A Pose-Invariant Face Recognition System using Linear PCMAP Model

Kazunori Okada[1,3] Christoph von der Malsburg[1,2] Shigeru Akamatsu[3]
[1] Laboratory of Computational and Biological Vision, USC, Los Angeles, U.S.A.
[2] Institut für Neuroinformatik, Ruhr-Universität Bochum, Bochum, Germany
[3] Human Information Processing Research Laboratories, ATR, Kyoto, Japan
HNB228 Los Angeles, CA 90089-2520 U.S.A.
kazunori@selforg.usc.edu

Abstract

We propose a novel pose-invariant face recognition system using a manifold representation for human faces with pose variations (linear PCMAP model) as the entry format for a database of known persons. The model's generalization capability for unknown head poses enables a continuous coverage of the pose parameter space, providing high approximation accuracy for pose estimation (analysis) and transformation (synthesis). With this model as the entry format for the database, the head pose of each known face is aligned to an arbitrary head pose of an input face, resulting in a pose-invariant recognition. Experimental results with 3D facial models recorded by a Cyberware scanner show that the recognition performance of our model against pose variations is superior to that of a single-view model and is equivalent to that of a multi-view model within a limited pose range in test samples.

1 Introduction

In this study, we present a representation and processing model of human faces with head pose variations and apply it to a pose-invariant face recognition system.

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. We approximated these mappings by using a combination of linear systems: 1) subspaces of input representation spaces spanned by principal components (PC-subspace) [13, 14], and 2) linear transfer matrices between these subspaces and a pose parameter space. We call this model the *linear PCMAP model* [11]. 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 (analysis) and transformation (synthesis). The model's generalization capability for unknown head poses from a limited number of observable samples enables a continuous coverage of the pose parameter space, providing high approximation accuracy for the analysis and synthesis processes.

We also propose a novel pose-invariant face recognition system using the linear PCMAP model as the entry format for a database of known persons. The head pose of each known person in the database is aligned to an input pose by synthesizing model views whose pose is the same as the input. This synthesis uses an *analysis-synthesis chain* of learned models. As a result of the pose alignment, the recognition performance should improve against pose variations. Furthermore, there is no systematic limitation to particular discrete head poses because of the continuous coverage of the pose parameter space.

This paper is organized as follows. Section 2 gives a formal description of the linear PCMAP model and its relation to a number of previous studies. In section 3, we propose a novel pose-invariant face recognition system using the linear PCMAP model as an entry format for a database of known persons. Our system is analyzed in comparison to other standard systems in experiments with 3D facial models recorded by a Cyberware scanner. Finally, we conclude this paper by discussing the results of the analyses and our future work in section 4.

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 [4, 15, 8].

We denote a training data set by $(\vec{v}^m, \vec{\theta}^m)_{1,\ldots,M}$, 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 $(\vec{x}^m)_{1,..,M}$ and $(\vec{j}^{m,n})_{1,..,M;1,..,N}$ are independently subjected to PCA resulting in a set of PCs as orthonormal bases of shape and texture representation spaces, $(\vec{y}^p)_{1,..,P}$ and $(\vec{b}^{s,n})_{1,..,S;1,..,N}$, 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 $(\vec{y}^p)_{1,..,P_o}$ and $(\vec{b}^{s,n})_{1,..,S_o^n}$ 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,

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

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

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

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 $\vec{r}^n = \langle \vec{j}^n - \vec{j}^{0,n} | \vec{b}^{s,n} \rangle_{1 \le s \le S_o}$. 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, exploiting the correlation between the shapes and textures of faces. 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,

$$\begin{aligned} &K : (\alpha, \beta, \gamma) \mapsto \\ &(cos(\alpha), sin(\alpha), cos(\beta), sin(\beta), cos(\gamma), sin(\gamma)) \ . \end{aligned} \tag{3}$$

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,

$$\Phi = Q \cdot H \ , \tag{4}$$

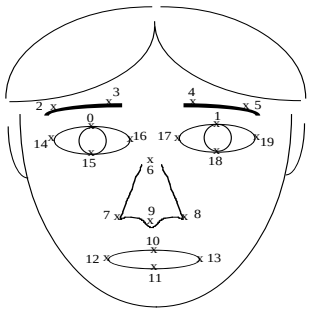

Figure 1: Definition of Facial Landmarks.

