

Multi-view Face Recognition by Nonlinear Tensor Decomposition

Chunna Tian^{1,2} and Guoliang Fan²

²*School of Electrical and Computer Engineering
Oklahoma State University, USA
chnatian@lab202.xidian.edu.cn*

Xinbo Gao¹

¹*School of Electronic Engineering
Xidian University, Xi'an China
chnatian@lab202.xidian.edu.cn*

Abstract

We discuss a new multi-view face recognition method that extends a recently proposed nonlinear tensor decomposition technique. We use this technique to provide a generative face model that can deal with both the linearity and nonlinearity in multi-view face images. Particularly, we study the effectiveness of three kinds of view manifold for multi-view face representation, i.e., the concept-driven, data-driven and hybrid data-concept-driven view manifolds. An EM-like algorithm is developed to estimate the identity and view factors iteratively. The new face generative model can successfully recognize face images captured under unseen views, and the experimental results provide the new method is superior to the traditional TensorFace-based algorithm and the view-based PCA method.

1. Introduction

Face recognition is emerging as an active research topic in the areas of pattern recognition and computer vision research. It is a challenging topic because natural face images are formed by the interaction of multiple factors related to the imaging condition, scene structure and human status, which broaden the variety of facial geometries, illumination, expressions and head poses *etc* [11]. In this work, we are interested in finding a compact, data-driven and generalizable face representation model for multi-view face recognition. Moreover, the model parameters can be interpreted explicitly in terms of physical variables. It was shown that the 3D morphable model [1] is capable of describing the 3D characteristics of human faces and results in promising face recognition results. In the case when the 3D face model is not available and only face images of multiple views are given, 2D appearance/view modeling is needed, which can be broadly divided into two categories, i.e. the linear models and the nonlinear ones.

In early works, principal component analysis (PCA) was applied to each view. Then the test image was decomposed on the view-base basis matrices (View-based PCA, i.e., VPCA) to locate its view and identity [7]. Bilinear analysis was introduced in [9] to represent views and identities under a uniform basis and separate “style” (view or pose) and “content” (identity) factors. The 2-mode analysis was extended to multilinear analysis [3] to deal with multi-factor variational face recognition (TensorFace) [11]. Then the PCA in the multilinear model is substituted by Independent Component Analysis (ICA) [10] to maximize the statistical independence of the representational components. In general, the linear methods have limited capability to cope with the nonlinear nature within the multi-view face images.

To deal with nonlinear pose/view variations, kernel learning and manifold learning techniques were introduced to multi-view face modeling. The kernel-based methods map the data from the original feature space to a high dimensional space [12] to capture the higher-order statistics of face appearances. Lee *et. al.* used PCA to describe the local linearity of faces within each view and introduced a probabilistic transition matrix to represent a nonlinear view manifold [6]. Raytchev *et. al.* proposed an Isomap-based view manifold generation for face images [8]. Lee *et. al.* [4] proposed a nonlinear tensor decomposition approach for silhouette-based gait tracking that combines both manifold learning and multi-linear tensor analysis to accommodate both multi-factor and nonlinearity in observations.

In this work, we significantly extend the method proposed in [4] for multi-view face recognition. More importantly, we will examine the effectiveness of concept-driven, data-driven and the hybrid data-concept-driven view manifolds for nonlinear tensor decomposition. An EM-like algorithm is developed to estimate the view and identity iteratively. The experimental results show a significant improvement over the traditional multilinear tensor-based methods, such as TensorFace in [11] and the VPCA method in [7].

2. The Proposed Algorithm

Three issues are addressed in this section: (1) How to develop a compact and general multi-view face representation? (2) How to obtain an accurate view manifold for multi-view face representation? (3) How to estimate the identity and view iteratively?

2.1 Nonlinear Tensor Decomposition

We briefly discuss nonlinear tensor decomposition proposed in [4] in the context of multi-view face modeling. This method can provide a compact low-dimensional representation of multi-view face images by exploring both the linear (e.g., the identity) and nonlinear (e.g., the view) structures in the high-dimensional data. We refer the readers to [4] for more details.

