Transform-Invariant PCA: A Unified Approach to Fully Automatic Face Alignment, Representation, and Recognition

Weihong Deng, Jiani Hu,
Jiwen Lu, *Member, IEEE*, and Jun Guo

Abstract—We develop a transform-invariant PCA (TIPCA) technique which aims to accurately characterize the intrinsic structures of the human face that are invariant to the in-plane transformations of the training images. Specially, TIPCA alternately aligns the image ensemble and creates the optimal eigenspace, with the objective to minimize the mean square error between the aligned images and their reconstructions. The learning from the FERET facial image ensemble of 1,196 subjects validates the mutual promotion between image alignment and eigenspace representation, which eventually leads to the optimized coding and recognition performance that surpasses the handcrafted alignment based on facial landmarks. Experimental results also suggest that state-of-the-art invariant descriptors, such as local binary pattern (LBP), histogram of oriented gradient (HOG), and Gabor energy filter (GEF), and classification methods, such as sparse representation based classification (SRC) and support vector machine (SVM), can benefit from using the TIPCA-aligned faces, instead of the manually eye-aligned faces that are widely regarded as the ground-truth alignment. Favorable accuracies against the state-of-the-art results on face coding and face recognition are reported.

Index Terms—Face alignment, face coding, face recognition, eigenfaces, principal component analysis

1 Introduction

As early as 1987, Sirovich and Kirby first found that faces can be represented efficiently as a mean face plus a weighted linear combination of the eigenvectors of a convariance matrix of face images [1]. In this context, Turk and Pentland [2] developed a well-known Eigenfaces method, where the eigenfaces define a "face space" which drastically reduces the dimensionality of the original space, and face detection and recognition are then carried out in the reduced space. While undoubtedly successful in appearance based recognition, the theoretical foundation for the use of eigenfaces is less clear [3]. In practice, automatically detected faces are often subjected to random transformations, such as translation, rotation, and scaling, in images. In these cases, the eigenface method possibly produces severely blurred components that mostly account for the transformations and ignore the more interesting and useful structure. To address this problem, eigenface based approaches, as well as other face-related studies [4], [5], [6], have aligned the faces by the similarity transformation defined by landmarks such as two eye centers. This handcrafted alignment makes the recognition performance largely depend on the accuracy of landmark localization [7], [8]. Further, even all the facial landmarks have been precisely manually marked, it cannot guarantee that the resulting aligned faces are optimized for recognition. In this sense, a fundamental limitation of current face recognition methods is the lack of the connection between the face alignment and face representation.

In this paper, we develop a transform-invariant PCA (TIPCA) technique which aims to automatically learn the eigenface bases by characterizing the intrinsic structure of the human faces that are invariant to the in-plane transformations of training images. To achieve this objective, TIPCA alternately aligns the image ensemble and derives the optimal eigenspace in a manner that the mean square error (MSE) between the aligned images and their reconstructions is minimized. The optimization is effectively solved by iteratively: 1) creating the eigenspace of the aligned image ensemble using PCA; and 2) aligning each image to the eigenspace using simultaneous inverse compositional algorithm. The resulting TI-eigenspace defines a unified coordinate system for various applications on face alignment, representation, and recognition.

The effectiveness of the TIPCA technique is successfully tested on the large-scale FERET image ensemble involving the facial images of 1,196 subjects. Experimental results validate *the mutual promotion between image alignment and eigenface coding, which can eventually improve the recognition performance*. On one hand, improved alignment of the images leads to a compact image coding. On the other hand, the TI-eigenspace that excludes the transform-related components helps precise image alignment. For the recognition application, by aligning the training and test images to the unified TI-eigenspace, the transformation variation among images is minimized.

It should be noted that the alternating optimization between eigenspace and alignment has been explored in image coding, first developed by Schweitzer for holistic image [9], and then extended for the active appearance model by Baker et al. [10]. However, due to the difficulty in avoiding bad local minima, their works [9], [10] were limited to encode and align the small image ensemble of the same face, object or scene, and were not applicable to the multi-class recognition problem. Compared with previous works, the contributions of this paper are as follows.

- We develop a practical optimization procedure that is effective to simultaneously encode and align a large ensemble of thousands of faces under complex variations. The proposed low-to-high dimensional eigenspace alignment strategy helps the alternating optimization of TIPCA to find the good local minimum to accurately align complex image ensembles. The MSE between the aligned images and their reconstructions keeps decreasing by more iterations, and, finally, is 30 percent lower than that of the manually eye-aligned images.
- By aligning the gallery and probe images to a unified TI-eigenspace, we develop a fully automatic recognition system, and show that the recognition performance keeps improving as more iterations are taken at the training stage of TIPCA, which provides a paradigm for improving fully automatic recognition performance by the close relationship among image alignment, representation, and recognition.
- Extensive experiments are conducted to demonstrate that state-of-the-art invariant feature descriptors, such as local binary patterns (LBP), histogram of the oriented gradient (HOG) and Gabor, and classification methods, such as sparse representation-based classification (SRC) and support vector machine (SVM), can benefit from using the TIPCA-aligned faces, instead of the manually eye-aligned faces that have been used by almost all the current studies on face coding, recognition, and classification as the *ground-truth alignment*.

- *W. Deng, J. Hu, and J. Guo are with Pattern Recognition and Intelligent System Laboratory, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, PO Box 186, Beijing, 100876, China.*
  E-mail: {whdeng, jnhu, guojun}@bupt.edu.cn.
- *J. Lu is with the Advanced Digital Sciences Center, 1 Fusionopolis Way, #08-10 Connexis North Tower, 138632 Singapore.*
  E-mail: jiwen.lu@adsc.com.sg.

Manuscript received 18 Nov. 2012; revised 27 May 2013; accepted 7 Sept. 2013. Date of publication 2 Oct. 2013; date of current version 12 May 2014. Recommended for acceptance by M.Tistarelli.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2013.194

The rest of this paper is organized as follows: Section 2 describes the algorithm steps of TIPCA, Section 3 introduces the applications on image alignment, representation, and recognition by TIPCA, and Section 4 provides experimental results and discussions. Section 5 summarizes our conclusions and predicts future works.

