Pose-Invariant Face Recognition: Representing Known Persons by View-based Statistical Models

Kazunori Okada[†] and Christoph von der Malsburg[†,‡]

†Computer Science Department
University of Southern California
Los Angeles, CA 90089-2520
kazunori@organic.usc.edu

‡Institut für Neuroinformatik
Ruhr-Universität Bochum
Bochum, D-44801 Germany
malsburg@organic.usc.edu

Running Head: Pose-Invariant Face Recognition

Corresponding Author:
Kazunori Okada
University of Southern California,
Hedco Neuroscience Building 228,
Los Angeles, CA 90089-2520
213-740-7428 (tel)
213-740-5687 (fax)

Abstract

We present a framework for pose-invariant face recognition using parametric linear subspace models as stored representations of known individuals. Each model can be fit to an input, resulting in faces of known people whose head pose is aligned to the input face. The model's continuous nature enables the pose alignment to be very accurate, improving recognition performance, while its generalization to unknown poses enables the models to be compact. As a demonstration, recognition systems with two types of parametric linear model are compared using a database of 20 persons. The experimental results showed our system's robust recognition of faces with ± 50 degree range of full 3D head rotation, while compressing the data by a factor of 20 and more.

List of Symbols

- $\mathcal{A}$, $\mathcal{S}$: Analysis and synthesis mapping
- $\mathcal{M}$: Analysis-synthesis chain mapping
- $\mathcal{SS}$, $\mathcal{TS}$: Shape and texture synthesis mapping
- Ω: View-based statistical model
- LM: Linear PCMAP model
- PM: Parametric piecewise linear subspace (PPLS) model (:=$\{LM_k\}$)
- $\vec{v}$: Vectorized facial image
- $\vec{x}$: Shape vector of a facial image
- $\vec{u}_x$: Average shape vector of $\{\vec{x}\}$
- $\vec{j}^n$: Texture vector (Gabor jet) sampled at node n
- $\vec{u}_j^n$: Average texture vector of $\{\vec{j}^n\}$
- $\vec{\theta}$: 3D angle vector of a head in an image
- $\vec{u}_\theta$: Average angle vector of $\{\vec{\theta}\}$
- N: The number of facial landmarks
- Y, B^n: Shape and texture models
- F, G, H^n: Shape-to-pose, pose-to-shape, shape-to-texture transfer matrices
- P_0, S_0: The number of principal components included in shape and texture models
- $\mathcal{R}$: Image reconstruction operator
- $\mathcal{K}$: Trigonometric functional transformation
- K: The number of local models in the PPLS model
- $\vec{w}$: Weight vector for the PPLS model
- σ_k: The k-th Gaussian width of the weight function
- $\vec{x}_i$: Shape vector estimate by the i-th iteration
- $\vec{\theta}_i$: Angle vector estimate by the i-th iteration
- η: Learning rate

1. Introduction

Past studies in the field of automatic face recognition (for reviews, see [32, 5]) have revealed that our utmost challenge is to reliably recognize people in the presence of image/object variations that occur naturally in our daily life [29]. Among others, head pose variation is one of most common variations because humans and their heads can move freely. Thus, handling of head pose variation is an extremely important factor for many practical application scenarios. There have been a number of studies which specifically addressed the issue of pose invariance in face recognition [2, 28, 18, 16, 33, 3, 1, 10, 8, 27, 13, 9, 12, 35, 26]. Despite the accumulation of studies and relative readiness of the problem, scientists to date have not yet achieved reliable pose invariance especially when there is no control over subjects, one must deal with an unlimited range of full 3D pose variations.

One of most successful approaches towards pose-invariant face recognition is the multi-view approach [2, 3, 9, 35]. This approach is based on the multi-view gallery, which consists of multiple views of various poses for each known person. Pose-invariance is achieved by assuming that, for each input face, there exists a view with the same head pose as the input for each known person in the gallery. Such a multi-view gallery can be constructed by manually recording views for each person [2], by using a person-independent view-transformation to create novel views from a single view [3], or by rendering views with a 3D structural model [9, 35]. These studies have reported generally better recognition performance than other approaches, such as the single-view approach [18, 16, 13], in which each input image is transformed to a fixed, canonical head pose prior to nearest neighbor identification. The large size of the gallery is, however, a disadvantage of this approach. The recognition performance and the gallery size have a trade-off relationship; to improve the performance requires denser sampling of the continuous pose variation, increasing the gallery size. This increase of the gallery size makes it difficult to scale

the recognition systems to the large number of known people and makes the recognition process more time-consuming.

