Multiview Face Recognition: From TensorFace to V-TensorFace and K-TensorFace

Chunna Tian, Guoliang Fan, *Senior Member, IEEE*, Xinbo Gao, *Senior Member, IEEE*, and Qi Tian, *Senior Member, IEEE*

***Abstract*—Face images under uncontrolled environments suffer from the changes of multiple factors such as camera view, illumination, expression, etc. Tensor analysis provides a way of analyzing the influence of different factors on facial variation. However, the TensorFace model creates a difficulty in representing the nonlinearity of view subspace. In this paper, to break this limitation, we present a view-manifold-based TensorFace (V-TensorFace), in which the latent view manifold preserves the local distances in the multiview face space. Moreover, a kernelized TensorFace (K-TensorFace) for multiview face recognition is proposed to preserve the structure of the latent manifold in the image space. Both methods provide a generative model that involves a continuous view manifold for unseen view representation. Most importantly, we propose a unified framework to generalize TensorFace, V-TensorFace, and K-TensorFace. Finally, an expectation-maximization like algorithm is developed to estimate the identity and view parameters iteratively for a face image of an unknown/unseen view. The experiment on the PIE database shows the effectiveness of the manifold construction method. Extensive comparison experiments on Weizmann and Oriental Face databases for multiview face recognition demonstrate the superiority of the proposed V- and K-TensorFace methods over the view-based principal component analysis and other state-of-the-art approaches for such purpose.**

***Index Terms*—Manifold learning, multiview face recognition, nonlinear tensor decomposition, subspace analysis, TensorFace.**

Manuscript received October 27, 2010; revised April 29, 2011 and August 12, 2011; accepted August 24, 2011. Date of publication February 3, 2012; date of current version March 16, 2012. This work was supported in part by the National Natural Science Foundation of China under Grants 60832005 and 61172146, by the Ph.D. Programs Foundation of the Ministry of Education of China under Grants 20090203120011 and 20090203110002, by the Key Science and Technology Program of Shaanxi Province of China under Grant 2010K06-12, by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2009JM8004, and by the Basic Science Research Fund in Xidian University under Grant 72105470. The work of G. Fan was supported by the U.S. National Science Foundation (NSF) under Grant IIS-0347613. The work of Q. Tian was supported in part by the NSF Division of Information and Intelligent Systems under Grant 1052851, by the Faculty Research Awards from Google, FXPAL, and NEC Laboratories of America. This work has been presented in part in the International Conference on Pattern Recognition in December 2008, Tampa, FL, USA. This paper was recommended by Associate Editor Z. Li.

C. Tian and X. Gao are with the Video and Image Processing System Laboratory, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: chnatian@xidian.edu.cn; xbgao@mail.xidian.edu.cn).

G. Fan is with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078-5032 USA (e-mail: guoliang.fan@okstate.edu).

Q. Tian is with the Department of Computer Science, University of Texas, San Antonio, TX 78249 USA (e-mail: qitian@cs.utsa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2011.2169452

I. Introduction

FACE recognition is an active research topic in the areas of pattern recognition, computer vision, and machine learning. It is also a challenging topic because real-world face images are formed with the interaction of multiple factors on the imaging conditions, including illumination, head poses, etc. Three-dimensional face recognition that can explicitly accommodate various factors for accurate recognition by involving 3-D face models has recently been studied [1], [2]. In this paper, we focus on multiview face recognition based on 2-D face images. We aim to build a compact data-driven face representation model. Specifically, we study two main issues: 1) how to develop a general multiview face modeling framework that is able to handle multiple unseen views for face recognition and 2) how to estimate the view and identity parameters effectively given a face image with an unknown/unseen view. Correspondingly, our research involves two technical components. First, we want to explore the low-dimensional intrinsic structure, i.e., the view manifold, in the view subspace. It supports a continuous-valued view manifold, based on which we can build two generative multiview face models. Second, we propose an expectation-maximization (EM)-like algorithm to estimate the identity and view of a new face image iteratively. We verify the effectiveness of the proposed algorithms on two multiview face databases, i.e., the Oriental Face [3] and Weizmann [4] databases, and one multi-illumination face database, i.e., the PIE [5] database.

To handle the multifactor face modeling and recognition, an elegant multilinear tensor decomposition based on the TensorFace framework was proposed in [6]. However, TensorFace exhibits a bottleneck which is a limitation in dealing with unseen views due to the discrete nature of the view subspace. In this paper, we want to embed a continuous-valued nonlinear view manifold in the TensorFace framework and name it view-manifold-based TensorFace (V-TensorFace). V-TensorFace is able to deal with unseen views by nonlinear interpolation along the sequential viewpoints in the low-dimensional view subspace. However, the neighborhood relationship in the view manifold may not be well preserved in the high-dimensional face image space in V-TensorFace particularly for the interpolated viewpoints [7]. Thereby, we want to establish a bidirectional mapping between the view manifold and the image space that preserves the local neighborhood structure of views. For this purpose, we employ the kernel trick in tensor decomposition and develop a kernelized TensorFace (K-TensorFace) for multiview face modeling. Both TensorFace and V-TensorFace methods can be generalized from K-TensorFace.

In summary, our main contributions include the following: 1) a proposal for two multiview face modeling methods, i.e., V-TensorFace and K-TensorFace, both of which involve a continuous-valued view manifold to handle unseen views; 2) the casting of TensorFace, V-TensorFace, and K-TensorFace into a unified framework; and 3) the development of an iterative EM-like parameter estimation method for multiview face recognition via V-TensorFace and K-TensorFace. The rest of this paper is organized as follows. Section II briefly reviews the related work. Sections III and IV present TensorFace and V-TensorFace, respectively. In Section V, we propose K-TensorFace with an appropriate view manifold to cope with the nonlinear view variations. The general multiview face representation that unifies TensorFace, V-TensorFace, and K-TensorFace is presented in Section VI, where the EM-like parameter estimation method is introduced. Section VII shows the experimental results on three face databases which involve view and illumination variations. Conclusions and future research lines are discussed in Section VIII.

II. Related Work

A rich literature on face recognition across poses/views exists. The review in [8] summarized two trends of multiview recognition methods: generalizing the pose representation ability of the model or designing pose-invariant mechanisms to tolerate the pose variation. However, only several classical subspace-learning-based methods are analyzed in [8]. Our review focuses on the related work based on subspace learning to improve the pose generalization ability of 2-D multiview face recognition algorithms. We further categorize them into two groups, i.e., the linear and nonlinear models.

A. Linear Methods

Most traditional subspace analysis algorithms, such as the principal component analysis (PCA) [9], linear discriminant analysis (LDA), and their variants [10], [11], assume that data obey the Gaussian distribution. This assumption is applicable to face images with only one factor varied, for example, identity. However, the in-depth rotations of multiview faces always cause a misalignment of the face appearance, which results in the non-Gaussian distribution of face data. Thus, a “divide and conquer” strategy was used in view-based PCA (VPCA). The identity was estimated under each view-specific basis [12]. In [13], a tied factor analysis model was proposed to describe the view variations on face images. It contains two principal components learned iteratively from the training data: identity and mean-pose vectors. To obtain the identity-independent view representation, bilinear analysis was introduced to separate “style” (view) and “content” (identity) factors [14]. The bilinear model provides an efficient way for modeling bifactor interactions. It was used to address the 3-D face and facial expression recognitions jointly in [15]. The *2-mode analysis* was extended to multilinear analysis [16] to deal with the multifactor variational face recognition, which is named as TensorFace-based recognition [6]. Then, PCA in the multilinear model is substituted by independent component analysis [17] to maximize the statistical independence of the representational components. Lee *et al.* proposed a tensor-based active appearance model to represent the nonrigid view changes for recognition [18].

In the aforementioned algorithms, face images are rasterized in vectors. Some researchers aim to further explore the spatial information of images. Yang and Ye *et al.* proposed the 2D-PCA [19] and 2D-LDA [20], respectively, which use 2-D matrices instead of image vectors to construct the basis matrix of subspace analysis. Reference [21] extended 2D-PCA to the horizontal and vertical 2D-PCA-based discriminant analysis methods for face verification, which were less sensitive to face alignment. In [22] and [23], the matrices of face images are considered as second-order tensor data. The generalized eigenvector decomposition is applied to obtain the principal components for face recognition. In [24] and [25], matrix sequences with variations are regarded as high-order tensor data. The variations of the data are separated by tensor unfolding. Finally, LDA is incorporated for classification. The extension of the incremental learning to high-order tensor data is presented in [26]. A generalized tensor dimensionality reduction method is proposed in [27] to capture the redundancies across different modes of multidimensional data. Experimental results have shown the superiority of these tensor-based approaches over the corresponding vector-based ones, particularly for faces with small sample sizes. However, the linear method is bound to ignore the nonlinearity in multiview faces. This is a bottleneck for achieving a high-performance recognition system.

B. Nonlinear Methods

To improve the nonlinearity tolerance to the view, the kernel trick and manifold learning techniques were incorporated into multiview face modeling. The kernel-based methods, e.g., kernel PCA (KPCA) and kernel LDA (KLDA), map face images into the high-dimensional feature space to linearly separate the nonlinear distributions in the image space [28]–[30]. In [31], data were transformed into the feature space and tensorized. Then, N-mode singular value decomposition (SVD) was applied to achieve TensorFace in the feature space. Since the data in the feature space has an invisible structure, the high-dimensional TensorFace creates a difficulty in factorizing the unknown view of a new test image. Thus, Park *et al.* [32] proposed an individual kernel TensorFace that creates a one-to-many mapping from the individual identity space to the observed data space. Dot products are utilized to avoid the *curse of dimensionality*. However, the effect of kernel operations on nonlinear representation of views is unknown until experimental evaluation is executed [8].