2 Learning Transform Invariant Eigenspace

2.1 Learning Eigenfaces by PCA

Eigenfaces rely on the observation first made by Kirby and Sirovich that an arbitrary face image, denoted as $I \in \mathbb{R}^d$, can be compressed and reconstructed by adding a small number of basis images $\phi_j \in \mathbb{R}^d$,

$$I = \mu + \sum_{j=1}^{m} a_j \phi_j + e, \tag{1}$$

where μ is the average image, $\phi_1, \ldots, \phi_m$ are the ordered basis images derived from an ensemble of training images using principal component analysis. e represents noise components. The process of estimating the coding parameters $\mathbf{a} = (a_1, \ldots, a_n)^T$ is equivalent to projecting the image onto a linear subspace, which we can call the face space, i.e., $a_j = \phi_j^T (I - \mu)$. Turk and Pentland recognized that this set of coding parameters themselves could be used to construct a fast image matching algorithm.

In more detail, given a set of N example training images: I^i where $i = 1, 2, \ldots, N$, the formulation of eigenfaces is based on a general principle that the mean square error between input patterns and their reconstructions is minimized.

$$\arg\min_{\mu,\, \phi_j} \frac{1}{N} \sum_{i=1}^{N} \left(\min_{\mathbf{a}^i} \left\| I^i - \left(\mu + \sum_{j=1}^{m} a_j^i \phi_j \right) \right\|^2 \right). \tag{2}$$

While undoubtedly successful in appearance based recognition, the theoretical foundation for the use of eigenfaces is less clear. Formally, PCA assumes the face images, usually normalized in some way, such as co-locating eyes to make them comparable, are usefully considered as (raster) vectors [3]. However, the uncertainty on feature locations would makes eigenface bases characterize the transform-related components, rather than the intrinsic structures of the human face. In this sense, the fundamental limitation of current methodology is lack of the connection between the alignment of face images and the construction of face space. How to align the face such that the resulting face space could be as compact as possible is an interesting question.

2.2 Transform-Invariant PCA

For the clearness of the formulation, we represent the (unaligned) training images $I^i(\mathbf{x})$ and the basis images $\phi_j(\mathbf{x})$ in the pixel form, where $\mathbf{x} = (x, y)^T$ is a column vector containing the pixel coordinates, rather than the vector form. Let $\mathbf{W}(\mathbf{x}; \mathbf{p})$ denote the parameterized set of possible transformations, where $\mathbf{p} = (p_1, \ldots, p_n)^T$ is a vector of parameters. In TIPCA, the transformed image is represented as the linear combination of a small number of basis images as follows.

$$I(\mathbf{W}(\mathbf{x}; \mathbf{p})) = \mu(\mathbf{x}) + \sum_{j=1}^{m} a_j \phi_j(\mathbf{x}) + e(\mathbf{x}), \tag{3}$$

where the warp $\mathbf{W}(\mathbf{x}; \mathbf{p})$ takes the pixel $\mathbf{x}$ in the basis image $\phi_j(\mathbf{x})$ and maps it to the sub-pixel location $\mathbf{W}(\mathbf{x}; \mathbf{p})$ in the image I. Given a set of unaligned facial images $\{I^i\}_{i=1}^N$, we assume that the transformed images, denoted by $I^i(\mathbf{W}(\mathbf{x}; \mathbf{p}^i))$, reside near on a low-dimensional face space, and seek a set of basis images that minimize the sum of distance from the transformed images to the face space. In other words, the transform-invariant eigenfaces are learned based on a modified principle that minimizes the mean square error between *transformed patterns* and their reconstructions.

$$\begin{aligned} &\arg\min_{\mu,\, \phi_j} \frac{1}{N} \sum_{i=1}^{N} \left\{ \min_{\mathbf{p}^i,\, \mathbf{a}^i} \sum_{\mathbf{x}} \left[e^i(\mathbf{x}) \right]^2 \right\}, \text{ where} \\ &e^i(\mathbf{x}) = I^i(\mathbf{W}(\mathbf{x}; \mathbf{p}^i)) - \left[\mu(\mathbf{x}) + \sum_{j=1}^{m} a_j^i \phi_j(\mathbf{x}) \right]. \end{aligned} \tag{4}$$

As the introduction of the transform parameter $\mathbf{p}^i$ for each training image I^i, the minimization in (4) require more effort than computing eigenvectors of the covariance matrix. We solve it by iteratively optimize $\{\mu, \phi_j\}$ and $\{\mathbf{p}^i, \mathbf{a}^i\}$ in turn, assuming where necessary that estimates of the others are available. The training of TIPCA is initialized by the "coarse" eigenspace derived by applying standard PCA on the detected faces, and then start to learn the transform invariant eigenspace by alternately conducting the two following steps:

Step 1) Eigenspace based Alignment, i.e., fix μ, $\{\phi_j\}_{j=1}^d$ and optimize $\{\mathbf{p}^i, \mathbf{a}^i\}_{i=1}^N$. Given μ and $\{\phi_j\}_{j=1}^d$ that define a eigenspace, we use the simultaneous inverse compositional (SIC) algorithm[1] to optimize $\{\mathbf{p}^i, \mathbf{a}^i\}$ for each image I^i respectively so that the square error between the transformed image and its reconstruction is minimized. Specifically, the SIC algorithm performs a Gaussian-Newton gradient descent optimization simultaneously on the transform parameters $\mathbf{p}^i$ and the coding parameters $\mathbf{a}^i$. Let $\mathbf{q}^i$ be the concatenated parameter vector of $\mathbf{p}^i$ and $\mathbf{a}^i$, and the Jacobian (steepest descent) images of (4) is

$$\mathbf{J}(\mathbf{x}) = \left[\nabla\phi \frac{\partial \mathbf{W}}{\partial p_1^i}, \ldots, \nabla\phi \frac{\partial \mathbf{W}}{\partial p_n^i}, \phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x}) \right], \tag{5}$$

where $\nabla\phi = \nabla\mu + \sum_{j=1}^{m} a_j^i \nabla\phi_j$. In each step, the increment of the parameters is computed by

$$\Delta\mathbf{q} = -\left[\sum_{\mathbf{x}} \mathbf{J}^T(\mathbf{x}) \mathbf{J}(\mathbf{x}) \right]^{-1} \sum_{\mathbf{x}} \mathbf{J}^T(\mathbf{x}) e^i(\mathbf{x}), \tag{6}$$

where $e(\mathbf{x})$ is the square error with current parameters. At each step, the transform parameters are updated by $\mathbf{W}(\mathbf{x}; \mathbf{p}^i) \leftarrow \mathbf{W}(\mathbf{x}; \mathbf{p}^i) \circ \mathbf{W}^{-1}(\mathbf{x}; \Delta\mathbf{p})$ and the appearance parameters are updated by $\mathbf{a}^i \leftarrow \mathbf{a}^i + \Delta\mathbf{a}$. After limited steps, the square error between the transformed image and its reconstruction would converge to a local minimum with respect to $\mathbf{p}^i$ and $\mathbf{a}^i$.