$$Q = \Phi \cdot G \ , \tag{5}$$

$$R^n = Q \cdot F^n \ , \tag{6}$$

where $R^n = (\vec{r}^{1,n}, .., \vec{r}^{M,n})^t$, $Q = (\vec{q}^1, .., \vec{q}^M)^t$, $\Phi = (\vec{\varphi}^1, .., \vec{\varphi}^M)^t = (K(\vec{\theta}^1), .., K(\vec{\theta}^M))^t$. 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 an 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

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

and the pose-to-face mapping of the synthesis process is

$$\begin{array}{ccccccc} \vec{\theta}^a & \overset{K}{\rightarrow} & \vec{\varphi}^a & \overset{Eq.(5)}{\rightarrow} & \vec{q}^a & \overset{Eq.(6)}{\rightarrow} & \vec{r}^{a,1}, .., \vec{r}^{a,N} \\ & & & & \downarrow_{Eq.(1)} & & \downarrow_{Eq.(2)} \\ & & & & \vec{x}^a & & \vec{j}^{a,1}, .., \vec{j}^{a,N} \\ & & & & \searrow_R & & \swarrow_R \\ & & & & & \vec{v}^a & \end{array} \ . \tag{8}$$

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. [10], 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 the 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 [6, 16], which was developed by Poetzsch et al. [12], performs a reverse operation that reconstructs a facial image from synthesized shape and texture representations. This operation is symbolized by the operator R in formula 8.

By connecting the analysis and synthesis stages, we create 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 an *analysis-synthesis-chain* process, and we use it in the face recognition system described in the next section. See Okada et al. [11] for performance analyses of this linear PCMAP model with samples from video sequences.

2.2 Previous Studies

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 [9]. This algorithm, however, requires a priori knowledge of 3D facial structure and its application has been limited to a small number of discrete poses. 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. [8] have 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 either transformation or a generalization capability for unknown head poses.

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 2 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 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 pose variations. Furthermore, there is no systematic limitation to particular discrete head poses due to the continuous coverage of the pose parameter space as a result of 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.

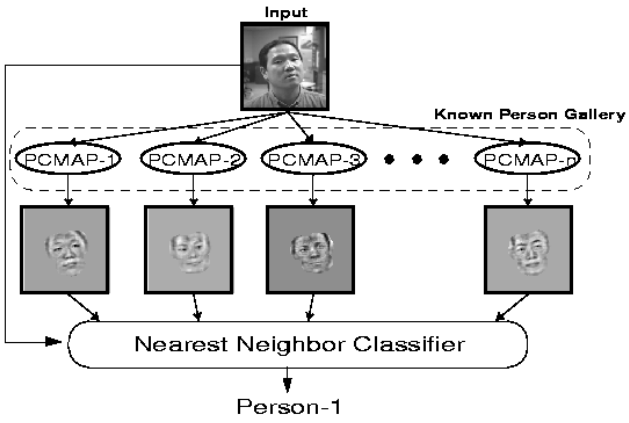

Figure 2: Pose-Invariant Recognition System with Linear PCMAP Models

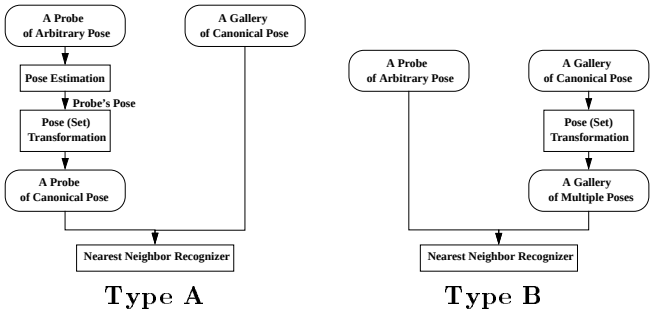

Figure 3: Two Previous Approaches for Pose-Invariant Face Recognition Systems.

