Locally Linear Regression for Pose-Invariant Face Recognition

Xiujuan Chai, Shiguang Shan, *Member, IEEE*, Xilin Chen, *Member, IEEE*, and Wen Gao, *Senior Member, IEEE*

***Abstract*—The variation of facial appearance due to the viewpoint (/pose) degrades face recognition systems considerably, which is one of the bottlenecks in face recognition. One of the possible solutions is generating virtual frontal view from any given nonfrontal view to obtain a virtual gallery/probe face. Following this idea, this paper proposes a simple, but efficient, novel locally linear regression (LLR) method, which generates the virtual frontal view from a given nonfrontal face image. We first justify the basic assumption of the paper that there exists an approximate linear mapping between a nonfrontal face image and its frontal counterpart. Then, by formulating the estimation of the linear mapping as a prediction problem, we present the regression-based solution, i.e., globally linear regression. To improve the prediction accuracy in the case of coarse alignment, LLR is further proposed. In LLR, we first perform dense sampling in the nonfrontal face image to obtain many overlapped local patches. Then, the linear regression technique is applied to each small patch for the prediction of its virtual frontal patch. Through the combination of all these patches, the virtual frontal view is generated. The experimental results on the CMU PIE database show distinct advantage of the proposed method over Eigen light-field method.**

***Index Terms*—Dense sampling, face recognition, locally linear regression (LLR), local patch, virtual frontal view.**

I. Introduction

FACE recognition has been studied for more than three decades. The state-of-the-art recognition technologies can achieve very high accuracy under restricted environment, such as frontal faces under controlled lighting conditions [1]–[3]. However, most of current face recognition systems fail under uncontrolled cases (e.g., outdoor with uncooperative subjects), since they are pretty sensitive to pose, lighting, occlusion, aging, and other variations. Especially, pose problem has been one of the bottlenecks for most current face recognition technologies.

Pose difference induces large variation of the appearance even for the same person. The distinction is often more remarkable than that caused by the difference of identity under the same pose. Therefore, the typical appearance-based methods, such as Eigenface [4], degrade dramatically when nonfrontal probes match against the enrolled frontal faces.

Many approaches have been proposed to deal with pose problem. Among them, the view-based methods are widely used [5]–[9]. For instance, view-based Eigenface had been proposed to extend the Eigenface to handle the pose problem. One disadvantage of the view-based method is that it usually needs multiple face images with different poses for each subject, which is often impractical for real-world applications.

Gross *et al.* propose the Eigen light-field (ELF) method to tackle the pose problem [10], [11]. This algorithm first estimates the ELFs of the subject's head from the input images. Matching between the probe and gallery is then performed by means of comparing the coefficients of the ELFs. Compared with view-based methods, ELF needs an extra independent training set (different from the gallery) that contains multiple images of varying poses for each subject. While in the recognition stage, one face is recognized even if he/she has only one image in the gallery. So far, the ELF method has achieved state-of-the-art results on the pose subsets of the CMU PIE face database.

Generating virtual view is another possible solution for pose-invariant face recognition. By generating virtual view, one can either normalize all the face images to a predefined pose (e.g., frontal) or expand the gallery (or the training set) to cover large pose variations. Simply speaking, there are two strategies to generate virtual view: 3-D model-based method [12]–[18] and learning-based method [19]–[25].

Since the variations in appearance caused by pose are closely related to the 3-D face structure, it is a natural idea to recover the 3-D model from the input 2-D face image. Thus, virtual views under any viewpoint are easily generated by using graphic rendering techniques [12], [13], [16]. For 3-D face model recovery, the 3-D morphable model (3-D MM) method [12], [13], proposed by Blanz and Vetter, is one of the most successful technologies. In this method, the prior knowledge of the face shape and texture is modeled by PCA. Then any novel face can be modeled by the linear combination of the prototypes, in which the corresponding shape and texture are expressed by the exemplar faces respectively. The specific 3-D face can be recovered automatically from one or more photographs by simultaneously optimizing the shape, texture and mapping parameters through an analysis-by-synthesis strategy. In FRVT 2002, virtual frontal view generation based on 3-DMM method significantly improved the performance of face recognition on pose variation tasks [1]. However, it is time consuming for most real-world applications. To reduce the complexity, Jiang *et al.* [18] propose a simplified

Manuscript received August 3, 2006; revised February 18, 2007. This work was supported in part by the National Natural Science Foundation of China under Contract 60332010 and Contract 60673091, in part by the 100 Talents Program of CAS, in part by the Natural Science Foundation of Beijing municipal under Contract 4061001 and Contract 4072023, and in part by ISVISION Technology Co., Ltd. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhigang (Zeke) Fan.

X. Chai is with the Harbin Institute of Technology, Harbin 150001, China (e-mail: xjchai@jdl.ac.cn).

S. Shan and X. Chen are with the Institute of Computing Technology and the Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, Beijing 100080, China (e-mail: sgshan@jdl.ac.cn; xlchen@jdl.ac.cn).

W. Gao is with the Harbin Institute of Technology, Harbin 150001, China, and also with Peking University, Beijing 100871, China, and the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (e-mail: wgao@jdl.ac.cn).

Digital Object Identifier 10.1109/TIP.2007.899195

version of 3-D MM to reconstruct the specific 3-D face from a frontal face. Lee [17] also realizes the 3-D model reconstruction by a 3-D deformable model which is composed of the edge model, the color region model and the wireframe model. Generic 3-D face model has been used in many papers to generate the virtual views to tackle the pose problem, such as [14] and [15]. The illumination Cone method [16] can also reconstruct the accurate shape and albedo for a specific person from at least seven images under a fixed pose but with different lighting conditions.

Unlike 3-D model-based methods, learning-based approaches generally try to learn how to estimate a virtual view directly in 2-D domain [19]–[25]. The active appearance model (AAM) fits an input face image to the prelearned face model, which consists of separated shape and appearance models [19]. By extending it to view-based AAM, novel views of the input face can be synthesized and used for pose-invariant face recognition [20]. Beymer and Poggio propose an example-based algorithm to synthesize novel views from single image and apply them to face recognition [22], [23]. Similar methods are proposed by Vetter *et al.* [24], [25], in which a face image is separated into shape vector and texture vector (shape-free face patch), and the linear object classes (LOC) [21] is applied to them respectively to generate virtual shape and texture under a novel pose by using a basis set of 2-D prototypes. Then the virtual "rotated" images are generated easily by combining the generated shape and texture. Evidently, the quality of the novel virtual view heavily depends on the accuracy of the face alignment, i.e., the separation of shape and texture. Unfortunately, building accurate pixel-wise correspondence between face images is still an open problem, which has prevented this method from further practical applications.