One solution to the trade-off problem is to represent each known person by a compact model. Given the multi-view gallery, each set of views of a known person can be used as training samples to learn such a model, reducing the gallery size while maintaining high recognition performance. The parametric eigenspace method of Murase and Nayar [20] and the virtual eigensignature method of Graham and Allinson [10] are successful examples of this approach. These methods represent each known person by compact manifolds in the embedded subspace of the eigenspace. Despite their good recognition performance, generalization capability is their shortcoming. Both systems utilized non-linear methods (cubic-spline for the former and radial basis function network for the latter) for parameterizing/modeling the manifolds. Such methods have a tendency to overfit peculiarities in training samples [4], compromising capability to generalize over head poses not given in training samples. This disadvantage must be overcome to avoid the curse of dimensionality problem [4], which hinders the collection of an appropriate set of training samples and the extension of the model to variations other than head pose.

Our investigation explores the model-based solution of pose-invariant face recognition using parametric linear subspace models as the representation format of known persons. The parametric linear subspace model is a type of view-based statistical model that learns the 3D nature of faces from 2D pictorial examples without actually reconstructing their 3D structure. Its parametric nature, which will be described in the next section, enables it to continuously cover the wide range of full 3D pose variation, thereby improving accuracy of the previous systems. On the other hand, its linear nature mitigates the problem for generalization of the non-linear methods, enabling it to learn from a few samples and to be compact. Our previous studies [27, 22] proposed a recognition system using the linear PCMAP (LPCMAP) model that

is a simple implementation of the parametric linear subspace model. The model's shortcoming was that its accuracy decreases as a wider range of head poses is considered. Our recent extension of the LPCMAP, the parametric piecewise linear subspace (PPLS) model [24, 25], mitigates the pose range limitation problem by piecing together a number of localized models for collectively covering a wide range of head poses accurately. The discrete local models are continuously interpolated, improving the structurally discrete methods such as the view-based eigenface by Pentland et al. [28].

This paper describes face recognition systems using the two parametric linear models and compares their recognition performance over the wide range of full 3D head rotation. In section 2, we introduce the framework of our recognition systems. Sections 3 and 4 briefly describe the LPCMAP and PPLS models, and section 5 gives the performance evaluation of the recognition systems.

2. Framework of Our Recognition Systems

The parametric linear subspace model consists of bidirectional, continuous, multivariate, mapping functions between a vectorized facial image $\vec{v}$ and 3D head angles $\vec{\theta}$. We call a mapping from the image to angles *analysis mapping*, and its inverse *synthesis mapping*. An application of the analysis mapping can be considered as *pose estimation* and that of the synthesis mapping as *pose transformation* or *facial animation*,

$$\begin{aligned} \mathcal{A}^{\Omega} &: \vec{v} \xrightarrow{\Omega} \vec{\theta}, \\ \mathcal{S}^{\Omega} &: \vec{\theta} \xrightarrow{\Omega} \vec{v}(\Omega). \end{aligned} \tag{1}$$

Ω denotes the model's data entities that are learned from training samples and also symbolizes a learned model. Each model is personalized by learning it with pose-varying samples of a specific person. Both analysis and synthesis mappings become specific to a person because they are constructed with the personalized model Ω that encodes specificities of the person's facial appearance. The synthesis mapping output $\vec{v}(\Omega)$ exhibits personal appearance that solely depends on Ω, while its head pose is given by an input. Details of these mappings will be described in the next two sections.

Given an arbitrary person's face, a learned model can be fit against it by concatenating the analysis and synthesis mappings. We call this model matching process the *analysis-synthesis-chain*,

$$\mathcal{M}^{\Omega} : \vec{v} \xrightarrow{\Omega} \vec{\theta} \xrightarrow{\Omega} \vec{v}(\Omega). \tag{2}$$

The output of the analysis-synthesis chain is called the *model view*. It provides a facial view of the person learned in Ω whose head pose is aligned to the input. This process not only fits a learned model to the input but also gives simultaneously a 3D pose estimate that can be used for other application purposes. Note that, because the analysis mapping is also personalized, fitting a model to a different person's face may introduce pose estimation errors, resulting in a sub-optimal model view. To overcome this shortcoming, $\mathcal{A}^{\Omega}$ needs to be replaced by a person-independent analysis mapping [21]. For the purpose of face recognition, however, this does not pose serious problems because the errors are small due to the geometrical proximity of different faces. Moreover, the sub-optimal model views of different people only help to single out the correct face.