In contrast to the kernel method, global and local embedding-based manifold learning methods, such as locally linear embedding (LLE) [33], ISOMAP [34], Laplacian Eigenmap [35]–[37], locality preserving projections [38], and maximization of the geometric mean of all divergences [39], aim to visualize the subtle structure of the data in a low-dimensional space. A unified understanding of these classic locality-preserving-based manifold learning algorithms is provided in [40] and is extended to sparse dimensionality reduction in [41]. When it comes to multifactor face representation, most works focus on building an appearance manifold for each identity and then use the point-to-manifold [42] or manifold-to-manifold [43] distance to evaluate the test data. To represent the view nonlinearity, Lee *et al.* used PCA to faces within each view to represent the local linearity of the view manifold and then connected the basis of each view with a probabilistic transition matrix learned from videos [44]. Raytchev *et al.* proposed an ISOMAP-based view manifold. They filtered each face image with a derivative of Gaussian filter to reduce the influence of identity on the view manifold [45]. We call the manifold derived directly from face images *data-driven view manifold*. The construction of a face model, which involves an independent nonlinear view manifold, is the focus of most current multiview face recognition research.

Most recently, the factorized manifold for observation analysis was proposed by introducing tensor decomposition to manifold construction. In [46], a quadrilinear space was obtained as an appearance manifold by applying tensor decomposition to the geometrical surface normals of the 3-D face model, where the discrete view subspace is represented as one of the linear spaces of the quadrilinear manifold. Lee *et al.* [47] proposed a nonlinear tensor decomposition approach for silhouette-based gait tracking, in which a continuous manifold for the camera view was built. They used the tensor decomposition to obtain discrete coefficients of the camera views and then applied the spline fitting to connect the view coefficients. In this paper, we extend the method in [47] to the view-manifold construction of faces and name it a *hybrid data-concept-driven view manifold*. It is applied to the proposed V- and K-TensorFace models. We examine the effectiveness of *concept-driven*, *data-driven*, and *hybrid data-concept-driven* view manifolds for the proposed K-TensorFace model. An EM-like algorithm is developed to estimate the view and identity iteratively. The experimental results show a significant improvement over the traditional TensorFace [6] and the VPCA [12] methods.

III. Review of TensorFace

View variations cause dramatic differences in face images, which have strong influence on the appearance-based face recognition. Thus, the study on view subspace seems necessary. Tensor decomposition provides a means to analyze the multiple factors of face images independently. The factorized model of multiview faces, which is called TensorFace, separates the view and identity factors with their corresponding subspaces. Interested readers are referred to [16] and [48] for the details on the tensor-decomposition-related multilinear algebra. The notations are explained as follows. Scalars are denoted by lowercase letters $(a, b, \ldots)$, vectors by bold lowercase letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices by bold uppercase letters $(\mathbf{A}, \mathbf{B}, \ldots)$, and higher order tensors by calligraphic uppercase letters $(\mathcal{A}, B, \ldots)$.

The tensor data can be regarded as a multidimensional vector. In multilinear algebra, each dimension of the N-order tensor $\mathcal{D} \in R^{(I_1 \times I_2 \times \ldots \times I_N)}$ is associated with a mode. Tensor decomposition through high-order SVD (HOSVD) [49] seeks for N orthonormal mode matrices. To obtain the TensorFace model of multiview faces, we rasterize face images along pixel, identity, and view variations as tensor $\mathcal{Y}$. Then, we apply HOSVD on it to get

$$\mathcal{Y} = \mathcal{C} \times_1 \mathbf{U}_{\text{pixel}} \times_2 \mathbf{U}_{\text{identity}} \times_3 \mathbf{U}_{\text{view}} \tag{1}$$

where $\times_n$ represents the mode-n product between core tensor $\mathcal{C}$ and mode matrix $\mathbf{U}_n$. The core tensor governs the interaction among mode matrices [6], [49]. $\mathbf{U}_{\text{pixel}}$ orthonormally spans the space of eigenimages. Rows of $\mathbf{U}_{\text{identity}}$ and $\mathbf{U}_{\text{view}}$ span the parameter space of different identities and views, respectively, which separates the influences of identity and view factors.

We extract an identity vector $\mathbf{p}^k$ from $\mathbf{U}_{\text{identity}}$ and a view vector $\mathbf{x}_i$ from $\mathbf{U}_{\text{view}}$ to represent person k and view i, respectively. TensorFace enables us to represent the training image $\mathbf{y}$ of the kth identity under the ith view based on $\mathcal{C} \times_1 \mathbf{U}_{\text{pixel}}$ as follows:

$$\mathbf{y}_i^k = \mathcal{C} \times_1 \mathbf{U}_{\text{pixel}} \times_2 \mathbf{p}^k \times_3 \mathbf{x}_i. \tag{2}$$

Given a test facial image $\mathbf{y}$, face recognition is used to identify the closest identity vectors in $\mathbf{U}_{\text{identity}}$. However, without the view information, we cannot calculate the identity vector of $\mathbf{y}$ based on (2). TensorFace assumes that the view vector of $\mathbf{y}$ is the ith row of $\mathbf{U}_{\text{view}}$, $i \in \{1, \ldots, V\}$. Thus, the test view is regarded as one of the training views. The identity vector $\mathbf{p}_i$ under view i can be obtained as

$$\mathbf{p}_i = \mathbf{y}\mathbf{B}^{-1} \tag{3}$$

where $\mathbf{B}$ is the flattened matrix of subtensor $(\mathcal{C} \times_1 \mathbf{U}_{\text{pixel}} \times_3 \mathbf{x}_i)$ along the identity dimension. Then, face recognition is used to find the nearest neighbor of $\mathbf{p}_i$ in the identity subspace $\mathbf{U}_{\text{identity}}$, i.e.,

$$k = \arg\min_{k,i} \|\mathbf{p}_i - \mathbf{p}^k\| \tag{4}$$

where $i \in \{1, \ldots, V\}$ and $k \in \{1, \ldots, K\}$. $\|\cdot\|$ denotes the Euclidean distance between two vectors.

TensorFace-based multiview face recognition traverses the discrete identity and view subspaces obtained from the training data to find a best match combination of identity and view vectors. The discrete nature of the view subspace in TensorFace lacks the power of representing the view structure. Thus, it may not handle the unseen views which are significantly diverging from the training ones. Moreover, the effective solution of the parameters instead of using an exhaustive search is another issue of TensorFace-based recognition.

IV. V-TensorFace

The key point of V-TensorFace is to build a view manifold which can represent the view nonlinearity and preserve the structure of the view subspace. Thus, V-TensorFace will lead to a better representation of multiview faces than TensorFace. There are two problems to be solved: 1) how to build such a view manifold and ensure its validity at the same time and 2) how to incorporate the manifold into the TensorFace model.

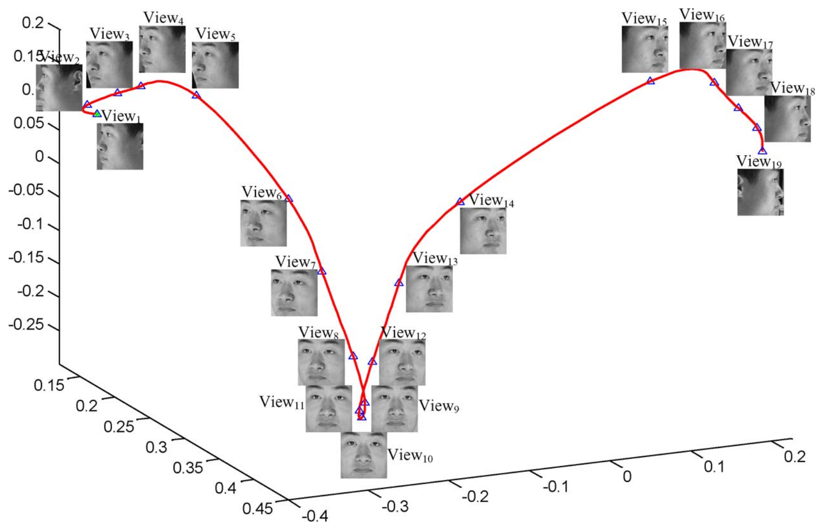


Fig. 1. Hybrid view manifold where only the first three dimensions are shown. The blue triangles denote the view coefficients learned by tensor decomposition. The red line represents the interpolated views in the continuous view manifold.

A. View-Manifold Generation and Verification

The view subspace in TensorFace is composed of the discrete view coefficients. However, we expect to build a continuous view manifold based on these low-dimensional discrete view coefficients, which preserves the neighboring structure of views in the high-dimensional space. In [47], a view-manifold generation method is proposed, which connects the view coefficients after applying Bicubic spline fitting to the discrete coefficients of the training views in the view subspace.

In our case, Fig. 1 shows the view manifold derived from the Oriental Face database. The generation details are given as follows. First, we apply HOSVD on the tensorized multiview face data as (1). The rows of $\mathbf{U}_{\mathrm{view}}$ are referred to as a finite set of view coefficients that will be used for spanning the view manifold, which are denoted by the blue triangles in Fig. 1. Since the sequential order of the views should be kept in the low-dimensional view subspace, we sort the view coefficients according to the order of the panning angles of face images to constitute the global structure of the view manifold. The local topology of the view manifold is generated by spline fitting method, which is demonstrated by the red line in Fig. 1. In this paper, we use $\mathcal{M} \subset R^e$ to denote the view manifold, where e is the dimensionality of the view subspace. $\mathcal{M}(\mathbf{x})$ represents the viewpoint on $\mathcal{M}$ projected from $\mathbf{x}$ in the space R^e.