In this paper, by formulating virtual view generation as a prediction problem, we propose a novel locally linear regression (LLR) method to efficiently generate the virtual frontal view from a given nonfrontal face image. Simply speaking, we partition the whole nonfrontal face image into multiple patches and apply linear regression to each patch to predict its corresponding frontal patch. The method is inspired by the idea that the linear mapping between nonfrontal patches and frontal patches maintains better than that of the global case in the case of coarse alignment. Compared with the approach in [25], LLR is more efficient since only simple linear regression is needed. In addition, it is much easier to implement, considering that LLR requires only the centers of the two eyes for alignment rather than accurate face alignment, as is mandatory for LOC method.

The basic ideas of LLR were initially reported in [26]. This paper further extends LLR from the original nonoverlapping patch partitioning to dense sampling and also reports more extensive and improved experimental results. The remaining part of the paper is organized as follows. Section II presents the details of the proposed LLR-based virtual view generation method. Section III provides the experimental results on CMU PIE database [28]. The conclusion and discussion are given in Section IV.

II. Virtual View Generation Based on Locally Linear Regression

Given a nonfrontal facial image, how do we generate its virtual frontal view? In this paper, this task is formulated as a general prediction framework, which predicts the mapping from a nonfrontal face to its frontal counterpart. We show that, in case the given samples are well aligned, there exists an approximated linear mapping between two images of one person captured under variable poses. This mapping is consistent for all persons, if only their facial images are pixel-wisely aligned. Unfortunately, the pixel-wise correspondence between images is still a challenging problem. Therefore, in most real-world face recognition systems, face images are only aligned coarsely based on very few facial landmarks, such as the two eye centers. In this case, the above-mentioned assumption of linear mapping no longer holds theoretically, since it becomes a complicated nonlinear mapping. What this paper tries to address is just the inability of global linear methods in the case of coarse alignment. To this end, we present a piecewise linear solution, LLR, as an approximation to the ground-truth nonlinear mapping. The main idea of the proposed method lies in the intuitive observation that, by partitioning the whole face surface into multiple patches, linearity of the mapping for each patch is increased because of the consistent normal and the better controlled alignment.

In this section, we first present the theoretical analysis and the basic ideas of the proposed method, based on which globally linear regression (GLR) and LLR are then described in detail.

A. Basic Ideas

In this section, we first try to formulate the relationship between a nonfrontal face image with a specific pose and its frontal counterpart based on the image formation model. We show that there exists a linear mapping between the nonfrontal face image and the frontal one for a specific face under the same illumination. Though the linear operators for a specific pose are different for distinct persons, they should quite similar due to the natural similarity of all the 3-D face shapes. Thus, a regression-based solution is intuitively applicable to estimate, for a novel person, the linear operator, which can then be used for virtual view generation.

Formally, given a specific face, we can safely assume that its 3-D surface is Lambertian. For a fixed lighting source $\vec{s}$, the intensity for each surface point (x, y, z) is independent of viewpoint and can be computed as

$$\Gamma(x, y, z) = \rho(x, y, z) \cos \alpha \tag{1}$$

where $\rho(x, y, z)$ is the albedo of the given point, α is the angle between the normal $\vec{n}(x, y, z)$ and the lighting directions $\vec{s}(x, y, z)$.

Let Γ be the vector concatenating all the intensities $\Gamma(x, y, z)$ of the surface points in scan-line order. Then, given a specific viewpoint, the 2-D face image $\mathbf{I}$ can be obtained from Γ by a linear orthogonal projection procedure, which is operated by selecting the visible points [23]. This procedure is expressed as

$$\mathbf{I} = \mathbf{D}\Gamma \tag{2}$$

where $\mathbf{D}$ is a matrix which drops points occluded under the given viewpoint Q, i.e., the projection operator. Evidently, the operator $\mathbf{D}$ depends on both the viewpoint and the 3-D structure of the specific face. Specifically, it should be a $m \times n$ matrix, where m is the pixel number in $\mathbf{I}$ and n is the surface point number in Γ. Herein, $\mathbf{D}$ is defined as $\mathbf{D}_{ij} = 1$, iff the jth surface

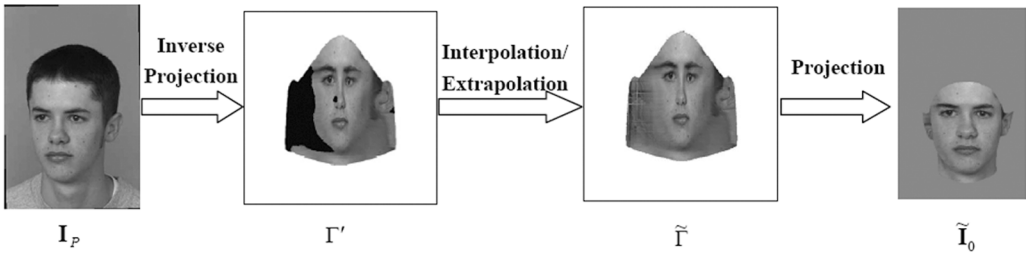


Fig. 1. Procedure for the generation of the virtual frontal view from one nonfrontal view.

point in Γ is visible under the viewpoint Q and is projected as the ith pixel in $\mathbf{I}$. For all other elements, $\mathbf{D}_{ij} = 0$.

In our case, assume that we only concern two distinct viewpoints, i.e., nonfrontal and frontal, indexed by P and 0, respectively. We obtain the frontal image vector $\mathbf{I}_0$ and nonfrontal one $\mathbf{I}_P$, respectively, as

$$\mathbf{I}_0 = \mathbf{D}_0\Gamma \tag{3}$$

$$\mathbf{I}_P = \mathbf{D}_P\Gamma. \tag{4}$$

To predict $\mathbf{I}_0$ from $\mathbf{I}_P$, we need to recover Γ from $\mathbf{I}_P$. However, this is obviously an ill-posed problem with no unique solution since m is always less than n. Intuitively, only the visible points are recoverable, while all the occluded surface points are not. Thus, we can define an occluded version of Γ with missing data as

$$\Gamma' = \mathbf{D}_P^T\mathbf{I}_P = \mathbf{D}_P^T\mathbf{D}_P\Gamma. \tag{5}$$

In (5), $\mathbf{D}_P^T\mathbf{D}_P$ is an $n \times n$ matrix and equal to a modified identity matrix whose ith "1" in diagonal is replaced to "0" if the ith surface point is occluded. Thus, Γ' is different from Γ only in the invisible surface points, whose intensities are set to "0" in Γ'.