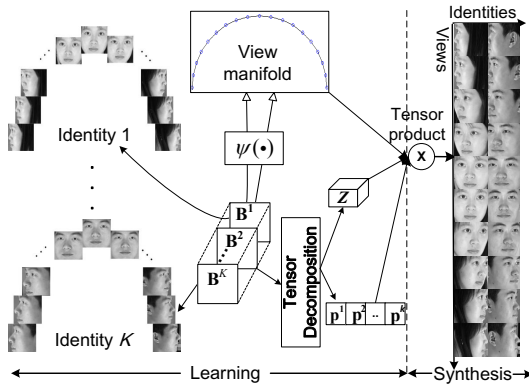


Figure 1. The learning and synthesis for the nonlinear tensor-based multi-view face representation.

Given a nonlinear view manifold, we can obtain N training view coefficients $\mathbf{x}_{1:N}$. We use $\mathbf{y}_{1:N}^{1:K}$ to represent the set of face images of K persons and N views. Then we can use the generalized Radial Basis Function (GRBF) to build a mapping relationship between $\mathbf{x}_{1:N}$ and $\mathbf{y}_{1:N}^{1:K}$ via a nonlinear kernel mapping $\psi(\mathbf{x}) = [\phi(\|\mathbf{x} - \mathbf{z}_1\|), \dots, \phi(\|\mathbf{x} - \mathbf{z}_N\|), 1, \mathbf{x}]$ and linear mapping matrices \mathbf{B}^k , where $\phi(\cdot)$ are Gaussian kernels and $\mathbf{z}_{1:N}$ are the kernel centers along the manifold. Then we can stack $\mathbf{B}^{1:K}$ to form $\mathbf{C} = [\mathbf{B}^1, \dots, \mathbf{B}^N]$. We can apply High Order Singular Value Decomposition (HOSVD) to tensorized \mathbf{C} and to abstract the low-dimensional identity coefficients $\mathbf{p}^k \in R^K$. This gives a multi-view face generative model that can synthesize $\mathbf{y}_i^k \in \mathbb{R}^d$ given an identity coefficient \mathbf{p}^k and view coefficient \mathbf{x}_i defined on the view manifold as

$$\mathbf{y}_i^k = \mathcal{Z} \times_2 \mathbf{p}^k \times_3 \psi(\mathbf{x}_i), \quad (1)$$

where \mathcal{Z} is a 3-order core tensor which governs the interaction between the view and identity. The process of model learning and image synthesis is shown in Fig. 2.

2.2 View Manifold Generation

We briefly discuss three kinds of view manifold generation for nonlinear tensor decomposition.

2.2.1 Concept-driven and Data-driven

The concept-driven manifolds ordinate from a conceptual design (Fig. 2(a)), while the data-driven ones are deduced from real data. For example, gait observations of a full walking cycle under a fixed view can be embedded on a 2-D circular-shaped manifold [2]. Moreover, gait observations from multiple views can be embedded on a 3-D torus [5]. These conceptual manifolds were shown effectively on exploring the intrinsic low-dimensional structure in the high-dimensional data. On the other hand, view manifolds can also be obtained from the training data by using the nonlinear dimensionality reduction methods like Locally Linear Embedding (LLE) [2] or Isomap [8]. However, those view manifolds are person-dependent, and cannot be used for multi-view face modeling that requires a commonly shared view manifold.

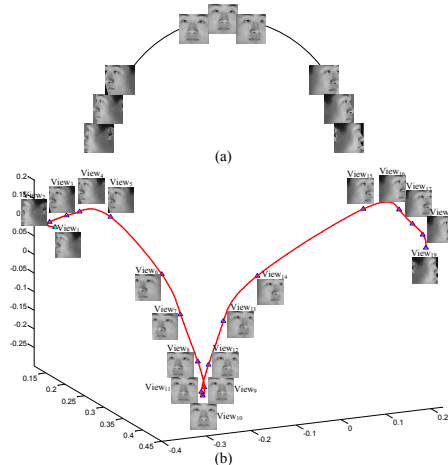


Figure 2. Two kinds of view manifolds. (a) The conceptual view manifold. (b) The hybrid view manifold where only the first three dimensions are shown.