Step 2) Eigenspace Update, i.e., fix $\{\mathbf{p}^i\}_{i=1}^N$ and optimize μ, $\{\phi_j\}_{j=1}^d$. If $\mathbf{p}^i$ is known, we can compute the transform $\mathbf{W}(\mathbf{x}; \mathbf{p}^i)$ for each input image I^i. The problem then reduces to a transformed version of principal component analysis. Specifically, we transform each image onto the aligned coordinate to give $I^i\left(\mathbf{W}(\mathbf{x}; \mathbf{p}^i)\right)$, stack it as a vector, and then perform PCA on these vectors, update μ to be the mean vector of the aligned ensemble, and ϕ_j, $j = 1, \ldots, d$ to be the eigenvectors of the covariance matrix with the d largest eigenvalues. Fig. 1 illustrates some example mean vectors and eigenvectors (in the image form) obtained during our experiment on the FERET database.

The alternating optimization of TIPCA terminates when the MSE in (4) stop to reduce.

2.3 Complexity Control

The iterative "image alignment–eigenspace update" procedure guarantees that the MSE can be reduced to be a local minimum.

1. The implementation details of the SIC algorithm, such as the image warping method, are described in [11], and a Matlab toolbox and the example code on a set of toy images are available in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.194.

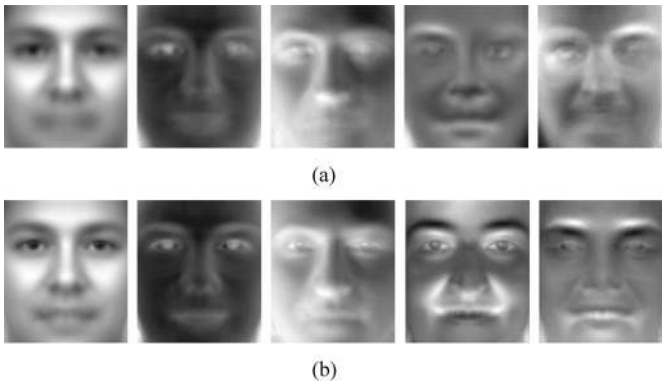

Fig. 1. The mean face and leading eigenfaces computed during the learning process of TIPCA. (a) those of the initialization. (b) those of the second iteration. Interestingly, the alternating optimization seems to "deblur" the basis images, which suggests that the alignment step is effective to reduce the transform-related components.

However, in complex problem with a large number of training faces, there may be millions of parameters and the algorithm tends to converge at a local minimum that is not good enough to address the subsequent representation and recognition tasks. In order to make TIPCA practical for real-world applications, we control the complexity of the optimization by the two following key strategies.

1) *Low-to-high dimensional eigenspace for alignment*. A important problem in TIPCA is the choice of m for alignment, which takes into account both the sufficient representation and the transform removal. If m is too small, the eigenspace cannot characterize enough variation of the image appearance that ensures the alignment algorithm to be converged. On the other hand, high dimensional eigenspace of the poorly-aligned images would include blur components that are misleading for alignment. To address this dilemma, the alignment step should select a relatively low dimension "blurred" eigenspace at initial iterations, which ensures the converge and, at the same time, excludes most blur components. Although the initial alignment is coarse, as the algorithm iterates, the alignment could become more and more precise. The precise alignment of the image ensemble makes principal eigenspace exclude the blur components, and thus allows the next alignment step to select higher dimensional "deblurred" eigenspace, which in turn benefit precise alignment. In summary, as the coarse-to-fine eigenspace is used for alignment, the mutual promotion of alignment and eigenface coding would iteratively reduce the MSE.

2) *Similarity transformation for alignment*. The goal of this paper is to make PCA invariant to image-plane transformation, while maintaining the clarity and spirit of eigenfaces and without resorting to more complex models, such as active appearance model [10] and morphable model [12]. Therefore, we prefer to focus on the deformations with few degrees of freedom, i.e., similarity transformations, which preserve linearity, angles and ratios of lengths. These geometric information (the relationship between facial features) are essential to the recognition of identity, gender, and expression. In addition, similarity transformation, which involves only four parameters, might simplify the optimization of the alignment and thus increase the converge rate for practical usages.

3 Unified Applications of Transform-Invariant PCA

The training stage of the TIPCA algorithm is an unsupervised iterative learning procedure with two outputs: an ensemble of aligned training images and a set of transform-invariant eigenfaces. Taking the former as the final result, TIPCA can be regarded as a approach to *batch image alignment*. More importantly, the set of TI-eigenfaces, which define a TI-eigenspace, provides an invariant appearance model leading to broad applications. This section details how the TI-eigenspace can be applied to *align*, *encode*, and *recognize* the unseen images.

3.1 TIPCA-Based Image Alignment

Image alignment aims to align a facial images, typically the output of the face detector, to the transform-invariant eigespace defined by the eigenfaces corresponding to the top m eigenvalues. This problem is well established in the computer vision domain, and we use the SIC algorithm because of its good converge rate [13]. Specifically, for an input image I, the SIC algorithm simultaneously recovers the transform parameter $\mathbf{p}$ and the appearance parameter $\mathbf{a}$ by solving following optimization problem:

$$\min_{\mathbf{p},\,\mathbf{a}} \sum_{\mathbf{x}} \left\{ I(\mathbf{W}(\mathbf{x};\mathbf{p})) - \left[\mu(\mathbf{x}) + \sum_{j=1}^{m} a_j \phi_j(\mathbf{x}) \right] \right\}^2. \tag{7}$$

The complexity of the alignment algorithm increases dramatically with a large m, but, fortunately, low dimensional TI-eigenspace, e.g., $m = 20$, is sufficient to perform precise alignment. For the recognition/classification problem, the gallery and test images should be aligned to the same TI-eigenspace to make them comparable within an unified coordinate.