The next step is to estimate Γ from Γ'. Relatively, the percentage of invisible points is small. Therefore, we can infer the intensities of the missing points by a linear interpolation/extrapolation operation $\Re$ based on the corresponding neighbor points. Thus, the estimation of Γ can be computed as follows:

$$\tilde{\Gamma} = \mathbf{D}_P^T\mathbf{I}_P + \Re\mathbf{I}_P = \left(\mathbf{D}_P^T + \Re\right)\mathbf{I}_P. \tag{6}$$

Here, $\Re$ is the neighborhood relation matrix of $n \times m$.

Finally, by substituting the estimation of Γ in (6) into (3), we can obtain

$$\begin{aligned}\mathbf{I}_0 &= \mathbf{D}_0\Gamma \\ &\approx \mathbf{D}_0\left(\left(\mathbf{D}_P^T + \Re\right)\cdot\mathbf{I}_P\right) = \left(\mathbf{D}_0\mathbf{D}_P^T + \mathbf{D}_0\Re\right)\cdot\mathbf{I}_P.\end{aligned} \tag{7}$$

Rewriting (7), we get the estimation of $\mathbf{I}_0$ as follows:

$$\tilde{\mathbf{I}}_0 = \mathbf{A}_P\mathbf{I}_P \tag{8}$$

where

$$\mathbf{A}_P = \mathbf{D}_0\mathbf{D}_P^T + \mathbf{D}_0\Re. \tag{9}$$

Consequently, we conclude that, given a nonfrontal face image of a specific face under certain pose, there exists an approximate linear mapping which transforms a nonfrontal face image into a frontal one. The above reasoning procedure is illustrated in Fig. 1, in which an example image is used to show how Γ', $\tilde{\Gamma}$, and $\tilde{\mathbf{I}}_0$ are obtained, respectively, from a nonfrontal face image.

It is obvious that, when the viewpoint is fixed, $\mathbf{A}_P$ is determined solely by the 3-D geometry of the specific person. Therefore, $\mathbf{A}_P$ should be distinct for different persons, unless these faces are well aligned in 3-D. In addition, the more similar the 3-D geometries of two faces are, the more similar the two mappings are. Fortunately, all faces are similar holistically in terms of the main organs' spatial configuration and their shapes. Thus, if only we align all the 2-D face images carefully according to some facial landmarks and with the aid of some generic 3-D face model, faces can be regarded as being aligned coarsely in 3-D. Hence, in this case, the linear mapping $\mathbf{A}_P$ for different persons can be considered to be equal approximately.

Based on the above analysis, when given a nonfrontal face $\mathbf{I}_P$, we can easily recover the virtual frontal face $\mathbf{I}_0$ if only $\mathbf{A}_P$ is known. Thus, the problem is converted to the estimation of the linear mapping $\mathbf{A}_P$. Theoretically, $\mathbf{A}_P$ can be computed according to (9) based on geometric inference, for instance, given a generic 3-D face model. In practice, however, this method is intractable due to the complicated occlusion and interpolation/extrapolation computation. In this paper, we turn to learning-based method to estimate the mapping $\mathbf{A}_P$, based on a training set containing some coarsely aligned sample pairs of nonfrontal view and its corresponding frontal view. Then, by using the learned mapping $\mathbf{A}_P$, a virtual frontal view can be generated from the single input nonfrontal face image. This idea is illustrated intuitively in Fig. 2.

Mathematically, the task to learn $\mathbf{A}_P$ is a typical prediction problem, which can be solved naturally by regression method. In this paper, the linear regression is applied to learn the mapping from examples, since, as analyzed above, the mapping to be solved is approximately linear in the case that all face images are aligned coarsely (based on the two eye centers in our case). Considering that the face images are processed as a whole, we call this implementation as GLR.The concrete description for GLR is given in Section II-B.

In addition, as mentioned above, the linear degree of the mapping $\mathbf{A}_P$ is highly related to the consistency of the 3-D face shapes. Therefore, the predicted $\mathbf{A}_P$ in the GLR method is expectably not very accurate; thus, its performance for virtual view generation is not quite satisfactory. To weaken this demerit, we

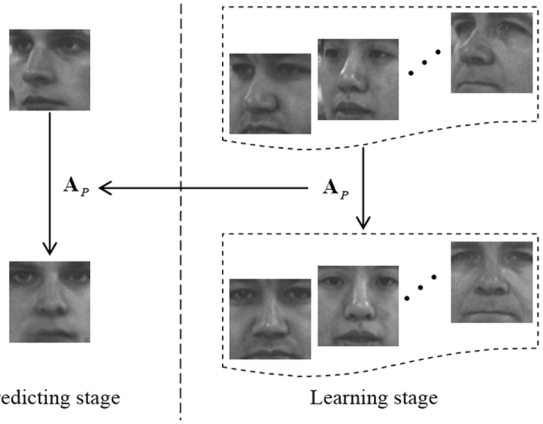


Fig. 2. Illustration of the proposed method on the virtual frontal view generation.

further propose a patch-wise version of GLR, called the LLR method. In LLR, faces are divided into patches, which results in more consistency in shapes and, thus, more linearity. LLR is described in detail in Section II-C.

B. Globally Linear Regression

Let $\{(\mathbf{\Phi}_0, \mathbf{\Phi}_P)\}$ be the training set, which is composed of the corresponding samples pairs under frontal pose and some specific nonfrontal pose. Here, $\mathbf{\Phi}_0 = (\mathbf{I}_0^1 \quad \mathbf{I}_0^2 \quad \cdots \quad \mathbf{I}_0^N)$ denotes the matrix whose ith column is the vectorized ith frontal face image in the frontal face set, and $\mathbf{\Phi}_P = (\mathbf{I}_P^1 \quad \mathbf{I}_P^2 \quad \cdots \quad \mathbf{I}_P^N)$ is the matrix formed by the corresponding nonfrontal face images under pose P. Note that $\mathbf{I}_P^i$ is the counterpart image vector of $\mathbf{I}_0^i$ from the same person but with different pose. Let m denote the number of pixels of an image. Theoretically, the appropriate mapping $\mathbf{A}_P$ can be sought by the following optimization procedure:

$$\min_{\mathbf{A}_P} \sum_{i=1}^{N} \left\| \mathbf{I}_0^i - \mathbf{A}_P \mathbf{I}_P^i \right\|. \tag{10}$$

Generally, $m \gg N$; thus, the linear mapping $\mathbf{A}_P$ can be estimated by the following linear regression procedure (least-square solution) [27]

$$\mathbf{A}_P = \mathbf{\Phi}_0 \mathbf{\Phi}_P^{\perp} \tag{11}$$

where

$$\mathbf{\Phi}_P^{\perp} = \left(\mathbf{\Phi}_P^T \mathbf{\Phi}_0\right)^{-1} \mathbf{\Phi}_P^T \tag{12}$$

is the pseudo inverse of $\mathbf{\Phi}_P$.