The assumption behind view-manifold topology construction is to preserve the intrinsic view structure of multiview face images. We verify this assumption as follows. Suppose the order of view coefficients is unknown, we hypothesize that we can still find the optimal order to connect all view coefficients by finding the shortest path that travels through them [50] because the shortest path can preserve the local smoothness of the view manifold that is essential to support valid view interpolation for face recognition under an unknown view.

We regard the view coefficients $\mathbf{x}$ as the points in a 1-D nonlinear path in the view subspace. Let $\mathbf{x}_t^l$ and $\mathbf{x}_t^{l+1}$ denote the adjacent l_{th} and $(l+1)_{\mathrm{th}}$ points in the t_{th} path, respectively. We define the distance $S_t^l(\mathbf{x}_t^l, \mathbf{x}_t^{l+1})$ as the Euclidean distance between $\mathbf{x}_t^l$ and $\mathbf{x}_t^{l+1}$ that is computed by ignoring the last few dimensions to remove the orthogonality among all view coefficients $\mathbf{x}$. Then, the shortest path P_v of V view coefficients is determined by

$$P_v = \min_t \sum_{l=1}^{V-1} S_t^l\left(\mathbf{x}_t^l, \mathbf{x}_t^{l+1}\right), \tag{5}$$

where V is the number of views. The solution to (5) is obtained by using the genetic algorithm due to its NP-complete nature [51]. The experimental results show that the order of the shortest path is exactly same as the intrinsic order of all given views. This confirms that the view manifold preserves the local smoothness and the correct global structure to support meaningful view interpolation.

B. View-Manifold-Based TensorFace Modeling

To represent the view nonlinearity, we embed the view manifold in TensorFace to build V-TensorFace. In this model, face image $\mathbf{y}$ of the kth identity under the ith view can be generated by (6)

$$\mathbf{y}_i^k = \mathcal{C} \times_1 \mathbf{U}_{pixel} \times_2 \mathbf{p}^k \times_3 \mathcal{M}(\mathbf{x}_i). \tag{6}$$

Here, $\mathbf{p}^k$ is the identity coefficient of the kth subject. $\mathcal{M}(\mathbf{x}_i)$ is the ith view coefficient on the view manifold $\mathcal{M}$. The major difference between (2) and (6) is the mode-3 multiplier. In (2), the mode-3 multiplier is a view coefficient selected from the discrete mode matrix $\mathbf{U}_{\mathrm{view}}$ (the blue triangles in Fig. 1). In (6), the mode-3 multiplier is a view coefficient sampled from the continuous view manifold $\mathcal{M}$ (the red line in Fig. 1). This manifold helps us to generate the face images of $\mathbf{p}^k$ under unseen views and locate the view of the face image under unseen views on the view manifold. In this model, tensor decomposition is used once, and the low-dimensional view nonlinearity and continuity are represented by the view manifold. We develop an iterative parameter searching method for an unknown image based on this manifold in Section VI.

V. K-TensorFace

The purpose of building a multiview face model is to find the latent structure of multiview faces then build a homomorphism mapping between the latent variable space and the image space. Most nonlinear dimensionality reduction approaches focus on preserving the local distance of the high-dimensional data in the latent space. For example, the view manifold in V-TensorFace preserves the local view adjacency relationship. However, this model does not guarantee that the local distances in the latent view subspace will be preserved in the image space, particularly for the interpolated unseen views. Since kernel mapping can provide a smooth interpolation from the low- to high-dimensional space, we introduce it to V-TensorFace model to build a smooth bidirectional mapping, resulting in the K-TensorFace model. In K-TensorFace, the manifold generation mechanism guarantees the locality preservation in the latent space, while the locality property of the adopted Gaussian kernel ensures the manifold structure preservation in the image space.

A. Kernelized TensorFace Modeling

We briefly discuss nonlinear tensor decomposition [47] in the context of multiview face modeling. This method can provide a compact low-dimensional representation of multiview face images by exploring both the linear (*e.g.*, the identity) and nonlinear (e.g., the view) variable spaces. We refer the readers to [47] for more details.

Given the multiview face images $\mathbf{y}_{1:N}^{1:K}$ of K persons, each person has N views, which are represented by $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ in the nonlinear view manifold. We use kernel $\phi(\cdot)$ of the generalized radial basis function (RBF) to build the nonlinear mapping between the view manifold and the high-dimensional image space as

$$y^j(\mathbf{x}) = \sum_{i=1}^{n} \omega_i^j \phi\left(\|\mathbf{x} - \mathbf{z}_i\|\right) + l(\mathbf{x}) b^j, \qquad (7)$$

where y^j denotes the jth element in $\mathbf{y}$. $\mathbf{z}$ are the kernel centers sampled from the view manifold, and ω_i is the weight of each kernel. b^j is the mapping coefficient of linear polynomial $l(\mathbf{x}) = [1, \mathbf{x}]$. The mapping between the multiview face images of the kth person and their low-dimensional view coefficients can be represented by

$$\begin{pmatrix} \psi(\mathbf{x}_1) \\ \psi(\mathbf{x}_2) \\ \vdots \\ \psi(\mathbf{x}_N) \\ \mathbf{A} \end{pmatrix} \mathbf{D}^k = \begin{pmatrix} \mathbf{y}_1^k \\ \mathbf{y}_2^k \\ \vdots \\ \mathbf{y}_N^k \\ \mathbf{O}_{(e+1)\times d} \end{pmatrix} \qquad (8)$$

where $\mathbf{A} = [l(\mathbf{x}_1)^\top, l(\mathbf{x}_2)^\top, \ldots, l(\mathbf{x}_N)^\top, O_{(e+1)\times(e+1)}]$. The multiplication between $\mathbf{A}$ and $\mathbf{D}^k$ ensures their orthogonality and makes the solution of $\mathbf{D}^k$ well posed. $\psi(\mathbf{x}) = [\phi(\|\mathbf{x} - \mathbf{z}_1\|), \ldots, \phi(\|\mathbf{x} - \mathbf{z}_n\|),\ 1, \mathbf{x}]$ is a vector of dimension $1 \times (N + e + 1)$. $\mathbf{D}^k$ is a $(N + e + 1) \times d$ matrix. Its jth column is $[\omega_1^j, \omega_2^j, \ldots, \omega_N^j, b^j]$.

For the kth person under view i, the mapping can be represented by

$$\mathbf{y}_i^k = \psi(\mathbf{x}_i)\mathbf{D}^k. \qquad (9)$$

The linear part $l(\mathbf{x})$ of $\psi(\mathbf{x})$ supports the conditionally positive property of the RBF mapping [52], [53]. It also provides a means to solve each $\mathbf{x}$ without solving the inverse of the nonlinear part of $\psi(\mathbf{x})$. In (8) and (9), $\mathbf{y}$ is identity dependent. However, $\psi(\mathbf{x})$ is identity free. We deduce that the identity information is embedded in the linear matrix $\mathbf{D}$ in this mapping. To extract the identity information, we stack $\mathbf{D}^{1:K}$ to form a tensor $\mathcal{D} = [\mathbf{D}^1, \ldots, \mathbf{D}^K]$. Then, we apply HOSVD to $\mathcal{D}$ to abstract the low-dimensional identity coefficients $\mathbf{p}^k \in R^K$. This results in the generative multiview face model K-TensorFace. In this model, $\mathbf{y}_i^k \in R^d$ can be synthesized by identity $\mathbf{p}^k$ and view $\mathbf{x}_i$ coefficients as

$$\mathbf{y}_i^k = \mathcal{C} \times_1 \mathbf{U}_{pixel} \times_2 \mathbf{p}^k \times_3 \psi(\mathbf{x}_i) \qquad (10)$$

where $\mathcal{C}$ is a 3-order core tensor.

The process of model learning and image synthesis is shown in Fig. 2. In the learning phase, the multiview face images $\mathbf{y}^k$ are mapped to their low-dimensional viewpoints $\mathbf{x}$ in the manifold $\mathcal{M}$ according to the RBF mapping in (9). The localized Gaussian kernels in RBF mapping attempt to preserve the latent structure of view manifold in the high-dimensional image space even for the interpolated viewpoints. Then, the linear matrix $\mathbf{D}$ of each person is stacked to form a tensor $\mathcal{D}$. We decompose $\mathcal{D}$ with HOSVD to obtain the identity coefficients $\mathbf{p}^k$ and the core tensor $\mathcal{C}$. Finally, the multiview face images are synthesized by (10) in the synthesis part.

B. Discussion on the Choice of View Manifolds

The RBF mapping can build a nonlinear mapping between face images and different kinds of view manifolds. The performance of K-TensorFace will be significantly affected by the involved view manifold. In order to cope with the view nonlinearity in the K-TensorFace model, we briefly discuss three kinds of view manifold generation methods: data-, concept-, and hybrid data-concept-driven view manifolds.