2.2.2 Hybrid Data-concept-driven View Manifold

In [4], an interesting view manifold was generated by using a spline fitting method that connects the view coefficients according to their intrinsic order in the view coefficient space obtained by the tensor decomposition, which is illustrated as Fig. 2(b). We name it a *hybrid data-concept-driven manifold*, whose generation process involves both the training data and a conceptual design. One may ask what about the case when the view

order is not available. We did the following experiment. If we regard the learnt view coefficients as the vertexes of a graph, the shortest path linking all vertexes leads to a view manifold with the correct intrinsic structure (i.e. the view order). This finding supports the possibility of unsupervised learning of the view manifold.

2.2.3 Discussion on the Choice of View Manifolds

The performance of nonlinear tensor decomposition discussed in Section 2.1 will be significantly affected by the view manifold involved for finding the nonlinear mapping between the view coefficients ($\mathbf{x}_{1:N}$) and multi-view face images ($\mathbf{y}_{1:N}^{1:K}$). The basic requirement is that the view manifold must be shared by all persons, i.e., person-independent. This makes data-driven manifolds not usable. Although the conceptual manifold meets this requirement, but it may not fully capture the intrinsic nonlinear structures about the view in the high-dimensional data due to the pure conceptual design.

Therefore, the hybrid view manifold seems to be a better choice that has a proper balance between generality and specificity. Essentially, when the hybrid view manifold is involved, HOSVD will be used twice. First, it is used to abstract the view vectors based on which the hybrid view manifold is formed via spline fitting. Then this view manifold is used for nonlinear tensor decomposition where HOSVD is used again to abstract the identity coefficients as discussed in Section 2.1.

2.3 Model Parameter Estimation

Given a view manifold $\mathcal{V} \subset \mathbb{R}^e$, we use $\mathcal{V}(\mathbf{x}_0)$ to denote the view point on \mathcal{V} projected from $\mathbf{x}_0 \in \mathbb{R}^e$. Given a new input $\mathbf{y} \in \mathbb{R}^d$, we need to estimate its view $\mathbf{x} \in \mathcal{V}$ and identity parameter \mathbf{p}^s by minimizing certain error function. We propose an EM-like algorithm to solve this problem iteratively. Since identity coefficients are more sensitive than view coefficients to new observations, we start with view initialization.

Initialization: Given a test image \mathbf{y} , we can obtain K (the number of persons) view coefficients $\mathbf{x}_k^0 \in \mathbb{R}^e$ ($k = 1, \dots, K$) with respect to each identity coefficient \mathbf{p}^k by solving the linear part of $\psi(\cdot)$ in Eq. (1). For each \mathbf{y} , we can synthesize its view coefficient by $\mathbf{v}' = \sum_k p_{\mathcal{V}}(\mathbf{x}_k^0) \mathbf{x}_k^0$ where $p_{\mathcal{V}}(\mathbf{x}_k^0)$ are the probability of \mathbf{x}_k^0 belonging to \mathcal{V} and can be determined under the Gaussian assumption as:

$$p_{\mathcal{V}}(\mathbf{x}_k^0) \propto \exp\{-\|\mathbf{x}_k^0 - \mathcal{V}(\mathbf{x}_k^0)\|^2 / (2\sigma^2)\}, \quad (2)$$

where $\sum_k p_{\mathcal{V}}(\mathbf{x}_k^0) = 1$ and σ^2 is a pre-set variance controlling the algorithm sensitivity. Then the view coefficient is initialized by $\mathbf{x} = \mathcal{V}(\mathbf{v}')$.