3.2 Previous Studies

A number of previous studies addressed the issue of robustly recognizing faces with pose variations using a nearest neighbor recognizer. These studies estimated an input's identity by finding the most similar entry to the input from a known person's database or *gallery*. Such systems can be categorized into two types of approaches as illustrated in figure 3.

The type A approach utilizes a gallery representing each known person with a single view. We call this type the *single view model* (SVM). In order to compensate for pose variations, the head pose of an input view and gallery entries can be matched by transforming the input's pose to a *canonical pose*. Recognition systems by Maurer and von der Malsburg [9] and Lando and Edel-

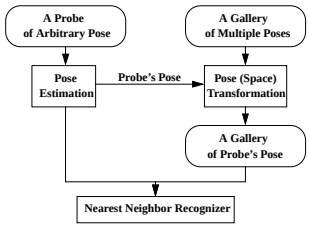

Figure 4: Our Pose-Invariant Face Recognition System.

Figure 5: Known Persons in Database.

man [7] are based on this approach.

The type B approach utilizes a gallery representing each known person with multiple views. We call this type the *multi view model* (MVM). An input's identity is estimated by finding a personal entry that contains the view most similar to the input. This multi-view gallery can be constructed from a single view gallery using a class specific transformation as shown in the figure. Beymer and Poggio proposed two systems of this type: one by manually creating a multi view gallery [2] and the other by synthesizing a multi-view gallery from a single view gallery using a class specific transformation (parallel deformation) [3].

Figure 4 illustrates our approach. This approach combines the SVM and MVM in that 1) it uses estimated information of an input pose similar to the SVM, and 2) it represents each known person by using knowledge derived from multiple views of the person similar to the MVM. With the information derived from inputs, a search space within the gallery can be greatly reduced in comparison to the MVM. Moreover, this model-based gallery is more compact than the MVM, which simply represents each known person with a set of multiple views. A special treatment for the canonical pose is not required in our approach, while it is required for both the SVM and MVM. Such knowledge is learned directly from sample statistics. Whereas the SVM and MVM cover the viewing sphere discretely, our approach provides a continuous coverage of the viewing sphere.

3.3 Experiments

3.3.1 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, shown in figure 5) are randomly picked from a 3D facial model database of Japanese faces developed at ATR. For each model, test and training samples are generated by rendering 2D view snapshots while explicitly rotating the 3D face model [5]. For the training samples, each model is rotated along only one axis at a time as shown in figure 6. Along each axis, 248 snapshots are generated, so there are 744

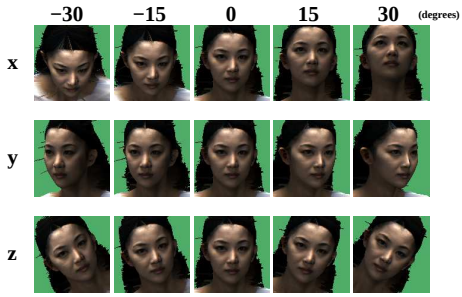

Figure 6: Training Samples.

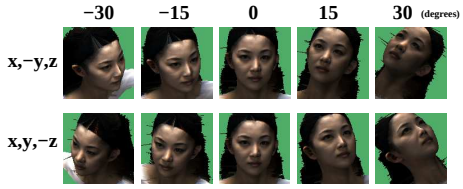

Figure 7: Test Samples.

training samples for each person. For the test samples, each model is simultaneously rotated along three axes, as shown in figure 7. 186 samples are generated for each person. 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. The head pose of each frame is directly given from the model's rotation angles. The test and training samples are appropriate for our system's evaluations since there are no measurement errors of head pose angles and landmark locations.

3.3.2 Results

Figure 8 shows the result of the performance analysis of our system of recognizing faces with pose variations. The proposed system is compared with two standard systems: i) each entry of a known person's database is represented by a single frontal view of the person (SVM), and ii) each entry is represented by multiple views of the person (all training samples used to train linear PCMAP models, MVM).E The bars in the figure show the percentages of correct identifications by the three systems when the range of head angles in test samples is limited to ±5, ±10, ±15, ±20, and ±30 degrees, respectively. The recognition rates of our system are constantly better than or equivalent to the SVM. Our system's performance is equally good in comparison to the MVM within the pose range of ±20 degrees. However, the MVM outperformed our system beyond the ±20 pose range.