View manifolds can be deduced from the training data using nonlinear dimensionality reduction methods like LLE [54] or ISOMAP [45]. However, those view manifolds are person dependent and cannot be used for multiview face modeling that requires a commonly shared view manifold. The concept-driven manifolds originate from an ideal conceptual design. For example, gait observations of a full walking cycle under a fixed view can be embedded on a 2-D circular-shaped manifold [54]. Moreover, gait observations from multiple views can be embedded on a 3-D torus [55]. These conceptual manifolds were shown effectively on exploring the intrinsic low-dimensional structure in the high-dimensional data in [54] and [55]. In our case, multiview face images rotating from the leftmost to the rightmost are obtained by placing cameras with equal angle spacing along a semicircular bracket. Hence, conceptually, we design a view manifold on a semicircle (see Fig. 3). In V-TensorFace, the view coefficients in $\mathbf{U}_{\text{view}}$ construct the

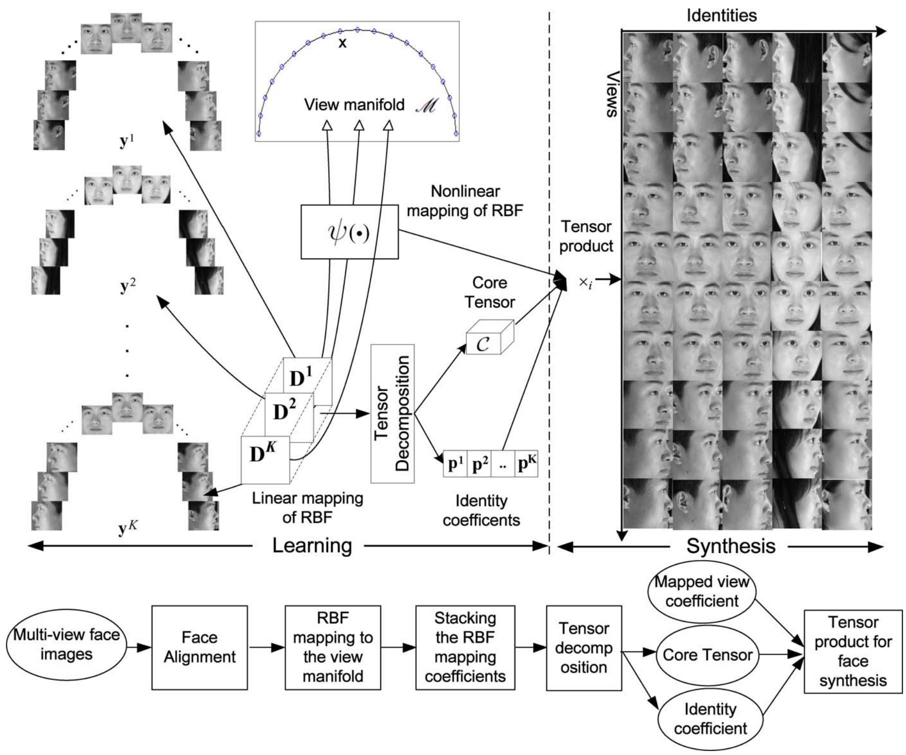


Fig. 2. Learning and synthesis for the nonlinear tensor-based multiview face representation.

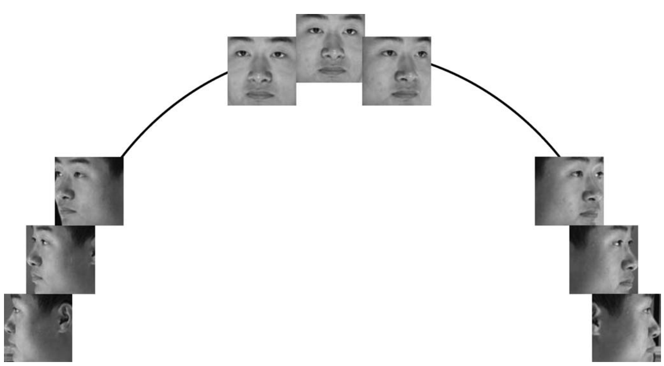


Fig. 3. Demonstration of the conceptual view manifold.

global nonlinear structure of the view manifold (data driven). Moreover, the view coefficients interpolated by spline fitting construct the locally linear structure of the view manifold (concept driven), which use the prior knowledge of manifold smoothness. It is a hybrid data-concept-driven manifold generation method (see Fig. 1). The comparison of these three manifold generation methods is given in Table I.

A basic requirement of learning K-TensorFace is that the view manifold must be shared by all persons, which means the view manifold is invariant to identity changes. This makes identity-dependent data-driven view manifolds unusable here. The conceptual view manifold uses the prior knowledge of view rotating in the physical space. Although it is purely determined by the conceptual design, which is independent to the identity changes, it may not necessarily capture the intrinsic nonlinear view structure of the multiview face images. Therefore, the hybrid view manifold seems to be a better choice since it has a proper balance between the generality and the specificity.

C. More Discussion on K-TensorFace

The training steps of K-TensorFace involving the hybrid view manifold are given in Table II. In this algorithm, HOSVD will be used twice. First, it is used to extract the view coefficients, based on which the hybrid view manifold is learned via spline fitting. Then, this view manifold is used for nonlinear tensor decomposition where HOSVD is used again to extract the identity coefficients. However, in [54] and [55], HOSVD is used once where only a data-driven pose manifold is involved. Most importantly, K-TensorFace is a general framework for multiview face modeling that can be degenerated to TensorFace or V-TensorFace with appropriate kernel selection. Regarding the parameter estimation, a similar EM-like algorithm [54], [55] is involved in our work. In particular, we made two improvements that further improve the face recognition accuracy. One is a new three-step initialization scheme that facilitates the convergence of the EM-like algorithm. The other is the discussion of two domains for face recognition, the image domain (ID), and the coefficient domain (CD). In [54] and [55] and TensorFace [49], the recognition task was done in CD only. We will compare and analyze the two domains for face recognition.

VI. General Framework for Multiview Face Representation

We cast the TensorFace-, V-TensorFace-, and K-TensorFace-based multiview face models in a general framework to help readers understand the relationship among them. Then, we give an iterative parameter searching method for face recognition.

TABLE I
COMPARISON OF VIEW-MANIFOLD GENERATION METHODS

Methods	Generated from	Pros	Cons
Data-driven manifold	Training data	Reflect the low-dimensional view changes of real data	Person-dependent The views in the low-dimensional space could be mixed for different persons
Concept-driven manifold	Conceptual design	Identity independent Intuitive with prior knowledge	Not necessarily capture the intrinsic nonlinear structures of the real data
Hybrid data-concept-driven manifold	Training data and conceptual design	Reflect view changes of the real data Identity independent	The accuracy of the interpolated views in the hybrid manifold depends on the dense of the training views

TABLE II
TRAINING PROCEDURE OF K-TENSORFACE INVOLVING THE HYBRID VIEW MANIFOLD

Input: Multi-view face images $\mathbf{y}_{1:N}^{1:K}$ of K persons under N views

Step$_1$: Rasterize $\mathbf{y}_{1:N}^{1:K}$ along the variations of pixel, identity and view as a tensor data $\mathcal{Y}$.

Step$_2$: Apply HOSVD to $\mathcal{Y}$ as Eq.(1) to extract the view matrix $\mathbf{U}_{view}$.

Step$_3$: Sort the view coefficients in $\mathbf{U}_{view}$ according to the given intrinsic view order.

Step$_4$: Span the hybrid view manifold via spline fitting according to the sorted view coefficients in Step$_3$.

Step$_5$: For $k = 1, 2, \ldots, K$, build the mapping between $\mathbf{y}_{1:N}^{k}$ and their corresponding low dimensional view coefficients $\mathbf{x}_{1:N}$ by Eq.(8).

Step$_6$: Arrange the linear mapping matrix $\mathbf{D}$ of K persons as a tensor $\mathcal{D}$, and apply HOSVD on it to obtain the generative model in Eq.(10).

Output: The generative K-TensorFace model in Eq.(10).

A. General Multiview Face Representation

There are two main differences among TensorFace, V-TensorFace, and K-TensorFace. One is the view representation, and the other is the mapping between the view subspace and image space. In the following, we want to revisit TensorFace and V-TensorFace from the K-TensorFace point of view to examine the aforementioned two main differences (see Table III).

To get TensorFace, we define $\psi(\mathbf{x}) = \mathbf{x}$ in (10), where $\mathbf{x} \in \mathbf{U}_{\text{view}}$. The nonlinear view variation cannot be accurately represented in TensorFace because of the sparse view representation. For this purpose, V-TensorFace is proposed, for which we define $\psi(\mathbf{x}) = \mathcal{M}(\mathbf{x})$ in (10). The interpolated views in the manifold $\mathcal{M}$ may not be preserved very well in the image space. Thus, the Gaussian kernel $\phi(\cdot)$ was incorporated to build a smooth bidirectional mapping between the view manifold and the image space. It results in the K-TensorFace model where $\psi(\mathbf{x}) = [\phi(\|\mathbf{x} - \mathbf{z}_1\|), \ldots, \phi(\|\mathbf{x} - \mathbf{z}_N\|),\ 1, \mathbf{x}]$ and $\mathbf{x} \in \mathcal{M}$. Both V-TensorFace and K-TensorFace involve a view manifold, with which we develop an iterative EM-like parameter searching method. V-TensorFace can be regarded as the linearization of K-TensorFace. When the view manifold in V-TensorFace is substituted with the discrete view coefficients, V-TensorFace degenerates into TensorFace. In this perspective, we call K-TensorFace a unified multiview TensorFace model.

B. Parameter Estimation

In the previous section, we have described the generation of a multiview face with the V- and K-TensorFace models. It is also necessary to model the inverse process. In other words, given a view manifold $\mathcal{M} \subset R^e$ and a new input $\mathbf{y} \in R^d$, we need to estimate its view parameter $\mathbf{x} \in \mathcal{M}$ and identity parameter $\mathbf{p}^s$ by minimizing the reconstruction error. Unfortunately, the solution of each parameter depends on the existing of another one. The EM-like algorithm is suited to this type of "chicken-and-egg" problem. Usually, the face variations caused by illumination and viewing direction may be larger than the interperson variation in distinguishing identities [8], [56]. As Fig. 4 shows, the distance between images with different identities under the same view (D_i) may be smaller than that between images with the same identity but different views (D_v). Therefore, identity coefficients are more sensitive than view coefficients to new observations, and thus, we start with view initialization.