3.2 TIPCA-Based Feature Extraction (Encoder)

Feature extraction aims to encode the image by identifying the most expressive features, i.e., the eigenvectors with the largest eigenvalues $\phi_1, \ldots, \phi_d$, while those with small eigenvalues are assumed to contain noise and are cut off accordingly. Furthermore, in order to achieve the transform-invariant property, the feature extraction of TIPCA are conducted by two separated procedures: 1) align the image by solving (7) with a selected dimension m, as detailed in Section (3.1), and 2) project the aligned image vector onto the leading d TI-eigenvectors.

$$a_i = \sum_{\mathbf{x}} \phi_i^T(\mathbf{x})[I(\mathbf{W}(\mathbf{x};\mathbf{p})) - \mu(\mathbf{x})],\ i = 1, \ldots, d. \tag{8}$$

The number of principal components for subsequent reconstruction or recognition is typically user-defined. In face recognition, the aligned image could be normalized to zero mean and unit length for better invariance to illumination before projected to the eigenspace.

3.3 TIPCA-Based Image Reconstruction (Decoder)

In the Eigenfaces method, the principal components and eigenvectors (eigenfaces) can be combined to reconstruct the image of a face. Similarly, TIPCA can be used to reconstruct a face image in the following way.

$$I(\mathbf{W}(\mathbf{x};\mathbf{p})) \approx \mu(\mathbf{x}) + \sum_{j=1}^{m} a_j \phi_j(\mathbf{x}). \tag{9}$$

In addition, TIPCA also extracts the transform parameters $\mathbf{p}$. The original (unaligned) image can be recovered by backwards transforming the reconstructed aligned image of (9) from the aligned coordinate to the original coordinate as following:

$$I(\mathbf{x}) \approx \mu\big(\mathbf{W}^{-1}(\mathbf{x};\mathbf{p})\big) + \sum_{j=1}^{m} a_j \phi_j\big(\mathbf{W}^{-1}(\mathbf{x};\mathbf{p})\big), \tag{10}$$

where a pixel $\mathbf{x}$ in the aligned images is mapped to the original pixel $\mathbf{W}^{-1}(\mathbf{x};\mathbf{p})$.

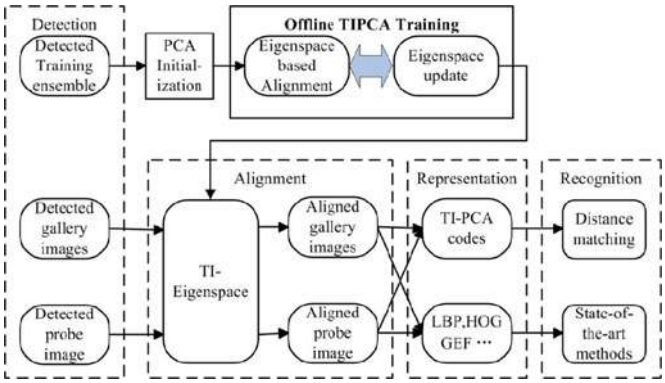

Fig. 2. The TIPCA based framework for fully automatic face alignment, representation, and recognition. In the training stage, the alternating optimization of TIPCA reduce the transform-related component within the eigenspace progressively, and finally output a transform-invariant eigenspace. In the test stage, the test images are aligned and projected to the TI-eigenspace for representation and recognition. State-of-the-art recognition algorithms can benefit from aligning the gallery and probe images to a unified TI-eigenspace.

3.4 TIPCA-Based Fully Automatic Face Recognition

By combining the TIPCA-based image alignment and feature extraction, a fully automatic eigenfaces based recognition algorithm can be readily figured out, as illustrated in Fig. 2. In the training stage, a TI-eigenspace is first automatically learned from an ensemble of training images, and the TI-principal components of those gallery images are then extracted, as detailed in Section 3.2, and stored. In the testing stage, the TI-principal components of the probe image is first extracted. Finally, the nearest neighbor classifier is used for classification. In our experiments, the distances between two arbitrary feature vectors, $\mathbf{a}^i, \mathbf{a}^j \in \mathbb{R}^d$ used in our experiment are defined as follows:

$$\begin{aligned} d_{Ed} &= (\mathbf{a}^i - \mathbf{a}^j)^T(\mathbf{a}^i - \mathbf{a}^j), \\ d_{Md} &= (\mathbf{a}^i - \mathbf{a}^j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{a}^i - \mathbf{a}^j), \\ d_{Wc} &= \frac{(\mathbf{a}^i - \mathbf{a}^j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{a}^i - \mathbf{a}^j)}{\|\mathbf{a}^i\| \cdot \|\mathbf{a}^j\|}, \end{aligned} \tag{11}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is the covariance matrix of the training data. For the decorrelated principal components, $\boldsymbol{\Sigma}$ is diagonal and the diagonal elements are the (eigenvalues) variance of the corresponding components. Ed, Md, Wc defines the euclidean distance, Mahalanobis distance, whitened cosine distance, respectively.

Beyond the direct matching of the eigenface codes, TIPCA can benefit various recognition methods via precise image alignment. By automatically learning the TI-eigenfaces from the training ensemble, and aligning both the gallery and the probe images to an unified eigenspace defined by TI-eigenfaces, any subsequent recognition method would benefit from the precise alignment. In this manner, *TIPCA can be incorporated with most state-of-the-art recognition algorithms*, besides the eigenface based approaches, and makes them operated in a fully automatic way. Some applications will be demonstrated in the following experiment section.

4 Experiments

In this section, we evaluate the effectiveness of TIPCA on image alignment, coding, and recognition using 3,307 facial images of 1,196 subjects from the gray-level FERET database, which is a standard testbed for face recognition technologies [14]. The tested images display diversity across gender, ethnicity, and age, and were acquired without any restrictions imposed on expression, illumination and accessories (See Fig. 3 for examples). Specifically, the experiment follows the standard data partitions of the FERET database:

- *gallery training set* contains 1,196 images of 1,196 people.
- *fb probe set* contains 1,195 images taken with an alternative facial expression.
- *fc probe set* contains 194 images taken under different lighting conditions.
- *dup1 probe set* contains 722 images taken in a different time.
- *dup2 probe set* contains 234 images taken at least a year later, which is a subset of the dup1 set.