Figure 1 illustrates our framework for pose-invariant face recognition. The framework employs the parametric linear model as the representation of a known person. We call a database of P known people, as a set of learned personalized models $\{\Omega_p | p = 1, .., P\}$, the *known-person*

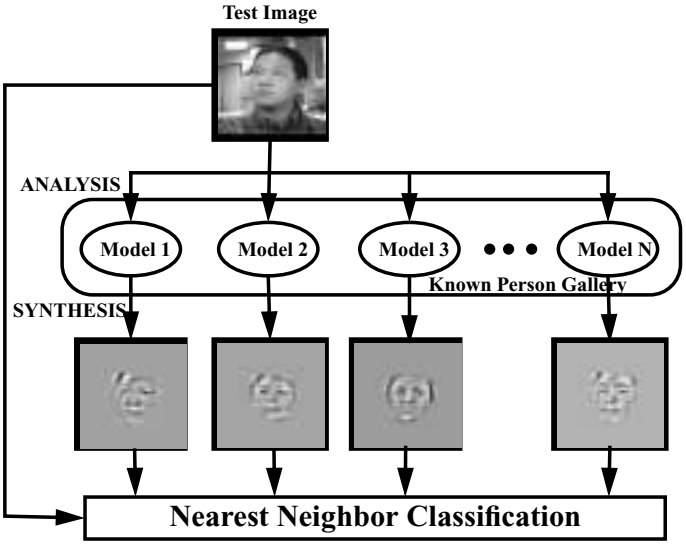

Figure 1. Recognition framework with parametric linear models.

gallery. Given a test image of an arbitrary person with an arbitrary head pose, each model in the gallery is matched against the image by using its analysis-synthesis-chain process. The process results in pose-aligned model views of all known persons. After this pose alignment, the test image is subjected to a nearest neighbor classifier for identification.

Figure 2 compares our system with a multi-view system using a gallery of three known persons. The top row displays model views of three learned models; the bottom row displays the best views of each known person that are most similar to the test image. Decimal numbers shown adjacent to the images denote their similarity to the test. Because the test image's head pose was not present in the gallery, the best views do not have the same head pose as the test, resulting in a wrong identification for the multi-view system. On the other hand, our system, in which each model is learned by the same samples stored in the multi-view gallery, identifies the test image correctly. This is realized by the model's continuous and generalizable nature, which results in model views whose head pose is better aligned to the test than the multi-view system.

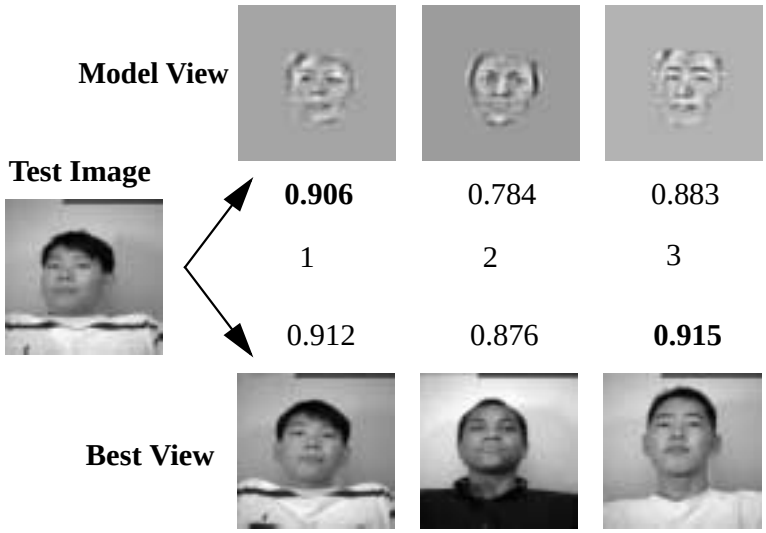

Figure 2. An example of face recognition with pose variation.