Initialization: It contains three steps, as shown in Fig. 5. For the illustration purposes, conceptual manifold is used in Fig. 5. It can be the hybrid data-concept-driven manifold.

a) *View estimation under each identity*. Given a test image $\mathbf{y}$, we assume its identity coefficients are $\mathbf{p}^k$, $k = 1, \ldots, K$, of the training data. Then, K view coefficients $\mathbf{x}_k^0 \in R^e$ with respect to each identity are solved by taking the linear part of $\psi(\cdot)$ in (11), in which $\mathbf{D}^k$ is the flattened matrix of subtensor $\mathcal{C} \times_1 \mathbf{U}_{pixel} \times_2 \mathbf{p}^k$

$$\psi\left(\mathbf{x}_k^0\right) = \mathbf{D}^{k^{-1}}\mathbf{y}. \tag{11}$$

b) *View coefficient synthesis*. For each $\mathbf{y}$, we synthesize its view coefficient by $\mathbf{x}' = \sum_k p(\mathbf{x}_k^0)\mathbf{x}_k^0$, where $p(\mathbf{x}_k^0)$ denotes the probabilities of $\mathbf{x}_k^0$ belonging to $\mathcal{M}$. It can be determined under the Gaussian assumption as

$$p\left(\mathbf{x}_k^0\right) \propto \exp\left\{-\left\|\mathbf{x}_k^0 - \mathcal{M}\left(\mathbf{x}_k^0\right)\right\|^2/(2\sigma^2)\right\}, \tag{12}$$

where $\sum_k p(\mathbf{x}_k^0) = 1$ and σ is a preset variance controlling the algorithm sensitivity. $\mathcal{M}(\mathbf{x}_k^0)$ is the viewpoint on the manifold projected from $\mathbf{x}_k^0$.

c) *View projection to the manifold*. The view coefficient is initialized by projecting $\mathbf{x}'$ to the manifold as $\mathbf{x} = \mathcal{M}(\mathbf{x}')$.

Identity Estimation: Given the estimated view coefficient $\mathbf{x}$, we calculate $\psi(\mathbf{x}) = [\phi(\|\mathbf{x} - \mathbf{z}_1\|), \ldots, \phi(\|\mathbf{x} - \mathbf{z}_N\|),\ 1, \mathbf{x}]$. The identity vector $\mathbf{p}^s$ of $\mathbf{y}$ in (10) is solved as follows:

$$\mathbf{p}^s = \mathbf{X}^{-1}\mathbf{y}, \tag{13}$$

TABLE III
COMPARISON OF MULTIVIEW FACE MODELS

Model type	Mathematical function of $\psi(\cdot)$	View representation	Mapping type	Parameter searching
TensorFace	$\mathbf{x}$	View coefficients in $\mathbf{U}_{view}$	Linear	Exhausting search
V-TensorFace	$\mathscr{M}(\mathbf{x})$	View manifold $\mathscr{M}$	Linear	EM-like iteration
K-TensorFace	$[\phi(\|\mathbf{x}-\mathbf{z}_1\|),\ldots,\phi(\|\mathbf{x}-\mathbf{z}_N\|),\ 1,\mathbf{x}]$	View manifold $\mathscr{M}$	Nonlinear	EM-like iteration

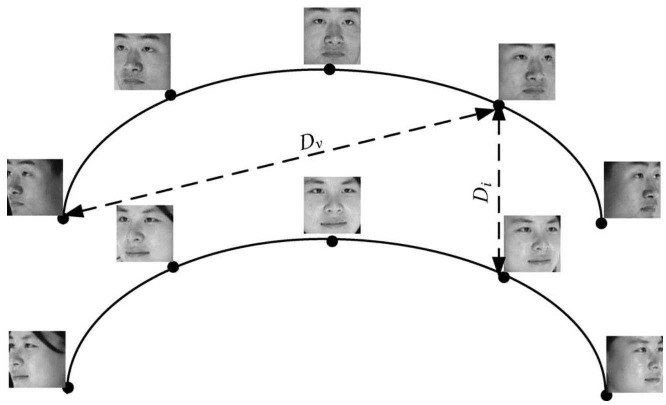


Fig. 4. Distance comparison between the identity and view variations. D_v is the distance between two images of the same identity with different views. D_i is the distance between two images of the same view with different identities.

where $\mathbf{X}$ is the flattened matrix of subtensor $\mathcal{C} \times_1 \mathbf{U}_{\text{pixel}} \times_3 \psi(\mathbf{x})$. Then, identity recognition can be performed in the CD in terms of $\mathbf{p}^k$ as

$$k_{\text{CD}} = \arg\min_k \|\mathbf{p}^s - \mathbf{p}^k\|, \tag{14}$$

or we consider $\mathbf{y}$ as drawn from a Gaussian mixture model centered at the reconstructed identity class k. Therefore, the likelihood function of observation $p(\mathbf{y}|k,\mathbf{x})$ belonging to person k can be formulated as

$$p(\mathbf{y}|k,\mathbf{x}) \propto \exp \left\{-\left\|\mathbf{y} - \mathcal{C} \times_1 \mathbf{U}_{\text{pixel}} \times_2 \mathbf{p}^k \times_3 \psi(\mathbf{x})\right\|^2 / (2\sigma^2)\right\}. \tag{15}$$

With the equal probability assumption of $p(k)$ and $p(k|\mathbf{x})$, the posterior probability of identity k is reduced to

$$p(k|\mathbf{x},\mathbf{y}) = p(\mathbf{y}|k,\mathbf{x}) / \sum\nolimits_k p(\mathbf{y}|k,\mathbf{x}). \tag{16}$$

Then, identity recognition can be performed in the reconstructed ID as

$$k_{\text{ID}} = \arg\max_k p(k|\mathbf{x},\mathbf{y}). \tag{17}$$

We will test both schemes with experiments. A new identity vector for further view estimation can be obtained as $\mathbf{p}^s = \sum_k p(k|\mathbf{x},\mathbf{y})\mathbf{p}^k$.

View Estimation: Given the test image $\mathbf{y}$ and the estimated $\mathbf{p}^s$, we can solve a new view coefficient $\mathbf{x}' \in R^e$ in (10) based on (11). Then, the updated view coefficient that is constrained on the view manifold is obtained by $\mathbf{x} = \mathcal{M}(\mathbf{x}')$. Therefore, the identity and view coefficients can be solved iteratively until the termination condition is met. The iteration between identity and view estimation is shown in Fig. 6.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we briefly introduce the experimental databases first, i.e., Oriental Face [3], Weizmann [4], and PIE [5]. To validate the proposed multiview face model, we report the results obtained on multiview faces in the Weizmann and Oriental Face databases. The view and identity recognition results of TensorFace and V-TensorFace on both databases are given to show the effectiveness of the hybrid data-concept-driven view manifold. To distinguish the influence of different view manifolds on K-TensorFace, we recognize the identity of Oriental Face database in both ID and CD with the concept-driven view manifold [concept manifold (CM)] and hybrid data-concept-driven view manifold (HM). Finally, we extend the proposed method to multi-illumination face recognition. The manifold generation method is verified on the illumination (order unknown) manifold construction on the PIE database.

A. Databases

In the Oriental Face database, we use 1406 face images of 74 oriental people, and each one has 19 views in the range of $[-90°,\ldots,90°]$ for pan sampled at around the $10°$ interval along a semicircle. Some examples are given in Fig. 7(a). Weizmann face database contains 28 individuals, and each one has five views, as shown in Fig. 7(b). In the PIE database, the frontal face images of 60 persons obtained by Camera 27 are chosen. Each person has 21 illuminations [see Fig. 7(c)] caused by 21 flash firing variation with the room lights off. Since most subspace-learning-based face recognition methods are sensitive to spatial misalignments [57], all images are shifted to align the eyes and nose tip position found manually. The automatic alignment can be realized by reconstructing from the misaligned images [58]. This method is applicable in our method to achieve automatic face alignment. The images in the Weizmann database are in the size of 64×50 pixels since the faces in this database are a bit longer than others. The rest of the images are in the size of 50×50 pixels.

B. Experiments on Weizmann Database

There is a small range of view variation in the Weizmann database. We use the leave-one-out-style cross-validation to recognize face images under unfamiliar views. The algorithm performance is compared with VPCA, TensorFace, the proposed V-TensorFace, and K-TensorFace. Since it is hard to build a concept view manifold for only five views, we only test K-TensorFace with the hybrid manifold (HM). Tables IV and V show the recognition results of view and identity, respectively.

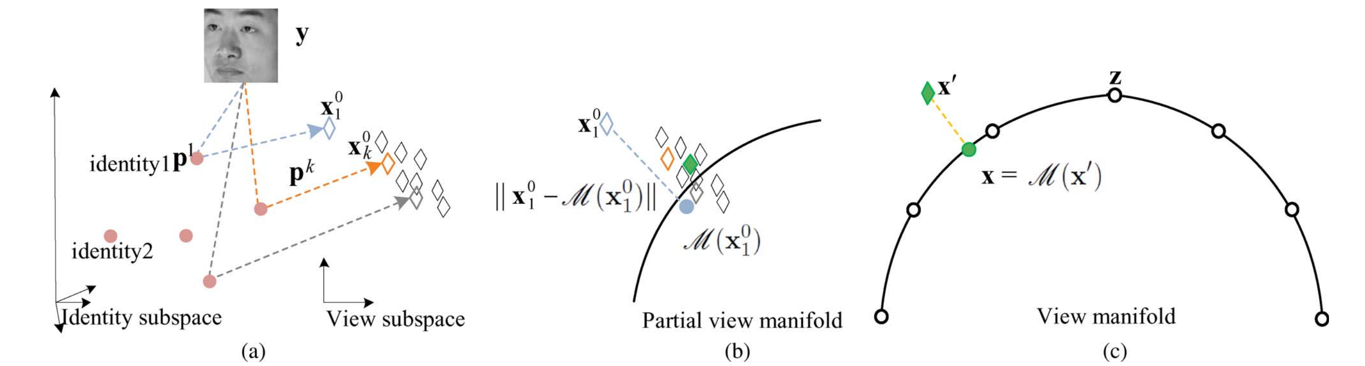


Fig. 5. Demonstration of initialization procedure. (a) View estimation under each identity. We assume that the identity vector of $\mathbf{y}$ is $\mathbf{p}^k$, $k = 1, \ldots, K$. (b) View coefficient synthesis. We only show part of the view manifold. (c) View projection to the manifold. The view manifold can be the conceptual view manifold or the hybrid one.