Practical face recognition algorithms commonly consist of two parts: alignment (normalization) and recognition. In the influential FERET'97 evaluation, partially automatic algorithms are given the coordinates of the eye centers for normalization [14]. Since then, the eye-aligned facial images have become the *de facto* standard for face recognition research. To ensure the reproducibility of our results, the ground-truth eye coordinate file of the FERET database is used, and the publicly available CSU face identification evaluation system [15] was utilized to provide the eye-aligned images, which registers the two eye centers at (30, 45) and (100, 45) in a 150 × 130 facial image. Fig. 3a shows some *eye-aligned faces* which are used in our experiments, and one can see from the figure that the intra-personal variability of this database is complex. Even though we have used the manually labelled eye coordinates of the FERET distribution, a few faces are not well aligned due to the slight errors of manual label (See the second row of Fig. 3a).

Our algorithm starts with facial images detected by the common face detectors. Viola and Jones$A$ face detector,[2] which outputs a square bounding box indicating the predicated center of the face and its scale, is applied for its stable performance and high speed. Given a detected face image of the width w, we crop the face according to the eye locations[3] of $(0.305w, 0.385w)$ and $(0.695w, 0.385w)$ using the CSU face identification evaluation system [15]. The cropped and scaled face images of a standard size 150 × 130, which subsequently is referred to as "*detected faces*", is illustrated in Fig. 3b. These detected faces are used for the initialization of TIPCA learning.

4.1 Learning Transform-Invariant Face Space

The alternating optimization of TIPCA starts with the detected faces of the 1,196 gallery images, and the iterative learning of our experiment involves 23 iterations. As the algorithm iterates, the dimension of the eigenspaces used in the alignment step increases from 20 to 100 as detailed in Fig. 4 a. As the detected faces display natural variations on translations, scales, and rotation angles, they are expanded by 30 pixels all around, forming a 210 × 190 input image, for the SIC algorithm to seek the optimal similarity transformation. At each iteration, the SIC algorithm is initialized with the 150 × 130 bounding box of the detected face, and its maximum number of gradient descent steps is set to 20.

To monitor the effectiveness of the alternating optimization, in each iteration, we reconstruct the aligned images by their first 100 eigenfaces, and compute the MSE of all the 1,196 training images. Fig. 4b plots the MSE as a function of the number of iterations, which monitors how the objective function value (4) changes in each iteration. One can see from the figure that the MSE keeps deceasing, which clearly shows the mutual promotion of the eigenface coding and image alignment. In other

2. We use the OpenCV implementation of the Viola and Jones's face detector [16]. Since there is only one face in each image, we reduce the false alarms by reserving the bounding box of the maximum size in each image. The detector missed only six faces out of all the 3,307 images involved in our experiments, and we have manually completed these six bounding boxes.

3. They are roughly the averaged locations of the two eyes of the typical bounding faces determined by the VJ face detector.

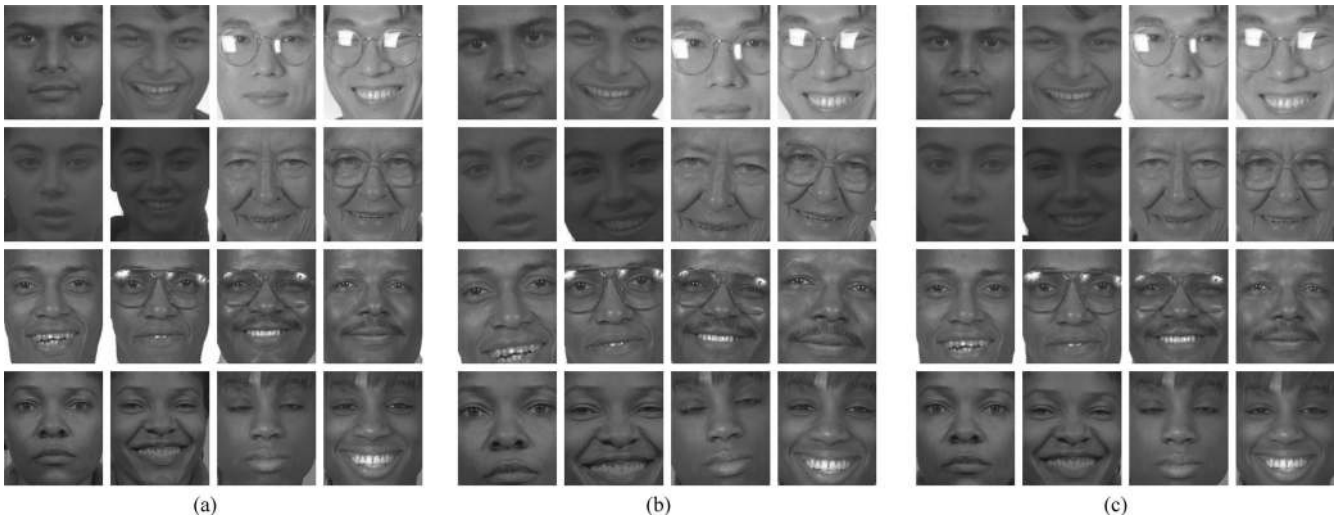

Fig. 3. Three types of aligned faces with the size of 150 × 130 used in our experiments. (a) manually eye-aligned faces which has been used in most studies on face recognition, gender and expression classification. (b) the detected faces, which are directly cropped and resized from bounding box of the face detector. (c) TIPCA-aligned faces, which are generated by aligning the images to a unified low-dimensional TI-eigenspace.

words, as the algorithm iterates, more and more intrinsic structures are represented by the TI-eigenfaces, and at the same time, the transformation among the images are largely reduced by the alignment step.

For comparison purpose, Fig. 4b also displays the MSE between the eye-aligned images and the reconstructions by their first 100 eigenfaces. As expected, the MSE of the eye-aligned faces are lower than that of the detected faces (the MSE at iteration 0), which explains the feasibility of eye-aligned faces for image coding and recognition. However, eye-based alignment is a heuristic approach without any theoretical justification, and we find that TIPCA-aligned faces are better reconstructed than eye-aligned faces after only one iteration. After 23 iterations, the MSE of the TIPCA-aligned faces is about 30 percent lower than that of eye-aligned faces.