Identity Estimation: Given \mathbf{y} and the estimated view coefficient \mathbf{x} , we can solve identity vector \mathbf{p}^s based on Eq. (1). Then identity recognition can be performed in coefficient domain (CD) in terms of \mathbf{p}^k as

$$k_{CD} = \arg_k \min \|\mathbf{p}^s - \mathbf{p}^k\|. \quad (3)$$

Or we can consider \mathbf{y} as drawn from a Gaussian mixture model centered at $\mathcal{Z} \times_2 \mathbf{p}^k \times_3 \psi(\mathbf{x})$ for each identity class k . Therefore, the likelihood function of observation $p(\mathbf{y}|k, \mathbf{x})$ belong to person k can be formulated as $p(\mathbf{y}|k, \mathbf{x}) \propto \exp\{-\|\mathbf{y} - \mathcal{Z} \times_2 \mathbf{p}^k \times_3 \psi(\mathbf{x})\|^2 / (2\sigma^2)\}$. (4)

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

$$p(k|\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|k, \mathbf{x}) / \sum_k p(\mathbf{y}|k, \mathbf{x}). \quad (5)$$

Then identity recognition can be performed in the reconstructed image domain (ID) as

$$k_{ID} = \arg_k \max p(k|\mathbf{x}, \mathbf{y}). \quad (6)$$

The major difference between (3) and (6) is that ID-based recognition involves the core tensor \mathcal{Z} that is not used by the CD-based one. We will test both schemes in the experiment. 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 based on Eq. (1) as $\mathbf{v}' \in \mathbb{R}^e$. Then the updated view coefficient that is constrained on the view manifold can be obtained by $\mathbf{x} = \mathcal{V}(\mathbf{v}')$. Therefore, the identity and view coefficients can be solved iteratively till the termination condition is met.

3 Experimental Results and Analysis

Our experiment were performed on the Oriental Face database, which has 1406 face images of 74 individuals under 19 viewpoints in the range of $[-90^\circ, \dots, 90^\circ]$ with 10° interval. All images were manually aligned. The leave-one-out cross-validation is used to recognize a face image of an unseen view instead of using multiple test images [11]. Two algorithms were compared here, TensorFace in [11] and VPCA in [7]. Basically, TensorFace overcomes the basis disunity and mode inseparability of VPCA, and our method further enhances TensorFace by involving a nonlinear view manifold.

We have five implementations for the generative model-based multi-view face recognition, depending on the view manifold (the concept manifold or the hybrid manifold, namely CGM or HGM) used, the domain used for recognition (the coefficient domain and the image domain, namely CD or ID), and the method used for recognition (non-iterative or iterative). In summary, we have 7 groups of experiments including that of Tensor-

Table 1. The comparison of recognition rates (%)

Methods Test data	VPCA	Tensor Face	CGM /CD	CGM /ID	HGM /CD	HGM /ID	HGM /Itera.
View ₂	60.81	62.10	48.65	39.19	75.68	74.32	77.03
View ₃	67.57	74.32	60.81	58.11	74.32	81.08	93.24
View ₄	71.62	68.92	86.49	87.84	63.51	67.57	81.08
View ₅	48.65	36.49	24.32	25.68	36.49	51.35	56.76
View ₆	40.54	35.14	74.32	67.57	40.54	35.14	41.89
View ₇	47.30	35.14	24.32	39.19	41.89	68.92	68.92
View ₈	39.19	47.30	62.16	40.54	50.00	60.81	71.62
View ₉	59.46	67.57	62.16	74.32	62.16	68.92	93.24
View ₁₀	66.26	75.68	68.92	79.73	79.73	85.14	79.73
View ₁₁	64.87	67.57	64.86	75.68	77.03	77.03	77.03
View ₁₂	47.30	35.14	37.84	71.62	45.95	67.57	68.92
View ₁₃	27.03	36.49	33.78	35.14	48.65	66.22	67.57
View ₁₄	37.84	22.97	13.51	16.22	32.43	52.70	54.05
View ₁₅	51.35	62.16	44.59	36.49	62.16	72.97	68.92
View ₁₆	74.32	78.38	55.41	55.41	87.84	87.84	89.19
View ₁₇	47.30	54.05	60.81	51.35	83.78	63.51	81.08
View ₁₈	56.76	57.75	45.95	39.19	60.81	47.30	56.76
Ave.-17	53.42	53.95	51.11	52.54	60.17	66.38	72.18
Ave.-13	56.13	58.49	54.78	57.49	65.49	70.79	77.34