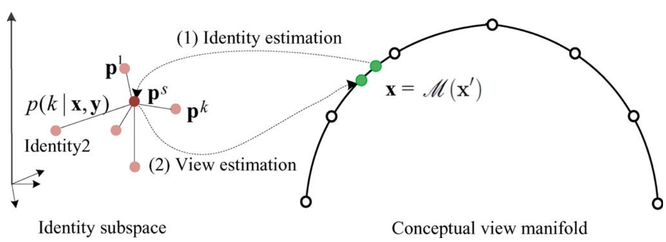


Fig. 6. Demonstration of the iteration between identity and view estimation.

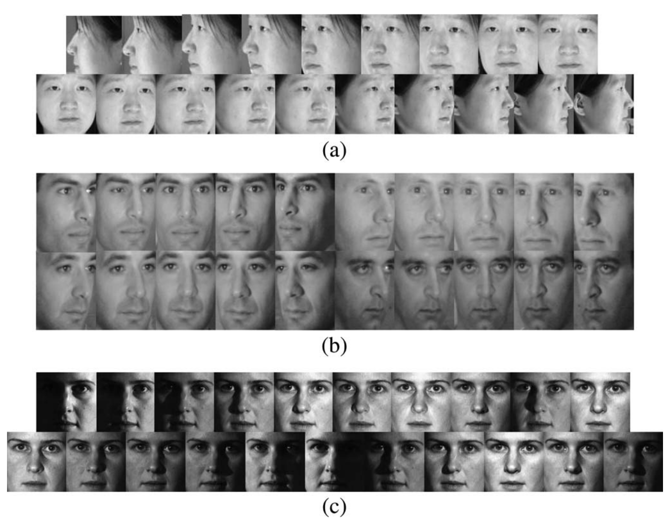


Fig. 7. Face examples in the multiview face databases. (a) Examples of each individual under 19 views in the Oriental Face database. (b) Examples of each individual under five views in the Weizmann database. (c) Multi-illumination face examples in the PIE database. Each individual under 21 illuminations in the PIE database.

Both VPCA and TensorFace are linear methods. If the test view is unfamiliar to the training ones, it is possible that the recognition rate of TensorFace is not higher than that of VPCA. However, tensor representation has the power of separating different influential factors. The by-product of TensorFace-based identity recognition is view recognition. However, determining an unfamiliar test view to a training view is not reasonable when the main features of the test view are away from the training ones. Therefore, it is necessary to build a view manifold, with which we can represent the view nonlinearity

TABLE IV
COMPARISON OF VIEW RECOGNITION ON THE WEIZMANN DATABASE (IN PERCENT)

Methods TestData	VPCA	TensorFace	V-TensorFace	K-TensorFace
View$_2$	100.00	100.00	100.00	100.00
View$_3$	96.43	92.86	100.00	100.00
View$_4$	100.00	82.14	100.00	100.00
Mean	98.81	91.67	100.00	100.00
Std.	2.06	8.99	0.00	0.00

TABLE V
COMPARISON OF IDENTITY RECOGNITION ON THE WEIZMANN DATABASE (IN PERCENT)

Methods	VPCA	Tensor	V-TensorFace		K-TensorFace		
TestData		Face	/CD	/ID	/CD	/ID	/Itera.
View$_2$	46.43	50.00	53.57	75.00	53.57	75.00	75.00
View$_3$	35.71	17.86	39.29	71.43	39.30	71.43	71.43
View$_4$	25.00	21.43	25.00	53.57	42.86	50.00	67.86
Mean	35.71	29.76	39.29	66.67	45.24	65.48	71.43
Std.	10.71	17.61	14.28	11.48	7.42	13.52	3.57

more accurately. In our parameter estimation strategy, view is recognized along the manifold. Only when the test view lies *between* its neighboring views we count it as a positive result. In TensorFace-based recognition, if the test view is close to its neighboring views, we count it as a positive result. The view recognition rates in Table IV show that the view manifold works better, which proves that the manifold results in a better multiview face representation.

In Table V, the face identification rates of V-TensorFace are about 10% and 37% higher than that of TensorFace in CD and ID, respectively. The recognition rate of K-TensorFace with iterative parameter estimation is 4.76% higher than that of TensorFace in ID. In both TensorFace and K-TensorFace, the recognition in ID is higher than that in CD. It is because the way to obtain $\mathbf{p}_s$ in CD needs a pseudoinverse operator while the recognition in ID avoids solving the matrix inverse. Basically, TensorFace overcomes the basis disunity and mode inseparability of VPCA. V-TensorFace improves TensorFace by representing the nonlinearity in the view subspace.

TABLE VI
COMPARISON OF RECOGNITION RATES OF FOUR METHODS ON ORIENTAL FACE DATABASE (IN PERCENT)

Methods	VPCA	Tensor	V-TensorFace			K-TensorFace				
TestData		Face	/CD	/ID	/Itera.	CM/CD	CM/ID	HM/CD	HM/ID	HM/Itera.
View_2	60.81	62.10	68.92	63.51	81.15	48.65	39.19	75.68	74.32	77.03
View_3	67.57	74.32	75.68	77.03	91.89	60.81	58.11	74.32	81.08	93.24
View_4	71.62	68.92	56.76	56.76	79.73	86.49	87.84	63.51	67.57	81.08
View_5	48.65	36.49	40.54	51.35	56.76	24.32	25.68	36.49	51.35	56.76
View_6	40.54	35.14	41.89	36.49	41.89	74.32	67.57	40.54	35.14	41.89
View_7	47.30	35.14	41.89	68.92	71.62	24.32	39.19	41.89	68.92	68.92
View_8	39.19	47.30	51.35	52.70	70.27	62.16	40.54	50.00	60.81	71.62
View_9	59.46	67.57	71.62	68.92	67.57	62.16	74.32	62.16	68.92	93.24
View_{10}	66.26	75.68	79.73	79.73	72.97	68.92	79.73	79.73	85.14	79.73
View_{11}	64.87	67.57	72.97	77.03	72.97	64.86	75.68	77.03	77.03	77.03
View_{12}	47.30	35.14	50.00	66.22	68.92	37.84	71.62	45.95	67.57	68.92
View_{13}	27.03	36.49	50.00	66.22	68.92	33.78	35.14	48.65	66.22	67.57
View_{14}	37.84	22.97	31.08	51.35	55.41	13.51	16.22	32.43	52.70	54.05
View_{15}	51.35	62.16	62.16	74.32	67.57	44.59	36.49	62.16	72.97	68.92
View_{16}	74.32	78.38	77.03	77.03	87.84	55.41	55.41	87.84	87.84	89.19
View_{17}	47.30	54.05	82.43	59.46	72.97	60.81	51.35	83.78	63.51	81.08
View_{18}	56.76	57.75	74.32	63.51	56.76	45.95	39.19	60.81	47.30	56.76
Mean-17	53.42	53.95	60.49	64.15	69.71	51.11	52.54	60.17	66.38	72.18
Mean-13	56.13	58.49	65.59	67.46	74.12	54.78	57.49	65.49	70.79	77.34
Std-13	13.79	15.67	13.61	8.47	9.17	16.48	18.40	15.41	10.82	10.70

K-TensorFace further enhances V-TensorFace by involving a nonlinear mapping between the view manifold and the face images. In this paper, the recognition rates of View_1 and View_5 are absent, because they are the starting point and endpoint of the view manifold, respectively. They were used to interpolate the inner viewpoints. Without them, we cannot extrapolate their related information. However, in this case, views can be determined by the hard division as VPCA and TensorFace do.

C. Experiments on the Oriental Face Database

The Oriental Face database contains a long range of view variations. On this database, experiments are compared among VPCA, TensorFace, V-TensorFace, and K-TensorFace. In V-TensorFace, the identity parameter is solved in both CD and ID without iteration and iteratively. In K-TensorFace, when the cases of CM and HM are combined with recognition in CD or ID, we have four experiments to test. They are K-TensorFace with CM/CD, CM/ID, HM/CD, and HM/ID, respectively, in which the parameter estimation is noniterative. In addition, we test K-TensorFace with iterative parameter estimation and HM (K-TensorFace HM/Itera.). The results are shown in Table VI.

In Table VI, the recognition rates of V-TensorFace are 6.44%–8.31% higher than those of TensorFace. It indicates that the manifold works better than the discrete view coefficients. Comparing the results in Tables V and VI, we can find that the identity recognition rate of TensorFace on the Oriental Face database is improved around 30% than that on the Weizmann database, which has slightly bigger view intervals. However, the proposed V- and K-TensorFace models are more robust than TensorFace to view sampling distance.

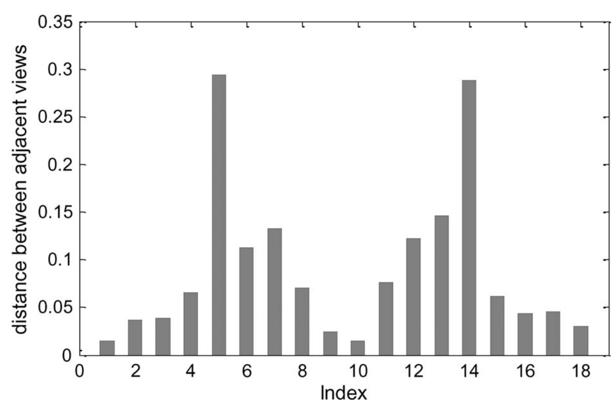


Fig. 8. Distance between two neighboring views in the view subspace.