4 Discussions

In this paper, we have presented a linear PCMAP model that is a manifold representation of 2D facial images with an explicit interface of pose variations. An advantage of our model is that both the analysis and synthesis pro-

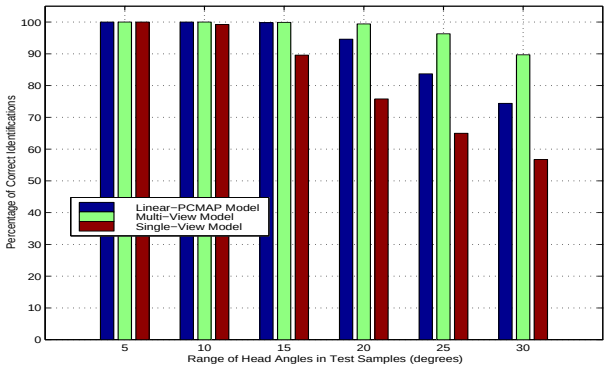

Figure 8: Percentages of Correct Identifications with 1) Linear PCMAP Model, 2) MVM, and 3) SVM as Entry Format of a Database of Known Persons

cesses continuously and smoothly cover the space of pose parameters by utilizing interpolation. Our model is capable of generalizing unknown poses from a limited number of training samples with a limited range of poses [11]. The model is also compact: the data compression ratio from a set of training samples to a learned model is approximately 60 [11].

We have 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 the SVM, which represents a known person with a single frontal view of the person. When a pose range in the test samples is within ± 20 degrees, the performance of our system is equivalent to the MVM. However, the MVM outperforms our system beyond the pose range. This is due to the fact that the effective range of the linear PCMAP model in which accurate approximations can be carried out is about ± 15 degrees [11]. The choice of the linear implementation of the PCMAP model provides the advantage of generalization for unknown poses, but unfortunately it leads to this limitation. One way to solve 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 research topics.

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 besides the identification task investigated in this paper. These scenarios include low-bandwidth visual communication systems, in which only the head pose information is sent over a network, or 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 the 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] David J. Beymer. Face recognition under varying pose. Technical Report A.I. Memo, No. 1461, Artificial Intelligence Laboratory, M.I.T., 1993.

[3] David J. Beymer and Tomaso Poggio. Face recognition from one example view. Technical Report A.I. Memo, No. 1536, Artificial Intelligence Laboratory, M.I.T., 1995.

[4] Ian Craw, Nicholas Costen, Takashi Kato, Graham Robertson, and Shigeru 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] Katsunori Isono and Shigeru 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.

[6] Martin Lades, Jan C. Vorbrueggen, Joachim Buhmann, Joerg Lange, Christoph von der Malsburg, Rolf P. Wuertz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE transactions on Computers*, 42:300–311, 1993.

[7] Maria Lando and Shimon Edelman. Generalization from a single view in face recognition. In *Proceedings of the International Workshop on Automatic Face and Ges ture Recognition*, pages 80–85, Zurich, 1995.

[8] 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.

[9] Thomas Maurer and Christoph 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.

[10] Thomas Maurer and Christoph 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.

[11] Kazunori Okada, Shigeru Akamatsu, and Christoph von der Malsburg. Analysis and synthesis of pose variations of human faces by a linear pcmap model and its application for pose-invariant face recognition system. submitted to Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 1999.

[12] M. Poetzsch, T. Maurer, L. Wiskott, and C. von der Malsburg. Reconstruction from graphs labeled with responses of gabor filters. In *Proceedings of the Internatinal Conference of Artificial Neural Networks*, pages 845–850, Bochum, 1996.

[13] 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.

[14] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[15] Thomas Vetter and Nikolaus Troje. A separated linear shape and texture space for modeling two-dimensional images of human faces. Technical Report TR15, Max-Plank-Institut fur biologische Kybernetik, 1995.

[16] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krueger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.