Once the linear mapping function $\mathbf{A}_P$ is estimated from the training set, when given any face image $\mathbf{I}_P$ with appointed pose P, its corresponding virtual frontal image $\mathbf{I}_0$ can be computed

$$\mathbf{I}_0 = \mathbf{A}_P \mathbf{I}_P = \mathbf{\Phi}_o \mathbf{\Phi}_P^{\perp} \mathbf{I}_P. \tag{13}$$

In the case all the faces are aligned accurately according to dense correspondence, Vetter's method [25] is equivalent to (11). However, the proposed method reaches the same goal by a different route. That is, (11) is deduced separately according to different problem formulation methods. In Vetter's method, it is reached from the definition of "linear object class" in 3-D, while in our method it is derived by formulating the task as a regression problem in 2-D image space.

To understand the above procedure further, we rewrite (13) as follows:

$$\mathbf{I}_0 = \mathbf{\Phi}_0 \boldsymbol{\alpha} \tag{14}$$

where

$$\boldsymbol{\alpha} = \mathbf{\Phi}_P^{\perp} \mathbf{I}_P \tag{15}$$

is a coefficient vector of N-dimension.

These two equations clearly show that the regression-based virtual view generation can be decomposed into two steps: one is the solving of the reconstruction coefficients of the input image $\mathbf{I}_P$ in the P pose image space via (15), and the other is the virtual frontal view prediction by using (14). Let us go a step further on (15) to understand the coefficient vector $\boldsymbol{\alpha}$ more intuitively. Easy to understand, it is actually the least square solution of the optimization problem minimizing the following residue function:

$$\varepsilon(\boldsymbol{\alpha}) = \left\| \mathbf{I}_P - \mathbf{I}_P^{\mathrm{Rec}} \right\|^2 \tag{16}$$

where

$$\mathbf{I}_{\mathrm{P}}^{\mathrm{Rec}} = \mathbf{\Phi}_P \boldsymbol{\alpha} = \sum_{j=1}^{N} \mathbf{I}_P^j \alpha_j \tag{17}$$

is the projection of $\mathbf{I}_P$ in the P pose image space. Intuitively, this optimization procedure aims at seeking a group of coefficients, which can best represent the input image in the P pose image space. Thus, from this analysis and (14) and (15), one can know that, in the proposed method, the frontal view corresponding to the input nonfrontal face image $\mathbf{I}_P$ with pose P is generated through a linear combination by using the same coefficients reconstructing the $\mathbf{I}_P$ in the P pose image space.

Note that, for each pose, the mapping matrix $\mathbf{A}_P$ can be learned at the training stage only one time. Therefore, the prediction of the frontal view from its nonfrontal view can be very efficient, if only the input nonfrontal face image is well aligned to the training samples. As is well accepted in face recognition domain, one can simply align all the faces under the same pose by fixing the eyes at the same positions, keeping the aspects of the face, and resizing to a predefined image size, and then the normalized nonfrontal face images as a whole are fed into the above prediction.

However, the virtual frontal view generated by GLR is not as realistic as expected, as shown in Fig. 5(c). The reason is that GLR treats the face image as a whole and uses only the eye centers and a general 3-D cylinder face model to perform the alignment. In other words, the fact that different person has different (local) geometric shape results in inaccurate alignment between faces leads to non-negligible difference among the linear mappings for different persons. Easy to understand, this will degrade the approximate effect of GLR. To improve the approximation further, we propose the LLR in next section.

C. Locally Linear Regression

As mentioned above, the unrealistic effect of GLR partially comes from the large difference of the mappings due to the non-negligible difference of various persons in terms of shape and albedo, since in this paper we align very coarsely the faces according to the eye positions for the purpose of practice. Another aspect to explain this imperfectness is from the viewpoint of the regression. As can be seen from above section, GLR actually predicts virtual frontal view by linear combination in the frontal view space but using the coefficients learned in the nonfrontal view space. Intuitively, this means, if some face in the training set is similar enough to the input face in terms of the normal and albedo of all the surface points, this specific face will be assigned a dominantly large weight, and, thus, the virtual frontal view of the input face can be better predicted. In practice, we can only collect a training set of limited size. Therefore, we may not be able to find totally similar faces for any input face, especially for the faces whose geometrical shapes are largely different with the general faces.

However, we notice that it is much easier to find two faces partially similar. Therefore, one natural way to eliminate the above-mentioned problem of GLR is to divide the whole face into many local patches and conduct regression patch by patch. Conducting regression in patch-wise mode leads to two additional merits.

First, in each surface patch, the normals and albedos of different persons are more similar compared with those of whole face region, in the case of coarse alignment. The reason behind is the natural spatial correlation because of the smooth of the face surface, which implies that small surface patches should be still similar even if the faces were not accurately aligned. Larger similarity of the surface patches leads to better linearity of the mapping we concern.

Second, the prediction problem for each patch becomes much easier than GLR because of the much lower dimension of patches. This is especially important when the given training set is of limited size, because the results of regression will generally be better if the feature dimension is reduced tremendously while the examples are unvaried.

Specifically, in the learning stage, the whole region of any frontal face is partitioned into many uniform blocks. Then, for each frontal block, its corresponding block in the nonfrontal face can be obtained by the aid of a generic 3-D cylinder face model. In the predicting stage, for a given nonfrontal face image, the same partitioning criterion as the training images with the same pose is applied to get multiple small patches. Then, for each of these nonfrontal patches, its corresponding frontal patch is predicted by using linear regression. Finally, all the virtual frontal patches are combined together to form the whole virtual frontal view. This procedure can be formally formulated as follows:

First, given the training set, we need to divide each face image into M blocks (rectangle patches). The uniform partitioning criterion is adopted for all the frontal faces. These patches can be either overlapped or adjacent. For each frontal patch, its nonfrontal counterpart is expected to contain surface points of the same semantics as those in the frontal patch. This semantic correspondence can be coarsely built by the aid of a general 3-D cylinder face model as shown in Fig. 3 in our case.