Even TensorFace and VPCA are comparable with regards to the identification rate. TensorFace represents multiview face images in a unified basis, which provides a general parametric face model with explicit physical meaning of different factors. Although the cameras are distributed at about 10° interval along a semicircle during the imaging of the oriental faces, these images do not obey a uniform distribution in the view subspace. We calculate the distance between the neighboring views in the view subspace of TensorFace using only the first three principal components of each view coefficient. Fig. 8 shows the result. The occlusions and changes of different neighboring view pairs are quite different. It is obvious that intervals $\text{View}_5-\text{View}_6$ and $\text{View}_{14}-\text{View}_{15}$ are too sparse [also see Figs. 1 and 7(a)] to support valid view interpolation. Thus, the recognition results in these intervals are usually lower than those in other intervals. Note that, in Table VI, *Mean-17* is the mean of the recognition rates of $\text{View}_2-\text{View}_{18}$. For further comparison, we ignore the

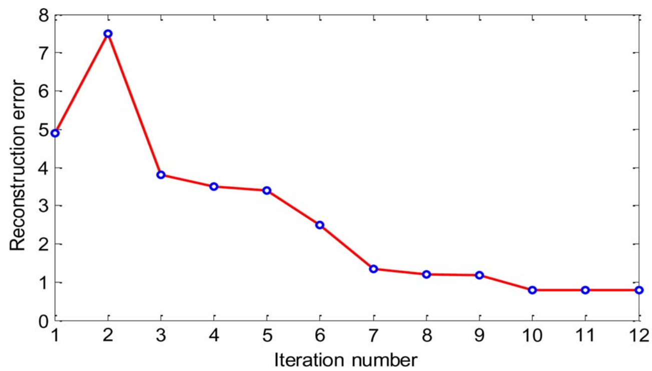


Fig. 9. Example of the converge curve of the proposed parameter estimation method.

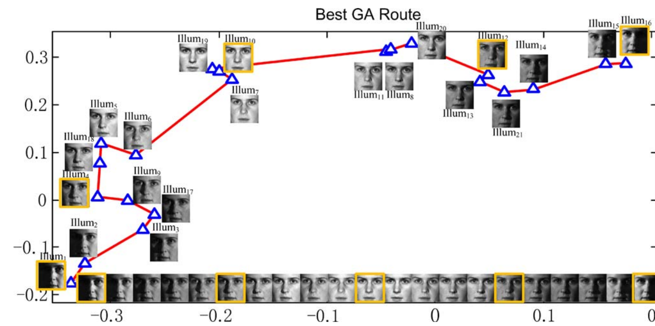


Fig. 10. Illumination order obtained by the GA shortest path in the illumination subspace on the PIE database. The faces connected by the red line are demonstrated at the bottom of the figure. The images in yellow squares are used for face synthesis and recognition.

intervals $\text{View}_5 - \text{View}_6$ and $\text{View}_{14} - \text{View}_{15}$ and use *Mean-13* to reflect the effectiveness of the view manifold.

In Table VI, K-TensorFace with CM performs worse than TensorFace in both CD and ID, which shows that the concept-driven view manifold is not effective in capturing the intrinsic structure of the view subspace. The obtained multiview images do not obey a uniform distribution in the view subspace, which is particularly obvious in the intervals $\text{View}_5 - \text{View}_6$ and $\text{View}_{14} - \text{View}_{15}$. The low recognition rates of HM in these intervals show that the performance of K-TensorFace depends on the distribution density of the manifold. Even the recognition rates are very different across views under the same method, the recognition rates of all the testing algorithms have similar variation tendencies across different views. If we can get more densely sampled multiview faces, the view manifold would be smoother, which will result in a better model description. In another perspective, this manifold generation method is useful for directing the camera distribution for the uniformly distributed multiview face capture.

The recognition rates of K-TensorFace with HM are improved by 6.75%–9.36% and 12.96%–14.66% than VPCA in CD and ID, respectively. Compared with TensorFace, the recognition rates of K-TensorFace with HM are improved around 7% and 12% in CD and ID, respectively. The iterative ID-based K-TensorFace method is 18.23%–18.85% better than TensorFace in identification. As seen from our intensive experiments, the iterative parameter estimation algorithm converges in 6–10 steps. An example of the converge curve is shown in Fig. 9. The convergence rate depends on the initialization. The V- and K-TensorFace models have almost the same recognition rates in CD. However, the recognition rate of K-TensorFace is around 3% higher than that of V-TensorFace in ID whether with or without the iterative parameter estimation. Since the recognition is executed in the reconstructed ID, we attribute this improvement to the smooth kernel mapping in K-TensorFace.

D. Validation of the GA Shortest Path-Based Manifold Generation Method

View manifold generation is used to preserve the sequential order of multiple views in the image space. The view order is easily observed. Thus, we test the GA shortest path-based manifold generation method on the data without knowing the view order. In Fig. 7(c), the illumination order of the 21 faces are unknown. Thus, we extend the manifold generation method to explore the manifold topology of illumination to validate its effectiveness. To obtain the identity-independent illumination coefficients, we apply HOSVD on the PIE data to get the mode matrix of illumination, which contains the coefficient vectors of 21 illuminations. We illustrate the first two dimensions of the illumination coefficients by the blue triangles in Fig. 10 and connect them with the shortest path. The faces connected by the red line are demonstrated in the lower part of the figure, which shows that the connected illumination coefficients are sorted in a reasonable illumination-varying order.

E. Synthesis and Recognition on the Illumination Data of PIE

Usually, when the illumination variation is smooth, the linear face recognition method achieves very good results. However, face recognition on sparsely sampled illuminations is a challenging task because of the nonlinear illumination changes. To handle this nonlinearity, we use K-TensorFace to build a multi-illumination face model. We adopt only five illuminations, namely, Illum_1, Illum_4, Illum_{10}, Illum_{12}, and Illum_{16} as the training data, which are marked with yellow squares in Fig. 10. The identity recognition is compared among VPCA, TensorFace, V-TensorFace, and K-TensorFace. In the last two methods, parameters are solved iteratively. The identity in each illumination is recognized by the leave-one-out-style cross-validation on those five illuminations. The results are given in Table VII. We can see that K-TensorFace with iteration has better results than TensorFace. The identity recognition rate is improved by around 13%, which means that K-TensorFace has better ability to handle the illumination nonlinearity.

To show the smooth projection from the illumination manifold to the image space in K-TensorFace, we synthesize new faces using K-TensorFace model under the interpolated illuminations in the manifold. In this experiment, the illumination manifold is derived from only five illuminations, i.e., Illum_1, Illum_4, Illum_{10}, Illum_{12}, and Illum_{16}, which are marked with

TABLE VII
IDENTITY RECOGNITION RESULTS ON MULTI-ILLUMINATION FACES (IN PERCENT)

Methods TestData	VPCA	TensorFace	V-TensorFace	K-TensorFace
$Illum_4$	80.00	56.67	73.33	71.67
$Illum_{10}$	60.00	71.67	76.66	90.00
$Illum_{12}$	78.33	88.33	60.00	93.33
Mean	72.78	72.22	69.99	85.00
Std.	11.10	15.84	8.82	11.66

Fig. 11. Multi-illumination faces generated along the illumination manifold. The images in yellow squares are the original faces in the PIE database, and the rest of the faces are synthesized along the illumination manifold.

yellow squares in Fig. 11. The remaining 36 images in Fig. 11 are the synthesized ones. We can see that the synthesized images look natural and realistic. Some of them are very close to the real images in the PIE database. The experiments in this section also show that the proposed general model can be extended to multifactor face recognition task.

VIII. CONCLUSION AND FUTURE WORK

In the multiview face recognition tasks, normally, linear methods cannot locate the unfamiliar intermediate views accurately. To represent the nonlinearity in view subspace, we embedded a novel view manifold to TensorFace and validate the effectiveness of the manifold with the shortest path method. To further explore the relationship between multiview faces and the view manifold, we extend the nonlinear tensor decomposition method to V-TensorFace for a compact generative face model building and discussed the effect of different view manifolds on this model. More importantly, we cast three kinds of TensorFace methods into a general framework, and this general method can be extended to the multifactor face representation. For face recognition purposes, a heuristic parameter searching method is developed. The experimental results show great promise for the new methods, wherein a hybrid view manifold is used. One interesting future research problem is to learn a 2-D view manifold that can accommodate both panning and tilting effects of multiview face images.

ACKNOWLEDGMENT

Portions of the research in this paper use the Oriental Face database collected under the research of the Artificial Intelligence and Robotics.