To evaluate the generalization ability of TIPCA, in each iteration, we align the probe images using the identical eigenspace that used in the alignment step, then reconstruct the aligned probe images by the first 100 eigenfaces (computed from the aligned training images at that iteration), and finally compute the MSE of the 2,111 probe images. The results are shown in Fig. 4c. Comparing Figs. 4c and 4b, we find 1) the MSE of the unseen probe images are higher than that of the training images; 2) the MSE of TIPCA-aligned faces is also notably lower than that of eye-aligned faces; and 3) the MSE of the probe images also generally deceases as the algorithm iterates. These results indicate the TIPCA has improved generalization ability to represent facial images than traditional eigenfaces based approaches.

Besides the MSE, we also measure the quality of reconstructed images by signal-to-noise ratio (SNR) [17]. Fig. 5 plots the average SNR as a function of dimension, i.e., the number of the components, used for reconstruction. When the dimension is larger than 100, the SNR of training set increase linearly as the dimension. For the test set, however, the SNR seems saturated. Similar to the results on reconstruction error, TIPCA outperforms PCA (by about 0.5-0.8 dB) on both the training and the testing image sets. On the unseen probe images, TIPCA achieves about twice the coding efficiency of PCA. Specifically, TIPCA uses 75 components to obtain an SNR of 7 dB while PCA requires about 150 components. To reach a SNR of 8 dB, TIPCA uses 250 components while PCA requires over 500.

To visualize the reconstruction effects of TIPCA, Fig. 6b shows five reconstructed images of a probe image using the first d (d = 20, 40, 60, 80, 100) TI-eigenfaces. The reconstructed images become clearer as the number of eigenfaces is increased. For comparison, Fig 6a shows the PCA based reconstruction on the same (eye-aligned) probe image, where the same number of eigenfaces, learned from the eye-aligned ensemble by PCA, were also used. Clearly, Fig. 6b displays more appearance details, such as the eyeglass frame and the texture of the beard, where Fig. 6a are blur. Although optimal for coding in the least MSE sense, PCA performs worse than TIPCA because of two possible

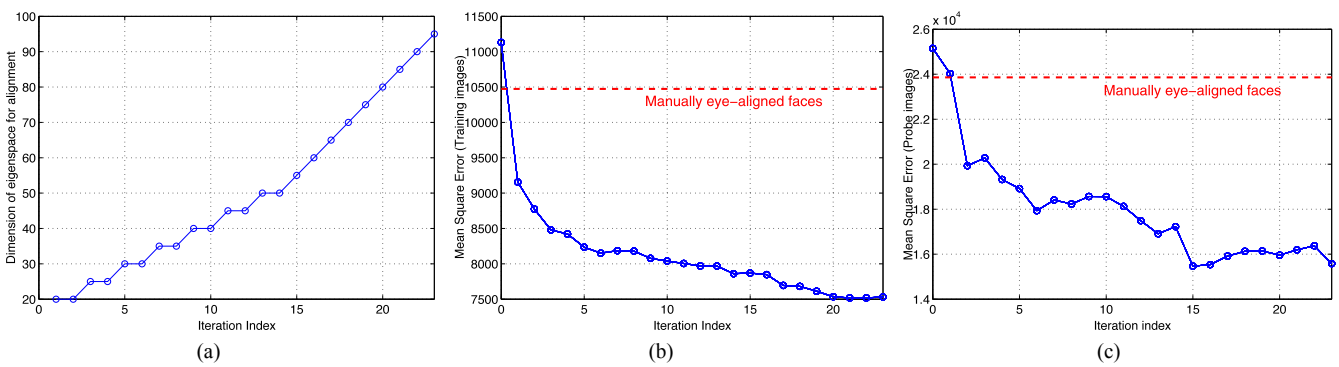

Fig. 4. The iterative learning of TIPCA. (a) the increasing dimension of the eigenspace used for the alignment step as the algorithm iterates. (b) the MSE of aligned training images by 100 dimensional eigenspace as the algorithm iterates. (c) the MSE of aligned (unseen) probe images by 100 dimensional eigenspace as the algorithm iterates. The horizontal red dashed lines in (b) and (c) indicate the MSE of the eye-aligned faces by traditional eigenface coding scheme.

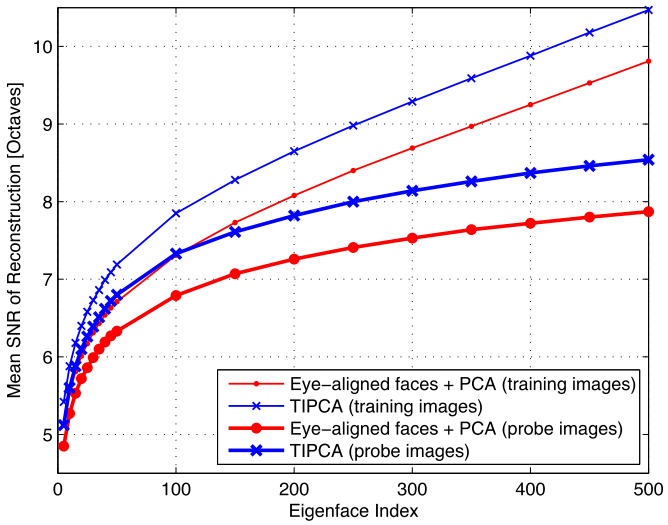

Fig. 5. The average signal-to-noise ratio as a function of number of eigenfaces used for reconstruction.

reasons: 1) the traditional eigenfaces characterize the transform-related components that contain in the eye-aligned training ensemble, and thus their linear combinations inevitably become blur; and 2) although aligned by manually labelling, the input (eye-aligned) image is not well aligned to the eigenspace. In practice, one or two pixel alignment error may cause reconstruction to be blur, such as that in eye-glass frame. Fig. 3c shows some examples of the precisely aligned faces by TIPCA which ensures efficient coding and high-quality reconstruction.

4.2 Appearance Based Face Recognition

This experiment evaluates whether the transform-invariant coding of TIPCA can directly improve the recognition accuracy. As in the common scheme, eigenfaces are constructed from a training set of face images and particular probe faces are recognized by comparing the principal components (eigenface weights). The number of principal components to remains is typically use-defined. Recognition performance will suffer from insufficient information if dimensionality is underestimated. On the other hand, an overestimate of dimension will introduce noisy components which also reduces performance [17]. Our empirical results validated that optimal recognition performance is achieved with a dimensionality roughly 150, using nearest-neighbor classification based on three popular distance measures defined in (11), namely the euclidean distance, the Mahalanobis distance, the whitened cosine distance.