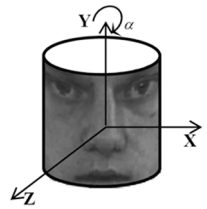


Fig. 3. Average 3-D cylinder face model.

Then, given an input image $\mathbf{I}_P$ whose pose is P, we partition it into M small patches $\mathbf{I}_P = (\mathbf{I}_{(1,P)} \quad \mathbf{I}_{(2,P)} \quad \cdots \quad \mathbf{I}_{(M,P)})$ as is done on the training images with P pose. Predicting the corresponding ith frontal patch $\mathbf{I}_{(i,0)}$ for the ith nonfrontal patch $\mathbf{I}_{(i,P)}$ follows two steps.

Step 1: Estimate the reconstruction coefficients for the ith patch in $\mathbf{I}_P$ in the specific patch space by

$$\boldsymbol{\alpha}_i = \boldsymbol{\Phi}^{\perp}_{(i,P)} \mathbf{I}_{(i,P)} \tag{18}$$

where $\boldsymbol{\Phi}_{(i,P)} = (\mathbf{I}^1_{(i,P)} \quad \mathbf{I}^2_{(i,P)} \quad \cdots \quad \mathbf{I}^N_{(i,P)})$ is the matrix containing the ith patch sampled from the training images with P pose.

Step 2: Compute the virtual frontal patch as follows:

$$\mathbf{I}_{(i,0)} = \boldsymbol{\Phi}_{(i,0)} \boldsymbol{\alpha}_i \tag{19}$$

where $\boldsymbol{\Phi}_{(i,0)} = (\mathbf{I}^1_{(i,0)} \quad \mathbf{I}^2_{(i,0)} \quad \cdots \quad \mathbf{I}^N_{(i,0)})$ is the matrix containing the ith patch sampled from the frontal training images.

After performing such prediction for each patch in the $\mathbf{I}_P$, we need to combine all the virtual frontal patches into a whole vector. As mentioned above, patches can be sampled densely with some overlapping, which can alleviate the blocking effect. In this case, for any pixel sampled by several patches simultaneously, its intensity is calculated as the mean of the specific pixels in these overlapping patches, as (20) shows

$$g_0(x,y) = \frac{\sum_{m=1,\cdots,M} g_{(m,0)}(x,y)}{\sum_{m=1,\cdots,M} \mathrm{Index}_{(m,0)}(x,y)} \tag{20}$$

where $\mathrm{Index}_{(m,0)}(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \mathbf{I}_{(m,0)} \\ 0, & \text{otherwise} \end{cases}$, and $g_{(m,0)}(x,y)$ is the recovered intensity value of the corresponding pixel in the mth patch $\mathbf{I}_{(m,0)}$, which is set to 0 if the point (x,y) is not presented in the mth patch.

The above two-step procedure is also illustrated in Fig. 4 by using an example image. The first step, i.e., the reconstruction step, is shown below the middle dashed line, while the second step, i.e., the prediction step is shown above the dashed line. Note that, as illustrated in Fig. 4, due to the pose variation, the sizes of the corresponding frontal and nonfrontal patches might be different.

III. Experimental Results

The proposed method is evaluated on CMU PIE face database. Two experiments are carried out to show the effectiveness

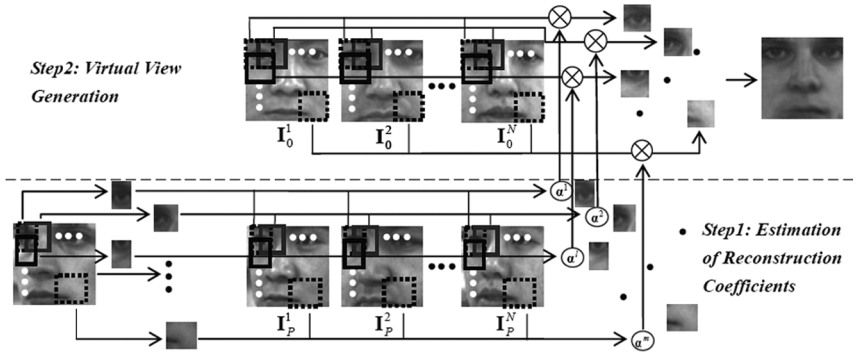


Fig. 4. Flow chart of the proposed virtual view generation method based on LLR.

TABLE I
SEVEN POSE CLASSES AND THE FACE EXAMPLES IN CMU PIE DATABASE

	7 pose classes						
	P37	P05	P07	P27	P09	P29	P11
Pose Angle	Yaw: Right 45°	Yaw: Right 22.5°	Pitch: Up	Frontal	Pitch: Down	Yaw: Left 22.5°	Yaw: Left 45°
Examples							

of the proposed LLR. In Experiment I, we visualize the generated virtual frontal views to show its reality. In Experiment II, we evaluate its ability to improve the performance of face recognition system. This section first introduces the dataset for validation and then details the two experiments in Sections III-B and III-C, respectively.

A. Database for Evaluations

In our experiments, seven pose subsets of CMU PIE database are used, which covers the pose yawing over ± 45 degree and the pitching variations in depth [28]. They are the pose set 37 and 11 (yawing about ± 45 degree), 05 and 29 (yawing about ± 22.5 degree), 07 and 09 (pitching about ± 20 degree), and 27 (near frontal), respectively. Each pose class includes 68 subjects. The pose class and face examples are given in Table I.

In our experiments, leave-one-out strategy is used for generating the virtual frontal views. Once the virtual frontal views are generated, the general face recognition system designed for frontal faces can be applied for face classification.

B. Virtual View Generation

First of all, visual reality of the virtual view should be checked since LLR aims at generating virtual frontal view from nonfrontal view. So, the following experiment is designed to generate some examples for demonstration. Before showing the results, we first detail the setup of the experiment.

In our experiment, face images are all normalized to the same size of 60×60 after fixing the eye positions and keeping the aspects of the face, as shown in Fig. 5. Given such a fixed face size, we still have two parameters to be determined: one is the size of the patches; the other is the sampling step for the patch. To see their effects, one example is shown in Fig. 5 and Table II. Fig. 5(a) is the input nonfrontal view from PIE P29 subset, while

(a) (b) (c)

Fig. 5. Example results for GLR.