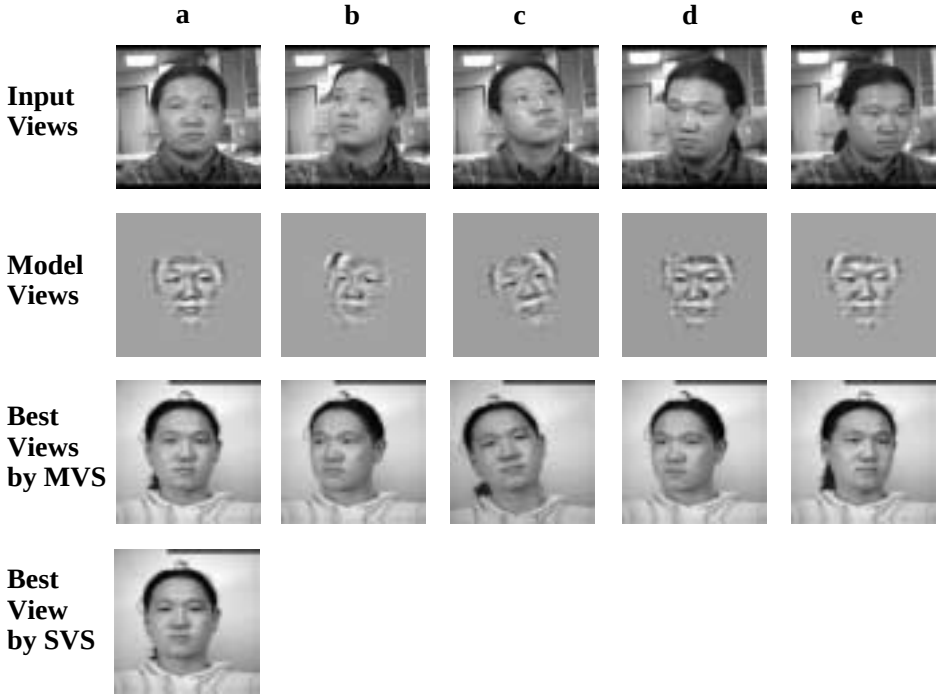

Figure 3. Comparison of three recognition frameworks in terms of pose-alignment ability. Model views shown in the second raw are given by our method. MVS: multi-view system; SVS: single-view system. See texts for their description.

Model Type	a	b	c	d	e	std.dev.
Model Views	0.915	0.871	0.862	0.891	0.878	0.0184
Best Views by MVS	0.930	0.872	0.876	0.913	0.897	0.0220
Best View by SVS	0.926	0.852	0.816	0.878	0.862	0.0359

Table 1. Cosine similarity values between the input and different types of model or best view in Figure 3.

Main difference of our framework from others is the continuous pose-alignment of model views to arbitrary inputs. Figure 3 and Table 1 illustrate this advantage in comparison with two other recognition frameworks: the multi-view (MVS) and single-view (SVS) systems. Given facial images with arbitrary head poses shown in the first raw, a parametric linear model, learned for this face, can provide model views whose head pose is well aligned to the inputs. MVS provides the most similar view (best view) to the input among the training samples used to learn the model, while SVS provides always the same frontal view that represents the person single-handedly. The figure shows that our model is appeared to provide better pose-alignment than the two other systems. Table 1 shows actual facial similarity values between the inputs and the three different types of model or best view. The standard deviation shown in the right column indicates the degree of pose invariance in each framework. The parametric linear model provided the smallest standard deviation among the three, demonstrating the model's favorable characteristics towards pose-invariance.

3. The Linear PCMAP Model

The LPCMAP model [22] is a parametric linear subspace model, which covers the whole parameter space of head poses by a single model. It consists of a combination of two linear systems:

1) *linear subspaces* spanned by principal components (PCs) of training samples and 2) *linear transfer matrices*, which associate projection coefficients of training samples onto the subspaces and their corresponding 3D head angles.

A LPCMAP model LM consists of the following data entities learned from training samples,

$$LM := \{\vec{u}_x, \{\vec{u}_j^n\}, \vec{u}_\theta, Y, \{B^n\}, F, G, \{H^n\}\}, \tag{3}$$

where $\vec{u}_x$ is an average shape representation as an array of 2D coordinates of N facial landmarks; $\vec{u}_j^1, .., \vec{u}_j^N$ are average texture representations as a set of Gabor jets sampled at the N landmarks [34, 23]; $\vec{u}_\theta$ is an average 3D head angle vector; Y is a *shape model* as a row matrix of the first $P_0 \leq 2N$ shape PCs; $B^1, .., B^N$ are *texture models* as row matrices of the first $S_0 \leq L$ texture PCs; and $F, G, H^1, .., H^N$ are shape-to-pose, pose-to-shape and shape-to-texture transfer matrices, respectively.

The analysis mapping function $\mathcal{A}(\vec{v})$ is given by relating the 3D head angles only to the shape representations,

$$\hat{\vec{\theta}} = \mathcal{A}^{LM}(\vec{v}) = \vec{u}_\theta + \mathcal{K}^{-1}(F \cdot Y \cdot (\mathcal{D}_x(\vec{v}) - \vec{u}_x)), \tag{4}$$

where $\mathcal{K}^{-1}$ extracts 3D angles from their trigonometric functions and $\mathcal{D}_x$ extracts a shape representation from a facial image.

