reconstruction cannot be per-



Figure 9. Synthesized Images with 10 Shape PCs and 20 Texture PCs. The Results for Unknown Poses

fect since the information of the gray level distribution is only available at coarse sampling points (20 landmarks in this case). These images, however, serve as references for the images reconstructed from synthesized representations.

Figure 8 displays examples of reconstructed images from synthesized shapes and textures using a shape model with 10 PCs and a texture model with 20 PCs. Each image in this figure corresponds to one in the figure 7. In this condition, the rotation variations along three axes, as well as the facial appearance presented in the training samples, are accurately captured.

Next, in figure 9, we show examples of reconstructed images from synthesized shapes and textures with *unknown poses*, in order to evaluate the model's generalization capability. The face is rotated by using a synthesis process of our model along all three rotation axes simultaneously between -15 degrees to $+15$ degrees in two different ways (A and B). The results show that the given rotation variations, which are not present in training samples, are also captured correctly. These results support the potential of our model for generalizing unknown poses. Both the shape and texture of largely rotated faces in this figure, however, seem to be more distorted. This suggests that the generalization capability of our model might be restricted to a range of head poses; the model does not seem to be able to extrapolate the pose variations.

3 Pose-Invariant Face Recognition System using Linear PCMAP Model

3.1 System Description

In this section, we present a novel face recognition system using the linear PCMAP model as an entry to a known person's gallery.

Figure 10 shows an overview of this recognition system. In this system, an arbitrary input is subjected to the analysis-synthesis-chain process, described in section 2, with each linear PCMAP model stored in the gallery. This results in

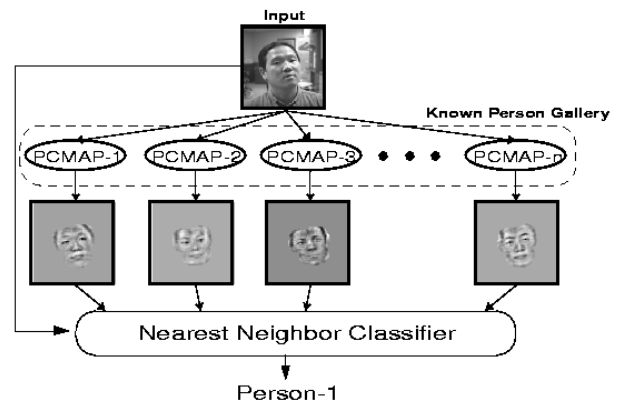


Figure 10. Pose-Invariant Recognition System with Linear PCMAP Models

model views of each known person whose pose is aligned to the input. After this pose alignment, we perform a nearest neighbor classification of the input with these model views. Because of the pose alignment, the recognition performance should improve against the pose variations. Furthermore, there is no systematic limitation to particular discrete head poses because of the continuous coverage of pose parameter space by using the linear PCMAP model. As long as the learned linear PCMAPs cover a sufficient range of head poses, an input with arbitrary poses can be processed without any pose restrictions.

3.2 Data Set

In this experiment, we use samples generated from 3D facial models recorded by a Cyberware 3030 scanner. Twenty models (10:female,10:male) are randomly picked from a 3D facial model database of Japanese faces developed at ATR. For each model, test and training samples in the same format as the previous experiment (test samples: 186, training samples: 248×3) are generated by rendering 2D view snapshots while explicitly rotating the 3D face model [6]. Locations of facial landmarks in various poses are determined by explicitly rotating 3D reference coordinates that are found manually for a frontal view of each model. These test and training samples are appropriate for our system's evaluations since there are no measurement errors of head pose angles and landmark locations.

3.3 Experiments

Table 1 shows the result of the performance analysis of

Table 1. Percentages of Correct Identifications with 1) Linear PCMAP and 2) Single-View Models as the Entry Format of a Database of Known Persons

Model	10 degs	15 degs	20 degs	25 degs	30 degs
L-PCMAP	100.0	99.8	94.6	83.7	74.4
S-View	99.2	89.6	75.8	65.0	56.7

our system of recognizing faces with pose variations. The proposed system is compared with a simplified system, in which each entry of a known person's database is represented by a single frontal view of the person (single-view model). Each column of the table shows percentages of correct identifications by the two systems when the range of head angles in test samples is limited to 10, 15, 20, and 30 degrees, respectively. Similar to the analyses for single persons presented in the previous section, head poses of the test samples are not present in training samples. The recognition rates of our system when head angles of test samples are limited within 20, 25, and 30 degrees are approximately 20 percent higher than those of a system with the single-view model. These results show that our system achieves high recognition rates with a wider range of head angles in test samples than a simple single-view model.

4 Discussions

In this paper, we proposed a linear PCMAP model that is a manifold representation of 2D facial images with an explicit interface of pose variations. This model was evaluated by a number of error analyses. The experimental results indicated high accuracy in approximation of the mapping between shape and pose. In the literature, only a few studies have reported quantitative analyses of pose estimation accuracy. The stereo based system by Xu and Akatsuka [24] resulted in an average angular error of 4.4 degrees when the range of pose variation was within ± 20 degrees. Choi et al [3] reported approximately 3 degrees average angular error within a range of ± 40 degrees in their 3D shape model fitting system using an EM algorithm. Our pose estimation results (approx. 0.9 degrees in ± 15 and 1.3 degrees in ± 20) outperformed these previous reports in a limited range of pose variations. Our shape synthesis also achieved sub-pixel accuracy on average when the pose range in the test samples was within ± 15 degrees. These results indicate that our model correctly approximates the mappings between shape and pose.