Face and VPCA. The results are shown in Table 1. Note that in Table 1 Ave.-17 means the average recognition rate of View₂ ~View₁₈. In Fig. 2(b), it is obvious that intervals View₅ ~View₆ and View₁₄ ~View₁₅ are too sparse to support valid view interpolation. For further comparison, we can obtain Ave.-13 that better reflects the effectiveness of view interpolation.

It is shown that TensorFace and VPCA are comparable. But TensorFace can separate the view and identity factors, and it also represents multi-view face images in a unified basis, which provides a general parametric face model with explicit physical meaning. The CGM is worse than TensorFace in both the CD and ID, which shows the concept-driven view manifold is not effective to capture the intrinsic structure of the view subspace. Because even face images are sampled with the same rotation interval, occlusion and appearance variations across views are actually quite different (See Fig. 2(b)). They are especially dramatic in View₅ ~View₆ and View₁₄ ~View₁₅ intervals. And the low recognition rates of HGM in these intervals show the performance of the proposed method depends on the distribution density of data in the manifold.

The recognition rates of HGM are improved 6.75%~9.36% and 12.96%~14.66% than VPCA in the CD and ID, respectively. Compared with TensorFace, that of HGM are improved around 7% and 12% in the CD and ID. The iterative ID-based HGM method is 18.23%~18.85% better than TensorFace. The main reason of the advantages of ID-based methods is that the reconstructed images in the ID involve the core tensor that carries the interaction between the view and identity factors, and is not used in the CD-based case.

4 Conclusions

We have extended a recently proposed nonlinear tensor decomposition method for multi-view face recognition that involves a continuous view manifold and provides a compact face generative model. More importantly, we have discussed the effect of different view manifolds on this face model. The experimental results show great promise of the new method where a hybrid view manifold is used. Our future studies will focus on how to optimize the view manifold via non-uniform sampling or advanced nonlinear interpolation methods.

Acknowledgements

This work is supported in part by the US NSF under Grant IIS-0347613 (G. Fan) and the Open-End Fund of National Laboratory of Pattern Recognition in China (X. Gao). Portions of the research in this paper use the oriental face database collected under the research of the Artificial Intelligence and Robotics (AI&R) at Xi'an Jiaotong University, China.

References

- [1] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. on PAMI*, 25(9):1063–1074, 2003.
- [2] A. Elgammal and C. S. Lee. Separating style and content on a nonlinear manifold. In *Proc. of IEEE CVPR*, 2004.
- [3] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. Multilinear singular value tensor decompositions. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.
- [4] C. S. Lee and A. Elgammal. Modeling view and posture manifolds for tracking. In *Proc. of IEEE ICCV*, 2007.
- [5] C. S. Lee and A. Elgammal. Simultaneous inference of view and body pose using torus manifolds. In *Proc. of ICPR*, 2006.
- [6] K. C. Lee and D. Kriegman. Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In *Proc. of IEEE CVPR*, 2005.
- [7] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. of IEEE CVPR*, 1994.
- [8] B. Raychev, I. Yoda, and K. Sakaue. Head pose estimation by nonlinear manifold learning. In *Proc. of ICPR*, 2004.
- [9] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12(6):1247–1283, 2000.
- [10] M. Vasilescu and D. Terzopoulos. Multilinear independent components analysis. In *Proc. of IEEE CVPR*, 2005.
- [11] M. A. O. Vasilescu and D. Terzopoulos. Multilinear image analysis for facial recognition. In *Proc. of ICPR*, 2002.
- [12] S. Zhou and R. Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel hilbert space. *IEEE Trans. on PAMI*, 28(6):917–929, 2006.