The synthesis mapping function $\mathcal{S}(\vec{\theta})$ is given by relating the 3D head angles to the shape

coefficients and the shape coefficients to the texture coefficients,

$$\begin{aligned}\hat{\vec{v}} &= \mathcal{S}^{LM}(\vec{\theta}) = \mathcal{R}(\hat{\vec{x}}, \{\hat{\vec{j}}^n | n = 1, .., N\}),\\ \hat{\vec{x}} &= \mathcal{SS}^{LM}(\vec{\theta}) = \vec{u}_x + Y^t \cdot G \cdot \mathcal{K}(\vec{\theta} - \vec{u}_\theta),\\ \{\hat{\vec{j}}^n\} &= \mathcal{TS}^{LM}(\vec{\theta})\\ &= \{\vec{u}_j^n + B^n \cdot H^n \cdot G \cdot \mathcal{K}(\vec{\theta} - \vec{u}_\theta) | n = 1, .., N\},\end{aligned} \tag{5}$$

where $\hat{\vec{x}}$ and $\{\hat{\vec{j}}^n | n = 1, .., N\}$ denote synthesized shape and texture representations, $\mathcal{R}$ reconstructs an image from a pair of the shape and texture representations [30], and $\mathcal{K}$ transforms 3D angles to a vector of their trigonometric functions.

Finally, the analysis-synthesis-chain function $\mathcal{M}(\vec{v})$ is given by concatenating Eq. (4) and Eq. (5),

$$\begin{aligned}\hat{\vec{v}} &= \mathcal{M}^{LM}(\vec{v})\\ &= \mathcal{R}(\mathcal{SS}^{LM}(\mathcal{A}^{LM}(\vec{v})), \mathcal{TS}^{LM}(\mathcal{A}^{LM}(\vec{v}))).\end{aligned} \tag{6}$$

4. The Parametric Piecewise Linear Subspace Model

The parametric piecewise linear subspace (PPLS) model [24] extends the LPCMAP model by using the piecewise linear approach [31]. It consists of a set of local linear models, each of which provides continuous analysis and synthesis mappings whose accuracy may be limited to a narrow parameter range due to its linearity. In order to cover a wide range of non-linear pose variation, the model pieces together a number of the local models distributed over the pose parameter space. The global mappings are constructed by weighted-averaging of outputs of the local mappings, maintaining their continuous nature and enabling them to generalize to unknown poses by interpolation.

A piecewise linear subspace model PM consists of a set of K local linear models in the format

of the above-described LPCMAP,

$$PM := \{LM_k | k = 1, .., K\}. \tag{7}$$

We assume that the local models are learned by training data sampled from appropriately distanced local regions of the *3D angle space*: the 3D parameter space spanned by the head angles. Each set of the local training samples is associated with a *model center*, the average 3D head angles $\vec{u}_\theta^{LM_k}$, which specifies the learned model's location in the 3D angle space. Missing components of shape representations due to large head rotations are handled by the *mean-imputation method* [17], which fills in each missing component by a mean computed from all available data at the component dimension.

The analysis mapping function of the PPLS model is given by averaging K local pose estimates with appropriate weights,

$$\hat{\vec{\theta}} = \mathcal{A}^{PM}(\vec{v}) = \sum_{k=1}^{K} w_k \mathcal{A}^{LM_k}(\vec{v}). \tag{8}$$

Similarly, the synthesis mapping function is given by averaging K locally synthesized samples with the same weights,

$$\begin{aligned} \hat{\vec{v}} &= \mathcal{S}^{PM}(\vec{\theta}) = \mathcal{R}(\hat{\vec{x}}, \{\hat{\vec{j}}^n\}) \\ \hat{\vec{x}} &= \mathcal{SS}^{PM}(\vec{\theta}) = \textstyle\sum_{k=1}^{K} w_k \mathcal{SS}^{LM_k}(\vec{\theta}), \\ \{\hat{\vec{j}}^n\} &= \mathcal{TS}^{PM}(\vec{\theta}) = \textstyle\sum_{k=1}^{K} w_k \mathcal{TS}^{LM_k}(\vec{\theta}). \end{aligned} \tag{9}$$

A vector of the weights $\vec{w} = (w_1, .., w_K)$ in Eq. (8) and Eq. (9) must be responsible for localizing the output space of the models, since the model's outputs themselves are continuous. For this purpose, we use a normalized Gaussian function of distance between an input pose and

each model center,

$$\begin{aligned} w_k(\vec{\theta}) &= \frac{\rho_k(\vec{\theta}-\vec{u}_\theta^{LM_k})}{\sum_{k=1}^{K}\rho_k(\vec{\theta}-\vec{u}_\theta^{LM_k})}, \\ \rho_k(\vec{\theta}) &= \frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{\|\vec{\theta}\|^2}{2\sigma_k^2}), \end{aligned} \tag{10}$$

where σ_k denotes the k-th Gaussian width. The weight value reaches maximum when the input pose coincides with one of the model centers; it decays as the distance increases. σ_k is set by the standard deviation of the 3D head angle vectors for LM_k and determines the extent to which each local model influences the outputs $\hat{\vec{\theta}}$ and $\hat{\vec{v}}$.