TABLE II
EXAMPLE RESULTS AND THE CORRESPONDING PSNR (DECIBELS) OF LLR WITH VARIOUS PATCH SIZE AND STEP

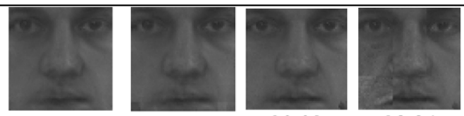
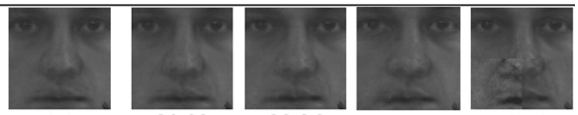
Patch Size	Step 1	Step 5	Step 10	Step 20	Step 30
10×10	 26.40	26.50	26.89		
20×20	 30.51	30.41	30.02	28.84	
30×30	 30.05	30.09	29.86	29.98	28.58

Fig. 5(b) is its ground truth frontal view from PIE P27 subset. Fig. 5(c) illustrates the prediction results of GLR, i.e., using the whole 60×60 patch for predicting (b) from (a). More prediction results of LLR with various patch sizes and sampling steps are shown in Table II. Three patch sizes (10×10, 20×20, and 30×30) and five sampling steps (1, 5, 10, 20, and 30) are evaluated for further analysis.

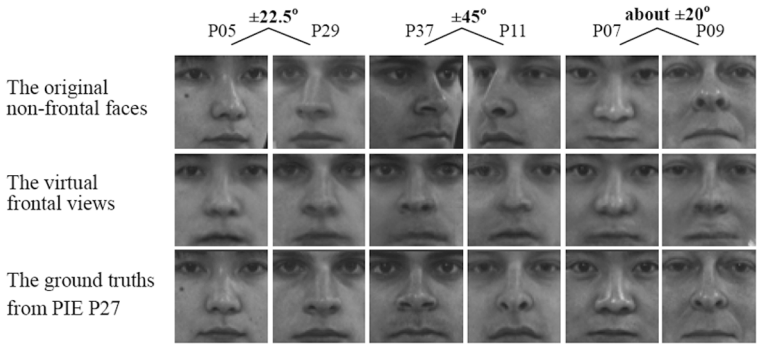


Fig. 6. Examples of LLR-based virtual frontal view generation (examples are from six CMU PIE pose sets respectively).

To evaluate the prediction accuracy quantitatively, in Table II, below each virtual frontal view, we also give the PSNR value of this prediction with the ground truth frontal face, i.e., Fig. 5(b). The PSNR is calculated using the follow equation:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\left[\frac{1}{\mathrm{WH}} \sum_{i=1}^{W} \sum_{j=1}^{H} [f(i,j) - f'(i,j)]^2\right]} \tag{21}$$

where $f(i,j)$ is the ground truth frontal view, $f'(i,j)$ is the predicted virtual frontal view, and W and H are the width and height of the image respectively.

Intuitively, the patch size should be neither too large nor too small. A too large patch may cause the break of linear assumption, while a too small patch might result in serious mis-alignment since we use a generic 3-D model. Therefore, a small patch may result in more artifacts, as can be observed from the examples in Table II, especially for patch size 10×10. On the contrary, large patch results in more blurring effect, especially for the nose and mouth part, as can be seen from the GLR prediction.

As for the sampling step, it is a tradeoff between over-smoothing and blocking effect, as can be seen from these examples. Especially, more blocking effects can be observed, when the patches are sampled without any overlapping, i.e., the step is as large as the patch size. On the contrary, a small step, such as 1 pixel, always results in very smoothing face images.

From Table II, one can see that, among these results, the PSNR is the largest for patch size 20×20 and step 1 or 5 pixels. Our experience with more example images shows that, from the point of view of both the visual effects and PSNR, patch size 20×20 with step 1 or 5 achieves the best results. Since step 1 is more time consuming than step 5, we use the patch size 20×20 and step 5 in our following experiments for pose-invariant face recognition. More examples are given in Fig. 6 to show the visual effect of LLR for different poses by using the patch size of 20×20 and step of 5 pixels. In Fig. 6, the images in the top row are the original pose images from the six PIE pose subsets. The middle row shows the virtual frontal views of the corresponding nonfrontal images in the top row. The images in the bottom are the real frontal images. From these results, one can see that the virtual frontal views are generally very similar to the real frontal face images even for the large rotation in depth (with yaw about 45°). Especially, the occlusion part in the nonfrontal view has been recovered very reasonably even for 45° rotation.

Additionally, in the preprocessing stage of our method, only two eye positions are needed to align the faces. Though, the coarse alignment results in some blurring effect for the mouth region, it greatly facilitates the fully automation of the proposed method. What is more, the virtual views are satisfactory not only visually but also for pose-invariant face recognition, as will be validated in next section.

C. Pose-Invariant Face Recognition From Virtual View

After nonfrontal face images are converted to virtual frontal view, pose-invariant face recognition can be easily achieved by using the virtual frontal views instead of all the nonfrontal face images. Easy to understand, such a strategy implies that the proposed method can be regarded as a preprocessing procedure independent of the following feature extraction and classifier design. Therefore, LLR can be combined with any face recognition technologies. In this paper, Fisherfaces method [29] is employed to validate the effectiveness of the proposed method, considering that Fisherfaces method has been one of the most successful face recognition approaches. We also compare the performance of LLR + Fisherface with that of ELF method.

1) Face Recognition With PCA + FLDA *Method:* In face recognition research, Fisher linear discriminant analysis (FLDA) has been recognized as one of the most successful methods. In FLDA, the input face image is transformed to a subspace where the between-class scatter is maximized and the within-class scatter is minimized by maximizing the Fisher separation criterion. When designing a FLDA classifier, one has to deal with the within-class scatter matrix carefully, because it may be singular. To avoid the singularity problem, PCA is first conducted to reduce the dimensionality to be less than N-C, where N is the number of training examples, and C is the number of classes. The PCA transformed features are then fed into the final FLDA for classification.

Both PCA and FLDA need training images to learn the subspaces. Strictly speaking, training set should be totally separated from the testing images. Since we aim to perform face recognition on CMU PIE database in our experiments; therefore, we do not use any images in the CMU PIE database for training both PCA and LDA. Instead, we select three subsets from the FERET pose database, i.e., "ba" (frontal), "be"(right rotation of 15°) and "bf"(left rotation of 15°)[30], to form the training set. There are totally 200 subjects in these three datasets, with one image per subject in each subset; thus, there are totally 600 training images

Fig. 7. Example images from the FERET database for the training of $\mathrm{PCA} + \mathrm{FLDA}$ method.