REFERENCES

[1] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
[2] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. J. Maybank, "Bayesian tensor approach for 3-D face modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1397–1410, Oct. 2008.
[3] [Online]. Available: http://www.aiar.xjtu.edu.cn/groups/face/English/HomePage.htm
[4] [Online]. Available: ftp://ftp.idc.ac.il/pub/users/cs/yael/Facebase/
[5] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database of human faces," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-02, 2001.
[6] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," in *Proc. Int. Conf. Pattern Recog.*, Quebec City, QC, Canada, 2002, vol. 3, pp. 511–514.
[7] X. Gao and C. Tian, "Multi-view face recognition based on tensor subspace analysis and view manifold modeling," *Neurocomputing*, vol. 72, no. 16–18, pp. 3742–3750, Oct. 2009.
[8] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognit.*, vol. 42, no. 11, pp. 2876–2896, Nov. 2009.
[9] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. Comput. Vis. Pattern Recog.*, Maui, HI, 1991, pp. 586–591.
[10] P. Belhumeur, J. Hespanha, and D. 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.
[11] X. Wang and X. Tang, "Random sampling LDA for face recognition," in *Proc. Comput. Vis. Pattern Recog.*, Washington, DC, 2004, vol. 2, pp. 259–265.
[12] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. Comput. Vis. Pattern Recog.*, Seattle, WA, 1994, pp. 84–91.
[13] S. J. D. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 970–984, Jun. 2008.
[14] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural Comput.*, vol. 12, no. 6, pp. 1247–1283, Jun. 2000.
[15] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "Bilinear models for 3-D face and facial expression recognition," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 3, pp. 498–511, Sep. 2008.
[16] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Multilinear singular value tensor decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
[17] M. Vasilescu and D. Terzopoulos, "Multilinear independent components analysis," in *Proc. Comput. Vis. Pattern Recog.*, San Diego, CA, 2005, pp. 547–553.
[18] H. Lee and D. Kim, "Tensor-based AAM with continuous variation estimation: Application to variation-robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1102–1116, Jun. 2009.
[19] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
[20] J. Ye, J. Ravi, and L. Qi, "Two-dimensional linear discriminant analysis," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, vol. 17, pp. 1569–1576.
[21] J. Yang and C. Liu, "Horizontal and vertical 2DPCA-based discriminant analysis for face verification on a large-scale database," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 4, pp. 781–792, Dec. 2007.
[22] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2005, vol. 18, pp. 1–8.
[23] D. Tao, X. Li, X. Wu, W. Hu, and S. J. Maybank, "Supervised tensor learning," *Knowl. Inf. Syst.*, vol. 13, no. 1, pp. 1–42, Sep. 2007.
[24] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
[25] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[26] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," *ACM Trans. Knowl. Discov. Data*, vol. 2, no. 11, pp. 1–37, 2008.

[27] H. Wang and N. Ahuja, "A tensor approximation approach to dimensionality reduction," *Int. J. Comput. Vis.*, vol. 76, no. 3, pp. 217–229, Mar. 2008.

[28] M. Yang, "Kernel Eigenfaces vs. kernel Fisherfaces: Face recognition using kernel methods," in *Proc. Conf. Autom. Face Gesture Recog.*, Washington, DC, 2002, pp. 215–220.

[29] Q. Liu, H. Lu, and S. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.

[30] S. Zhou and R. Chellappa, "From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 917–929, Jun. 2006.

[31] Y. Li, Y. Du, and X. Lin, "Kernel-based multifactor analysis for image synthesis and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Beijing, China, 2005, vol. 1, pp. 114–119.

[32] S. W. Park and M. Savvides, "Individual kernel tensor-subspaces for robust face recognition: A computationally efficient tensor framework without requiring mode factorization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1156–1166, Oct. 2007.

[33] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[34] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[35] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[36] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[37] S. Si, D. Tao, and K. Chan, "Evolutionary cross-domain discriminative Hessian eigenmaps," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1075–1086, Apr. 2010.

[38] X. He, S. Yan, Y. Hu, and H. Zhang, "Learning a locality preserving subspace for visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nice, France, 2003, vol. 1, pp. 385–392.

[39] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[40] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[41] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Mining Knowl. Discov.*, vol. 22, no. 3, pp. 340–371, May 2011.

[42] J. Zhang, S. Z. Li, and J. Wang, "Manifold learning and applications in recognition," *Intell. Multimedia Process. With Soft Comput.*, vol. 168, pp. 281–300, 2005.

[43] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *Proc. Comput. Vis. Pattern Recog.*, Anchorage, AK, 2008, pp. 1–8.

[44] K. C. Lee and D. Kriegman, "Online learning of probabilistic appearance manifolds for video-based recognition and tracking," in *Proc. Comput. Vis. Pattern Recog.*, Washington, DC, 2005, vol. 1, pp. 852–859.

[45] B. Raytchev, I. Yoda, and K. Sakaue, "Head pose estimation by nonlinear manifold learning," in *Proc. Int. Conf. Pattern Recog.*, Cambridge, U.K., 2004, vol. 4, pp. 462–466.

[46] Y. Xu and A. K. Roy-Chowdhury, "Learning a geometry integrated image appearance manifold from a small training set," in *Proc. Comput. Vis. Pattern Recog.*, Anchorage, AK, 2008, pp. 1–8.

[47] C. S. Lee and A. Elgammal, "Modeling view and posture manifolds for tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, 2007, pp. 1–8.

[48] K. G. Kolda, "Orthogonal tensor decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.

[49] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, vol. 1, pp. 447–460.

[50] X. Zhang and G. Fan, "Dual gait generative models for human motion estimation from a single camera," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1034–1049, Aug. 2010.

[51] [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/13680-traveling-salesman-problem-genetic-algorithm

[52] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.

[53] F. Girosi and T. Poggio, "Networks and the best approximation property," *Biol. Cybern.*, vol. 63, no. 3, pp. 169–176, Jul. 1990.

[54] A. Elgammal and C. S. Lee, "Separating style and content on a nonlinear manifold," in *Proc. Comput. Vis. Pattern Recog.*, Washington, DC, 2004, vol. 1, pp. 478–485.

[55] C. S. Lee and A. Elgmmal, "Simultaneous inference of view and body pose using torus manifolds," in *Proc. Int. Conf. Pattern Recog.*, Hong Kong, 2006, vol. 3, pp. 489–494.

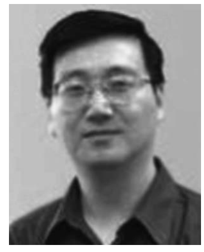
[56] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.

[57] W. Gao, B. Cao, S. Shan, and P. Yang, "Curse of misalignment in face recognition: Problem and a novel misalignment learning solution," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 314–320.

[58] S. Yan, H. Wang, J. Liu, and X. Tang, "Misalignment-robust face recognition," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1087–1096, Apr. 2010.

Chunna Tian received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 2002, 2005, and 2008, respectively.

From 2006 to 2007, she was with the Visual Computing and Image Processing Laboratory, Oklahoma State University, Stillwater, as a Visiting Student. Since 2008, she has been with the School of Electronic Engineering, Xidian University, where she is currently an Associate Professor of Signal and Information Processing. Her research interests include image processing, machine learning, and pattern recognition.



Guoliang Fan (M'01–SM'05) received the B.S. degree in automation engineering from Xi'an University of Technology, Xi'an, China, in 1993, the M.S. degree in computer engineering from Xidian University, Xi'an, in 1996, and the Ph.D. degree in electrical engineering from the University of Delaware, Newark, in 2001.

From 1996 to 1998, he was a Graduate Assistant with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been an Assistant and then Associate Professor with the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater. His research interests include image processing, computer vision, machine learning, and multimedia. He is an Associate Editor of the *EURASIP Journal on Image and Video Processing* and *ISRN Machine Vision*.

Dr. Fan is a recipient of the 2004 National Science Foundation (NSF) CAREER Award, the Halliburton Excellent Young Teacher Award in 2004, the Halliburton Outstanding Young Faculty Award in 2006 from the College of Engineering, OSU, and the Outstanding Professor Award from IEEE-OSU in 2008 and 2011.

Xinbo Gao (M'02–SM'07) received the B.S., M.S., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University, where he is currently a Professor of Pattern Recognition and Intelligent Systems and the Director of the Video and Image Processing System Laboratory. His research interests are computational intelligence, machine learning, computer vision, pattern recognition, and wireless communications. In these areas, he has published 4 books and around 100 technical articles in refereed journals and proceedings including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *Pattern Recognition*, etc. He is also on the editorial boards of journals including the *EURASIP Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier)

Dr. Gao is a Fellow of The Institution of Engineering and Technology. He served as General Chair/Co-chair or Program Committee Chair/Co-chair or member for around 30 major international conferences.

Qi Tian (M'96–SM'03) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1992, the M.S. degree in electrical and computer engineering from Drexel University, Philadelphia, PA, in 1996, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Urbana, in 2002.

He is currently an Associate Professor with the Department of Computer Science, University of Texas, San Antonio (UTSA). His research interests include multimedia information retrieval and computer vision. He has published over 120 refereed journal and conference papers. His research projects were funded by the National Science Foundation, Army Research Office, Department of Homeland Security, San Antonio Life Sciences Institute, Center for Infrastructure Assurance and Security, and UTSA. He took a one-year faculty leave and joined Microsoft Research Asia during 2008–2009. He is a Guest Editor of the *Computer Vision and Image Understanding*, *Pattern Recognition Letters*, *EURASIP Journal on Advances in Signal Processing*, *Journal of Visual Communication and Image Representation*, and is on the editorial boards of the *Journal of Multimedia* and the *Machine Visions and Applications*.

Dr. Tian is the recipient of faculty research awards from Google, NEC Laboratories of America, FX Palo Alto Laboratory, Inc., Akiira Media Systems, and Hewlett–Packard Laboratories. He was the Coauthor of a Best Student Paper in the IEEE International Conference on Acoustics, Speech, and Signal Processing 2006 and of a Best Paper Candidate in the Pacific Rim Conference on Multimedia 2007. He is also the recipient of the 2010 ACM Service Award. He has been serving as program chair, organization committee member, and technical program chair for numerous IEEE and ACM conferences including ACM Multimedia, Special Interest Group on Information Retrieval, International Conference on Computer Vision, International Conference on Multimedia and Expo, etc. He is a Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and is on the editorial board of the IEEE TRANSACTIONS ON CIRCUIT AND SYSTEMS FOR VIDEO TECHNOLOGY.