The analysis-synthesis-chain function $\mathcal{M}(\vec{v})$ is given by connecting an analysis output to a synthesis input,

$$\begin{aligned} \hat{\vec{v}} &= \mathcal{M}^{PM}(\vec{v}) \\ &= \mathcal{R}(\mathcal{SS}^{PM}(\mathcal{A}^{PM}(\vec{v})), \mathcal{TS}^{PM}(\mathcal{A}^{PM}(\vec{v}))) \end{aligned} \tag{11}$$

Note that Eq. (8) cannot be solved by evaluating its r.h.s. because the weights are computed as a function of an unknown $\vec{\theta}$. To overcome this problem, we formulate a gradient descent-based solution of the equation. Let a shape vector $\vec{x}$ be an input to the algorithm. Also let $\vec{x}_i$ and $\vec{\theta}_i$ denote the shape and angle estimates by the i-th iteration. The algorithm iterates the following formulae until the mean-square error $\|\Delta\vec{x}_i\|^2$ becomes sufficiently small.

$$\begin{aligned} \Delta\vec{x}_i &= \vec{x} - \vec{x}_i, \\ \Delta\vec{\theta}_i &= \textstyle\sum_{k=1}^{K} w_k(\vec{\theta}_i)\mathcal{A}'^{LM_k}(\Delta\vec{x}_i), \\ \vec{\theta}_{i+1} &= \vec{\theta}_i + \eta\Delta\vec{\theta}_i, \\ \vec{x}_{i+1} &= \textstyle\sum_{k=1}^{K} w_k(\vec{\theta}_{i+1})\mathcal{SS}^{LM_k}(\vec{\theta}_{i+1}), \end{aligned} \tag{12}$$

where η is a learning rate and $\mathcal{A}'$ is a slight modification of Eq. (4) that has a shape vector interface. The initial conditions $\vec{x}_0$ and $\vec{\theta}_0$ are given by the local model whose center shape $\vec{u}_x^{LM_k}$ is most similar to $\vec{x}$. In this article, we do not attempt to give a proof of this algorithm's

convergence to the global minima. It is obvious that the convergence property depends on the specific distribution of given local models in the 3D angle space. Within the experimental setting described in the next section, however, our convergence tests with the above-described initial conditions resulted in no trappings into local minima that were significantly distanced from the global minima.

5. Experiments

5.1. Data Set

For evaluating our system's performance over various head poses, we must collect a very large number of samples with controlled head poses, which is not an easy task. For mitigating this difficulty, we use 3D face models pre-recorded by a Cyberware scanner. Given such data, relatively faithful image samples with an arbitrary, but precise, head pose can easily be created by image rendering. We used 20 heads randomly picked from the ATR-Database [14], as shown in figure 4.

For each head, we created 2821 training samples. They consist of 7 local sample sets each of which covers a pose range of ± 15 degrees at one-degree interval. These local sets are distributed over the 3D angle space such that they collectively cover a pose range of ± 55 degrees along each axis of 3D rotations; their model centers are distanced by ± 40 degrees from the frontal pose (origin of the angle space). We also created 804 test samples for each head. In order to test the model's generalization capability to unknown head poses, we prepared the test samples whose head poses were not included in the training samples. Head angles of some test samples were in-between multiple local models and beyond their ± 15 degree range. They cover a pose range of ± 50 degrees. For more details of the data, see our previous reports [21, 24].

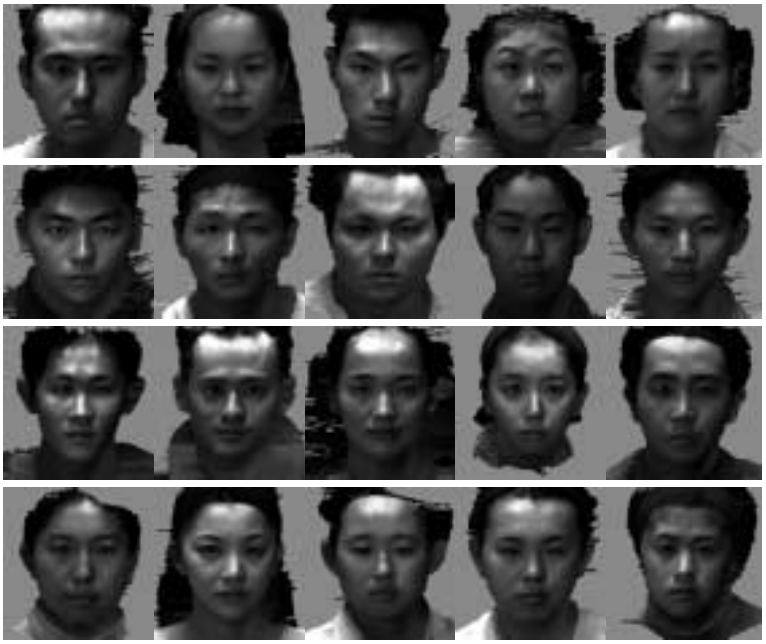

Figure 4. 20 frontal views rendered from the 3D face models.