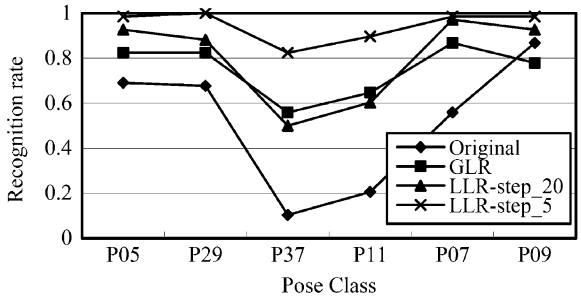


Fig. 8. Performance comparison of pose-invariant face recognition methods on PIE database.

of 200 subjects. These 600 face images are used to train both the PCA and FLDA subspaces. These subspaces are then used to extract features for CMU PIE face images. Some example training images are shown in Fig. 7. Considering that the imaging conditions between the FERET and CMU PIE databases are different, the evaluation results on PIE database are expected to have a good generalizability.

After PCA and FLDA models are obtained using the FERET face images, face recognition based on $\mathrm{PCA} + \mathrm{FLDA}$ is performed on the 68 subjects in the CMU PIE dataset. In the experiment, the frontal facial images in P27 dataset form the gallery, while the nonfrontal face images in P05, P07, P09, P11, P29, and P37 are used as probes to match against the frontal images in the gallery. We compare four different recognition modes: without preprocessing (i.e., using directly the original nonfrontal images as input), GLR-based frontal view generation, LLR frontal view generation without overlapping sampling (i.e., with 20×20 patch and 20 pixels step), and LLR frontal view generation sampling with 20×20 patch and 5 pixels step. Hereafter, these four modes are abbreviated as "Original," "GLR," "$\mathrm{LLR} - \mathrm{step\_20}$," and "$\mathrm{LLR} - \mathrm{step\_5}$," respectively. For the "Original" mode, the nonfrontal probe face images are directly projected to PCA and FLDA subspaces to extract features. For the other three modes, features are extracted by projecting the generated virtual frontal views into the PCA and FLDA subspaces. For all the methods, the extracted FLDA features are compared with those of the images in the gallery using cosine similarity, and nearest neighborhood classifier is exploited as the final classifier. The recognition results of these methods are shown in Fig. 8.

From Fig. 8, we clearly find that recognition performance can be impressively improved by using the virtual frontal views generated via LLR instead of the original nonfrontal face images. Especially, LLR based on dense sampling with 5 pixels step performs best. This is particularly obvious for face images with larger rotation, e.g., on P37 and P11, on which "LLR-step_5" mode outperforms distinctly the "original" mode, "GLR" and "LLR-step_20." In addition, compared with GLR, LLR generally performs better, if only patches are sampled densely with some overlapping. The possible reasons lie in that locally linear assumption is better satisfied and dense sampling can remove blocking effect efficiently. Note that, the GLR works also generally better than the "Original" mode with only one exception, i.e., on the P09 (looking down pose) subset.

TABLE III
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND THE ELF

Probe Sets	Methods	
	LLR-step5 with PCA+LDA	ELF method [11] (3-P Normalization)
P05	98.5%	88%
P29	100%	86%
P37	82.4%	74%
P11	89.7%	76%
P07	98.5%	100%
P09	98.5%	87%
Mean	94.6%	85.1%

2) Comparisons With the ELF Method: The ELF algorithm is one of the most representative methods for recognizing faces across poses [11]. Therefore, in this section, we compare the results of our "$\mathrm{LLR} - \mathrm{step\_5}$" with those of ELF on the same pose subsets of PIE database. The comparison results are shown in Table III. Note that in this table, the results of ELF are directly cited from [11], in which the experiment is done under the following configuration: All the face images are aligned according to three points, i.e., the two eyes and the nose tip. Half of the subjects in PIE database are randomly selected for training, and the recognition is performed on the rest 34 subjects. For these 34 subjects, images in the P27 subset form the gallery, while the images in the other pose subsets are matched against the gallery.

From Table III, one can find that the proposed method outperforms ELF method on all probe sets but the P07 subset. In summary, the average recognition rate of our method is 9.5% higher than that of the ELF method. In addition, compared with ELF method, our method needs only two eyes for alignment; thus, it is easier to be implemented. From these comparisons, we can safely conclude that the proposed LLR method can impressively facilitate the pose-invariant face recognition.

IV. SUMMARY AND FUTURE WORK

Aiming at pose-invariant face recognition, this paper proposes a method of generating virtual frontal view from a nonfrontal face image. We first justify that, for a specific face, there exists an approximate linear mapping between the frontal face image and its nonfrontal face image. By formulating the task as a prediction problem, we propose a solution based on linear regression and extend it to LLR. In LLR, face images are divided into densely sampled patches and linear regression is performed on these patches for better prediction. The effectiveness of the proposed method is evaluated on CMU PIE database from two aspects, i.e., the subjective visual reality of prediction, and the objective quantitative validation for pose-invariant face recognition. The results show that the proposed method can impressively improve the performance of face recognition across poses.

As formulated in this paper, given a nonfrontal face image, the proposed method is essentially an example-based learning strategy for the prediction of its frontal counterpart. Easy to understand, the mapping between frontal view and nonfrontal view should be a complicated nonlinear function. This paper only studied its approximation based on piecewise linear model; therefore, a natural alternative way is to predict using some nonlinear regression method or neural network. This is one of our future works.

It is easy to understand from the two-stage interpretation of regression that the proposed method is essentially a subspace-based method. In other word, the virtual frontal view is generated by a linear combination of the frontal views in the training set, while none of the pixel intensities in the virtual view are taken directly from the input side view. Compared with warping-based method, LLR handles occlusion naturally, but the virtual views might be not as realistic as warping-based method. Therefore, how to combine it with warping-based method is an interesting future work.

One of the shortcomings of the proposed method is that, as a view-based method, LLR requires separate models for each pose, which means it requires a lot of memory to store the learnt mapping matrices. In practice, we must first consider carefully how many separate pose models should be built and how these models can be compressed in order to save storage required.

Though the proposed method is easy to be implemented because only two eyes are needed for face alignment, the pose of the input nonfrontal face image is assumed to be known. This implies that one has to use a front-end procedure to estimate the pose of the input image. Fortunately, there have been many methods for pose estimation. It will be also our future work to integrate the pose estimation with the proposed method to construct a fully automatic pose-invariant face recognition system.