An advantage of our model is that both the analysis and synthesis processes continuously and smoothly cover the space of pose parameters by utilizing interpolation. The

experimental results showed that our model is capable of generalizing unknown poses from training samples with a limited range of poses. The model is also compact: the data compression ratio from a set of training samples to a learned model is approximately 60. Computational costs of the model is fair. The learning process of our model includes computationally intensive procedures such as PCA and SVD. The time taken to learn a single face was 4 minutes on a Sun SPARC20 workstation. In contrast to learning, the matching process of our model is computationally efficient. Only a fraction of a second is needed for pose estimation or transformation of a single face. Compared to linear class systems, our compact and continuous model not only provides a better overall fit for continuous pose variations in samples but also eliminates the requirement for operator's assistance and subject's collaboration, which is required when collecting samples with specific head poses. The generalization capability of our model is possibly due to our choice of simplified linear systems. However, there is a trade-off in that the effective range of pose variations becomes limited. One idea for solving this problem is to patch the whole parameter space with a set of local linear models. Therefore, a point in the parameter space can be interpolated with a number of neighboring local models. This is one of our future topics.

We also proposed a novel pose-invariant face recognition system using the linear PCMAP model as an entry format of a known person's gallery. Our recognition system postulates that pose-invariance can be achieved by giving a learning capability to the memory/knowledge systems, a known person's gallery in this case, instead of trying to find pose-invariant properties in input representations within a perceptual process.

The experimental results presented in this paper suggest that this system improves the recognition performance against pose variations in comparison to a simplified memory model that represents a known person with a single frontal view of the person. These results also imply that our model provides the correctness (preserves innate facial appearances while accurately analyzing/synthesizing head poses) of the texture synthesis process. A precise facial landmark finding or tracking system is required for the automation of this recognition system. The Maurer's system used in our experiment is one of the candidates for this task, but this front-end system should be improved.

The parameterization of our model with physical head angles provides a compact interface for other perceptual modules that is easy to interpret. This characteristic also provides a number of potential application scenarios for low-bandwidth visual communication systems, in which only the head pose information is sent over a network, or for tele-conferencing systems, in which facial orientations in a virtual space can be corrected to maintain eye contact.

Acknowledgments

The authors wish to thank Jan Wieghardt and Junmei Zhu for helpful discussions and Katsunori Isono for making his 3D facial model rendering system available for this study. This work has been supported by ONR grant N00014-98-1-0242.

References

- [1] D. Beymer, A. Shashua, and T. Poggio. Example based image analysis and synthesis. Technical Report A.I. Memo, No. 1431, Artificial Intelligence Laboratory, M.I.T., 1993.
- [2] D. J. Beymer and T. Poggio. Face recognition from one example view. Technical Report A.I. Memo, No. 1536, Artificial Intelligence Laboratory, M.I.T., 1995.
- [3] K. N. Choi, M. Carcassoni, and E. R. Hancock. Estimating 3d facial pose using the em algorithm. In *Face Recognition: From Theory to Applications*, pages 412–423. Springer-Verlag, Sterling, UK, 1998.
- [4] I. Craw, N. Costen, T. Kato, G. Robertson, and S. Akamatsu. Automatic face recognition: Combining configuration and texture. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 53–58, Zurich, 1995.
- [5] S. Edelman, D. Weinshall, and Y. Yeshurun. Learning to recognize faces from examples. In *Proceedings of the 2nd European Conference on Computer Vision*, volume 588, pages 787–791, 1992.
- [6] K. Isono and S. Akamatsu. A representation for 3d faces with better feature correspondence for image generation using pca. Technical Report HIP96-17, The Institute of Electronics, Information and Communication Engineers, 1996.
- [7] M. Lades, J. C. Vorbrueggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wuerzt, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE transactions on Computers*, 42:300–311, 1993.
- [8] M. Lando and S. Edelman. Generalization from a single view in face recognition. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 80–85, Zurich, 1995.
- [9] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:743–755, 1997.
- [10] T. Maurer and C. von der Malsburg. Single-view based recognition of faces rotated in depth. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 248–253, Zurich, 1995.
- [11] T. Maurer and C. von der Malsburg. Tracking and learning graphs and pose on image sequences. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 176–181, Vermont, 1996.
- [12] H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
- [13] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg. The bochum/usc face recognition system: And how it fared in the feret phase iii test. In *Face Recognition: From Theory to Applications*, pages 186–205. Springer-Verlag, Sterling, UK, 1998.
- [14] K. Okada, C. von der Malsburg, and S. Akamatsu. A pose-invariant face recognition system using linear pemap model. Technical Report HIP99-48, The Institute of Electronics, Information and Communication Engineers, 1999.
- [15] P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation. In *Face Recognition: From Theory to Applications*, pages 244–261. Springer-Verlag, Sterling, UK, 1998.
- [16] M. Poetzsch, T. Maurer, L. Wiskott, and C. von der Malsburg. Reconstruction from graphs labeled with responses of gabor filters. In *Proceedings of the International Conference of Artificial Neural Networks*, pages 845–850, Bochum, 1996.
- [17] T. Poggio and T. Vetter. Recognition and structure from one 2d model view: Observations on prototypes, object classes and symmetries. Technical Report A.I. Memo, No. 1347, Artificial Intelligence Laboratory, M.I.T., 1992.
- [18] L. Sirovich and M. Kirby. Low dimensional procedure for the characterisation of human faces. *Journal of the Optical Society of America*, 4:519–525, 1987.
- [19] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.
- [20] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1006, 1991.
- [21] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:733–742, 1997.
- [22] T. Vetter and N. Troje. A separated linear shape and texture space for modeling two-dimensional images of human faces. Technical Report TR15, Max-Planck-Institut fur biologische Kybernetik, 1995.
- [23] L. Wiskott, J.-M. Fellous, N. Krueger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.
- [24] M. Xu and T. Akatsuka. Detecting head pose from stereo image sequence for active face recognition. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 82–87, Nara, 1998.