Table 1 shows the face recognition performance in 150 dimensional eigenspace derived by PCA and TIPCA. PCA is evaluated using both eye-aligned faces and detected faces. Besides the finally optimized performance, we also test the intermediate results of TIPCA after 1, 5, 9, 13, 17 iterations. The averaged accuracy over four probe sets and the three distance measures keeps increasing as more iterations of TIPCA learning is applied. This finding indicates that the MSE in Fig. 4b and 4c is a effective indicator of the quality of face alignment for recognition. After five iterations, the average accuracy of TIPCA starts to surpass that of the eigenface approach based on eye-aligned face. Finally, TIPCA outperforms PCA by a margin of about 8 percent in average. The superiority of TIPCA seems more apparent when the latter two distances are applied. For instance, using the whitened cosine distance on the dup2 probe set, TIPCA boosts the accuracy of PCA by about 15 percent (from 27.8 to 42.3 percent). This may be because the latter two distances, which weight the low-variance components more heavily, makes the blur components of PCA to be more harmful for recognition.

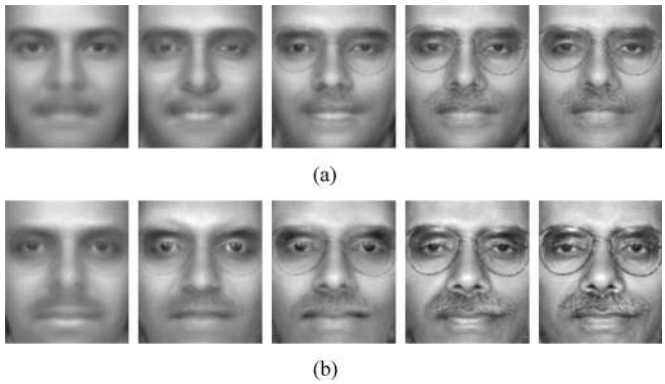

Fig. 6. Some reconstructed images based on (a) PCA and (b) TIPCA with the dimension $d = \{20, 40, 60, 80, 100\}$. Note that reconstructed region of PCA is manually defined by two eye centers, but that of TIPCA is automatically selected from the detected face.

By observing the finally optimized performance with TI-eigenspace learned from 23 iterations, we find that the performance differences using different eigenspace dimensions for alignment is not significant. This suggest that TIPCA can be applied in an efficient way using low dimensional eigenspace for alignment, while keeping highly accurate recognition performance. By comparing the performance with equivalent #D and different #I, one can find that better recognition performance can be achieved by more training iterations.

4.3 Face Recognition with Local Descriptors

By reducing the transform-related components among the aligned faces, TIPCA boosts the performance of appearance based recognition. However, the recognition accuracy is relatively low since the TIPCA based appearance features still suffer from the intra-personal variations caused by illumination, expressions, and occlusion. One possible solution is to apply local descriptors to represent the aligned faces rather than the pixel. It is interesting to evaluate the effectiveness of TI-Eigenfaces based alignment on face recognition with local descriptors, which are robust to mis-alignment by themselves.

In this experiment, we compare the TIPCA-aligned face with eye-aligned face for face recognition with three widely used local descriptors:

- *Local Binary Patterns* [18]: The basic idea of LBPs is that binary values are calculated from a pixel neighborhood and the binary values are concatenated to one binary value. The $LBP^{U2}_{8,2}$ operator [18] is adopted in 7×7 pixel cell, for each cell accumulating a local histogram of 59 uniform patterns over the pixels of the cell. The combined histogram entries form the representation, resulting a 23,364 ($22 \times 18 \times 59$) dimensional feature vector.
- *Histogram of the Oriented Gradient* [19]: The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients. The image is first divided into multiple 5×5 pixel cells, a local histogram of 18 signed gradient directions over the pixels of the cell are accumulated for each cell. For better invariance to illumination, "L2-Hys" contrast-normalization [19] with the threshold 0.2 is applied over each 10×10 pixel block. The combined histogram entries form the final 14,040 ($30 \times 26 \times 18$) dimensional feature vectors.
- *Gabor Energy Filters (GEF)* [20]: A family of five scales and eight orientations of Gabor filters are adopted. Each energy filter consists of a real and imaginary part which are squared and added to obtain an estimate of energy at a

TABLE 1
Comparative FERET Recognition Rates of PCA and TIPCA Using Three Popular Distance Measures of 150 Principal Components

Alignment and feature extraction	Euclidean				Mahalanobis				Whitened Cosine				Average
	fb	fc	dup1	dup2	fb	fc	dup1	dup2	fb	fc	dup1	dup2	
Eye-aligned face+PCA	**73.1**	**17.0**	30.1	12.8	68.7	45.4	28.7	17.1	75.4	60.3	39.1	27.8	41.29
Detected face+PCA	67.4	10.8	21.3	10.3	59.2	16.0	19.9	19.2	67.7	34.0	28.0	31.2	32.08
TIPCA (#I= 1, #D=20)	61.8	9.8	27.6	12.4	66.2	33.5	32.8	23.9	73.6	51.0	43.9	34.2	39.23
TIPCA (#I= 5, #D=30)	66.4	12.4	29.9	13.2	72.6	48.5	35.3	26.9	79.1	68.6	44.0	35.9	44.40
TIPCA (#I= 9, #D=40)	67.0	15.5	29.2	14.1	73.1	50.5	36.7	27.8	79.7	66.0	47.2	40.2	45.58
TIPCA (#I=13, #D=50)	68.7	14.4	31.2	12.8	74.7	52.1	37.0	26.1	80.7	68.0	48.6	39.3	46.13
TIPCA (#I=17, #D=70)	69.0	15.5	30.5	13.2	74.7	53.1	41.1	33.8	80.0	70.6	50.8	**43.2**	47.96
TIPCA (#I=23, #D=20)	69.8	15.5	31.2	13.7	75.4	51.0	40.9	28.6	80.9	67.0	**52.1**	**43.2**	47.44
TIPCA (#I=23, #D=30)	69.9	14.4	**32.8**	**15.8**	75.1	50.0	41.7	30.8	81.4	66.5	49.9	41.0	47.44
TIPCA (#I=23, #D=40)	71.0	14.4	31.0	15.4	75.1	50.5	40.0	30.8	80.8	69.1	49.6	39.3	47.25
TIPCA (#I=23, #D=50)	71.3	13.9	30.3	13.7	**76.3**	54.6	40.3	**34.6**	80.8	74.2	49.3	41.0	48.36
TIPCA (#I=23, #D=70)	71.0	12.9	31.4	**15.8**	76.1	54.6	40.6	29.1	81.6	70.1	51.7	42.3	48.10
TIPCA (#I=23, #D=100)	72.1	13.9	32.4	**15.8**	75.9	**56.7**	**42.7**	33.8	**81.5**	**75.8**	51.4	42.3	**49.52**