ACKNOWLEDGMENT

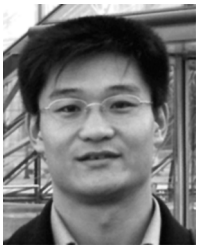
The authors would like to thank the Associate Editor and the anonymous reviewers, whose comments helped to improve the paper greatly.

REFERENCES

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, May 1995.
[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature surveys," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.
[3] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002: Evaluation report," presented at the FRVT, Mar. 2003, NIST.
[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurol.*, vol. 3, no. 1, pp. 71–96, 1991.
[5] H. Murase and S. Nayar, "Learning and recognition of 3D objects from appearance," in *Proc. Qualitative Vision Workshop*, New York, Jun. 1993, pp. 39–50.
[6] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspace for face recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Seattle, WA, 1994, pp. 84–91.
[7] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects form appearance," *Int. J. Comput. Vis.*, vol. 14, no. 1, pp. 5–24, 1995.
[8] S. McKenna, S. Gong, and J. Collins, "Face tracking and pose representation," in *Proc. Brit. Machine Vision Conf.*, Edinburgh, U.K., 1996, pp. 755–764.
[9] Z. Zhou, J. HuangFu, H. Zhang, and Z. Chen, "Neural network ensemble based view invariant face recognition," *J. Comput. Study Develop.*, vol. 38, no. 9, pp. 1061–1065, 2001.
[10] R. Gross, I. Matthews, and S. Baker, "Eigen light-fields and face recognition across pose," in *Proc. 5th Int. Conf. Auto. Face and Gesture Recognition*, Washington, DC, 2002, pp. 3–9.
[11] R. Gross, I. Mattews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.
[12] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. SIGGRAPH*, New York, 1999, pp. 187–194.
[13] V. Blanz and T. Vetter, "Face recognition based on fitting a 3-D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
[14] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in *Proc. 4th Int. Conf. Auto. Face and Gesture Recognition*, Grenoble, France, 2000, pp. 285–292.
[15] G. Feng and P. C. Yuen, "Recognition of head-&-shoulder face image using virtual frontal-view image," *IEEE Trans. Syst., Man, Cybern. A, Cybern.*, vol. 30, no. 6, pp. 871–883, Jun. 2000.
[16] A. S. Georghiades, P. N. Belhumeur, and D. J. Keiegman, "From few to many: Illumination cone models for face recognition under variable lighting and poses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
[17] M. W. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model," *Pattern Recognit.*, vol. 36, no. 8, pp. 1835–1846, 2003.
[18] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face reconstruction," *Pattern Recognit.*, vol. 38, no. 6, pp. 787–798, 2005.
[19] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[20] T. Cootes, K. Walker, and C. Taylor, "View-based active appearance models," in *Proc. 4th Int. Conf. Auto. Face and Gesture Recognition*, Washington, DC, 2000, pp. 227–238.
[21] T. Poggio and T. Vetter, "Recognition and structure from one 2-d model view: Observations on prototypes, object classes, and symmetries," Tech. Rep. A. I. Memo, Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, No. 1347, 1992.
[22] D. Beymer, "Face recognition under varying pose," Tech. Rep. A. I. Memo, Artif. Intell. Lab., Mass. Inst. Technol., Cambridge, No. 1461, 1993.
[23] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. 5th Int. Conf. Computer Vision*, 1995, pp. 500–507.
[24] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 733–742, Jul. 1997.
[25] T. Vetter, "Synthesis of novel views from a single face image," *Int. J. Comput. Vis.*, vol. 28, no. 2, pp. 103–116, 1998.
[26] X. Chai, S. Shan, X. Chen, and W. Gao, "Local linear regression (LLR) for pose invariant face recognition," in *Proc. 7th Int. Conf. Auto. Face and Gesture Recognition*, Southampton, U.K., Apr. 2006, pp. 631–636.
[27] J. M. Lattine, J. D. Carroll, and P. E. Green, *Analyzing Multivariate Data*. New York: Thomson.
[28] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th Int. Conf. Auto. Face and Gesture Recognition*, Washington, DC, 2002, pp. 46–51.
[29] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[30] P. Phillipse, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1103, Oct. 2000.

Xiujuan Chai received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2000 and 2002, respectively, where she is currently purusing the Ph.D. degree.

Her research interests include pattern recognition, image processing, face recognition, and biometrics.



Shiguang Shan (M'04) received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1997 and 1999, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China, in 2004.

He is currently an Associate Researcher and serves as the Vice Director of the Digital Media Center of the Institute of Computing Technology (ICT), CAS. He is also the Vice-Director of the ICT-ISVision Joint R&D Lab for Face Recognition. His research interests cover image analysis, pattern recognition, and computer vision. He is particularly focusing on face recognition-related research topics.

Dr. Shan received the China's State Scientific and Technological Progress Award in 2005.

Xilin Chen (M'00) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1988, 1991, and 1994, respectively. He was a Professor at the Harbin Institute of Technology from 1999 to 2005.

He was a Visiting Scholar at Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. He joined the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in August 2004. His research interests include image processing, pattern recognition, computer vision, and multimodal interface.

Dr. Chen has served as a program committee member for more than 20 international and national conferences. He has received several awards, including the China's State Scientific and Technological Progress Award in 2000, 2003, and 2005 for his research work.

Wen Gao (M'92–SM'05) received the B.Sc. degree in computer science from the Harbin University of Science and Technology, Harbin, China, in 1982, the M.Sc. degree in computer science from the Harbin Institute of Technology, Harbin, in 1985, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He joined the Harbin Institute of Technology in 1985, where he served as a Lecturer, Professor, and Head of the Department of Computer Science until 1995. He was with the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, from 1996 to 2005. During his career as a Professor with CAS, he was also appointed as Director of the Institute of Computing Technology, Executive Vice President of the Graduate School, as well as the Vice President of the University of Science and Technology of China. He is currently a Professor with the School of Electronics Engineering and Computer Science, Peking University, China. He has published four books and over 300 technical articles in refereed journals and proceedings in the areas of multimedia, video compression, face recognition, sign language recognition and synthesis, image retrieval, multimodal interface, and bioinformatics.

Dr. Gao is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA, Editor of the *Journal of Visual Communication and Image Representation*, and Editor-in-Chief of the *Journal of Computer* (in Chinese). He received the China's State Scientific and Technological Progress Awards in 2000, 2002, 2003, and 2005, respectively.