For each sample, the 2D locations of 20 landmarks of inner facial parts, such as eyes, nose and mouth, are derived by rotating the 3D landmark coordinates, initialized manually, and by projecting them onto an image plane. The explicit rotation angles of the heads also provide 3D head angles of the samples. The rendering system provides the self-occlusion information. Up to 10% of the total landmarks were self-occluded for each head.

5.2. Results

For comparison, we constructed four recognition systems with 20 known persons: 1) the single-view system (SVS), which represents each known person by a single frontal view, 2) the LPCMAP system with a gallery of LPCMAP models, 3) the PPLS system with a gallery of PPLS models, and 4) the multi-view system (MVS), which represents each person by various views of the person. The LPCMAP, PPLS and MVS are constructed by using the same 2821 training samples per person; the SVS serves as a base-line. For both models, P_0 and S_0 are set

Test Samples	Identification	Compression
SVS	59.9±10.6%	0.035%
LPCMAP	91.6±5.0%	0.74%
PPLS	98.7±1.0%	5%
MVS	99.9±0.2%	—

Table 2. Average correct-identification and relative compression rates for four different systems.

	PPLS	LPCMAP
Unknown: $\mathcal{M}(\vec{v})$	98.7±1.0%	91.6±5.0%
Known: $\mathcal{S}(\vec{\theta})$	99.3±0.7%	92.4±4.0%

Table 3. Identification rates when head pose of tests is unknown or given as ground-truth.

to 8 and 20, respectively. The PPLS models consist of 7 local models and perform 500 iterations with η set to 0.01 for each test sample. Each pair of views are compared by an average of normalized dot-product similarities between the corresponding Gabor jet's magnitudes.

Table 2 summarizes the results of our recognition experiments. Identification rates in the table are averaged over the 20 persons; the compression rates represent the size of the known-person gallery relative to the MVS. The results show that recognition performance of the PPLS system was more robust against the given pose variation (7% higher rate) than the LPCMAP system. Performance of our model-based systems was much better than the baseline and slightly lower than the MVS. The high identification rate of the MVS was perhaps due to the dense sampling in the 3D angle space and the virtue of our Gabor jet-based texture representation and similarity metric. Nonetheless, identification rates of the PPLS and MVS were almost the same while the former compressed the data by a factor of 20.

For some applications, head pose information of test faces can be given as ground-truth prior to identification. In this case, the pose-aligned model views can be created by our model's synthesis mapping function instead of the analysis-synthesis-chain. Table 3 compares average

identification rates of the two cases. The results show that the knowledge of head poses gave a slight increase in recognition performance, however the increase was minimal. The fact that the use of ground-truth pose information did not greatly improve recognition performance supports our assumption about the usage of personalized pose estimation as described in section 2.

6. Conclusion

This article presents a framework for recognizing faces with large 3D pose variations. It utilizes the parametric linear subspace model for representing each known person in the gallery. The analysis-synthesis-chain function of the models is used to fit them to an arbitrary input, resulting in pose-aligned model views of each known person. The continuous and generalizable nature enables our models to provide accurate pose-alignment and to be compact at the same time. The linear construction of the model was emphasized to facilitate its generalization to unknown head poses, however the intrinsic non-linearity of the problem decreased the model's accuracy. In order to accurately model the non-linearity while maintaining the system's linearity, we introduced the PPLS model which is based on the piecewise linear approach. The experimental results showed the robustness of our recognition system with the PPLS model against large and full 3D head pose variations covering ± 50 degree rotation along each axis. While significantly compressing the data size, the PPLS system performed better than the LPCMAP system and similar to an equivalent multi-view system, indicating the effectiveness of our recognition framework.

The statistics of our recognition experiments must, however, be treated carefully in terms of variations across different people because the number of known people in our experiments was relatively small and our samples included some artificialities which might accidentally increase the performance. Although we do not expect that the artificialities have greatly influenced

our system's performance, we must further evaluate our systems with a larger database of real faces.

Our recognition systems utilize pixel-wise landmark locations for representing facial shape and deriving head pose information. In reality, finding landmark locations in static facial images with arbitrary head pose is an ill-posed problem. Gabor jet-based landmark tracking system [19] can be used to provide an accurate landmark positions [21], however it requires the landmarks to be initialized by other methods. Pose-specific graph matching [15] provides an another solution but with much lower precision. As future work, we plan to develop a pose-invariant landmark finding system using our parametric linear models.

Although the presented work concentrated only on head pose variations, our future goal must address other types of variation such as illuminations and expressions for realizing more robust systems. There have been a number of recent progresses on both illumination variations [6, 9] and expression variations [7, 11]. However, an issue on combining the variation-specific solutions into a unified system that is robust against all types of variation has scarcely been investigated. We believe that our simple and general approach will benefit us for extending the presented system towards this goal.

Acknowledgments

The authors thank Shigeru Akamatsu and Katsunori Isono for making their 3D face database available for this study. and the developers of the FLAVOR libraries for providing the software platform. This study was partially supported by ONR grant N00014-98-1-0242.

References

[1] M. S. Bartlett and T. J. Sejnowski. Viewpoint invariant face recognition using independent component analysis and attractor networks. In *Neural Information Processing Systems: Natural and Synthetic*, volume 9, pages 817–823. MIT Press, 1997.

[2] D. Beymer. Face recognition under varying pose. Technical Report A.I. Memo, No. 1461, Artificial Intelligence Laboratory, M.I.T., 1993.

[3] D. Beymer and T. Poggio. Face recognition from one example view. Technical Report 1536, Artificial Intelligence Laboratory, M.I.T., 1995.

[4] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, New York, 1995.

[5] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

[6] P. Debevec, T. Hawkins, H. P. Tchou, C. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of Siggraph*, pages 145–156, 2000.

[7] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–988, 1999.

[8] S. Duvdevani-Bar, S. Edelman, A. J. Howell, and H. Buxton. A similarity-based method for the generalization of face recognition over pose and expression. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 118–123, Nara, 1998.

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 277–284, Grenoble, 2000.

[10] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In *Face Recognition: From Theory to Applications*, pages 446–456. Springer-Verlag, 1998.

[11] H. Hong. *Analysis, Recognition and Synthesis of Facial Gestures.* PhD thesis, University of Southern California, 2000.

[12] F. J. Huang, Z. Zhou, H. J. Zhang, and T. Chen. Pose invariant face recognition. In *Proceedings of Fourth International Conference on Automatic Face and Gesture Recognition*, pages 245–250, Grenoble, France, 2000.

[13] H. Imaoka and S. Sakamoto. Pose-independent face recognition method. In *Proceedings of IEICE Workshop of Pattern Recognition and Media Understanding*, pages 51–58, June 1999.

[14] K. Isono and S. Akamatsu. A representation for 3D faces with better feature correspondence for image generation using PCA. In *Proceedings of IEICE Workshop on Human Information Processing*, pages HIP96–17, 1996.

[15] N. Krüger, M. Pötzsch, and C. von der Malsburg. Determination of face position and pose with a learned representation based on labeled graphs. Technical report, Institut fur Neuroinformatik, Ruhr-Universität Bochum, 1996.

[16] 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.

[17] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data.* Wiley, New York, 1987.

[18] 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.

[19] 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.

[20] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[21] K. Okada. *Analysis, Synthesis and Recognition of Human Faces with Pose Variations.* PhD thesis, University of Southern California, 2001.

[22] K. Okada, S. Akamatsu, and C. von der Malsburg. Analysis and synthesis of pose variations of human faces by a linear PCMAP model. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 142–149, Grenoble, 2000.

[23] 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, 1998.

[24] K. Okada and C. von der Malsburg. Analysis and synthesis of human faces with pose variations by a parametric piecewise linear subspace method. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 761–768, Kauai, 2001.

[25] K. Okada and C. von der Malsburg. Parametric piecewise linear subspace method for processing facial images with 3D pose variations. In preparation, 2002.

[26] K. Okada and C. von der Malsburg. Pose-invariant face recognition with parametric linear subspaces. To appear in *the International Conference on Automatic Face and Gesture Recognition*, 2002.

[27] K. Okada, C. von der Malsburg, and S. Akamatsu. A pose-invariant face recognition system using linear pcmap model. In *Proceedings of IEICE Workshop of Human Information Processing*, pages 7–12, Okinawa, 1999.

[28] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. Technical report, Media Laboratory, M.I.T., 1994.

[29] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.

[30] M. Pötzsch, 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.

[31] S. Schaal and C. G. Atkeson. Constructive incremental learning from only local information. *Neural Computing*, 10:2047–2084, 1998.

[32] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell. Connectionist models of face processing: a survey. *Pattern Recognition*, 27:1209–1230, 1994.

[33] T. Vetter and N. 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.

[34] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[35] W. Y. Zhao and R. Chellappa. SFS based view synthesis for robust face recognition. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 285–292, Grenoble, 2000.