'#I' suggests the number of iterations taken in the training stage.
'#D' suggests the number of eigenspace dimensions used for alignment in the test stage.

particular location and frequency band. The response of each filter is downsampled by a factor of 64, and then normalized to zeros mean and unit length. The combined responses of the 40 filters result in a 12,160 ($19 \times 16 \times 40$) dimensional feature vector.

The local feature vectors are first normalized to zero mean and unit length, and then subjected to PCA for dimensionality reduction. The previous experiment has shown that the whitened cosine similarity distance performs best for eigenface code, we therefore apply it again to evaluate the performance of local feature based eigenface code. Different from the appearance based recognition, local feature based recognition does not suffer from low-variance components. Hence, we select the full dimensional, i.e., 1,195-dimensional, eigenface codes for recognition, and the results are tabulated in Table 2. As expected, the recognition accuracies are largely improved by 15-40 percent compared with the appearance based results in Table 1.

TIPCA-aligned face based local descriptors achieve successively increased accuracies as the algorithm iterates, start to outperform eye-aligned face based descriptors after five iterations, and finally boosts the average accuracy by about 3.5 percent. It is a significant improvement considering that the accuracies using eye-aligned face are already very high, especially on the fb and fc sets. The superiority of TIPCA-aligned face seems more apparent on the dup1 and dup2 probes. For instance, using HOG feature on the dup2 probe set, TIPCA-aligned faces boosts the accuracy of eye-aligned faces by about 11 percent (from 71.8 to 82.9 percent). Interestingly, the precise alignment by TIPCA could alter the relative recognition ability of the local descriptors. For instance, compared with LBP and HOG, GEF performs better on the most challenging dup2 probe set using eye-aligned faces, but becomes worse using TIPCA-aligned faces. This may be because the GEF is more robust to the mis-alignment by eye-based alignment, but become less discriminative when precise alignment is available.

4.4 Leave-Out Test on Generalization Capability

Previous experiments use the full set of gallery images for the training of TIPCA, which indicates all the identity-related information is encoded in the TI-eigenspace. However, in the large-scale face recognition/retrivial applications, it is difficult to collect all the gallery subject for model training. This experiment aims to test the generalization capability of TIPCA to align and represent unseen subjects. Specifically, the 234 images of dup2 probe set involves 75 subjects. We leave the corresponding 75 gallery images out of the training stage of TIPCA, and then compare the "leave-out" recognition performance with those reported in previous experiments.

TABLE 2
Comparative FERET Recognition Rates of Detected Face, Eye-Aligned Face, and TIPCA-Aligned Face Using the Whitened Cosine Distance Measure of Three Popular Local Descriptors

Alignment	LBP				HOG				GEF				Average
	fb	fc	dup1	dup2	fb	fc	dup1	dup2	fb	fc	dup1	dup2	
Eye-aligned face	**98.1**	93.3	80.5	70.9	**97.3**	94.3	79.9	71.8	95.6	99.0	75.2	75.6	85.96
Detected face	92.2	72.2	64.5	63.2	89.1	72.7	57.8	60.7	93.6	96.4	65.9	69.7	74.83
TIPCA (#I= 1, #D=20)	93.1	88.7	74.8	69.7	90.8	88.1	71.2	65.0	93.6	96.9	71.7	71.8	81.28
TIPCA (#I= 5, #D=30)	95.6	93.8	79.2	74.4	94.0	95.4	77.7	71.4	96.5	98.5	75.3	76.5	85.69
TIPCA (#I= 9, #D=40)	96.6	96.4	80.6	76.1	95.1	96.9	81.3	79.1	96.4	99.0	75.6	77.4	87.54
TIPCA (#I=13, #D=50)	97.1	97.4	83.2	78.2	96.4	97.9	81.9	77.8	96.6	**99.5**	76.3	77.4	88.31
TIPCA (#I=17, #D=70)	96.7	97.9	84.9	80.8	95.9	97.4	83.9	82.1	**96.8**	98.5	77.1	77.8	89.15
TIPCA (#I=23, #D=20)	96.9	95.9	**85.6**	79.1	95.4	**98.5**	**86.0**	80.8	96.5	98.5	79.5	79.5	89.35
TIPCA (#I=23, #D=30)	96.4	94.8	84.2	79.9	94.9	94.3	85.3	82.5	96.2	99.0	77.0	78.2	88.56
TIPCA (#I=23, #D=40)	96.7	96.4	84.8	**81.2**	95.8	96.4	85.0	**82.9**	96.5	**99.5**	76.9	78.6	89.23
TIPCA (#I=23, #D=50)	96.7	**98.5**	84.3	78.6	95.2	96.9	84.9	83.3	96.6	99.0	77.4	79.5	89.24
TIPCA (#I=23, #D=70)	96.9	**98.5**	85.0	79.9	95.7	**98.5**	84.9	**82.9**	96.6	99.0	**78.7**	**81.2**	**89.82**
TIPCA (#I=23, #D=100)	97.2	97.4	84.3	79.1	96.2	96.9	85.0	**82.9**	96.2	99.0	78.4	**81.2**	89.48

TABLE 3
Comparative FERET dup2 Recognition Rates with/without the Involved Subjects for TIPCA Training

TIPCA #I=23	Pixel	LBP	HOG	GEF
#D=20	43.2 / **43.6**	79.1 / **80.3**	80.8 / **82.9**	79.5 / **79.9**
#D=30	41.0 / **41.9**	**79.9** / 78.6	82.5 / 82.5	78.2 / **81.6**
#D=40	39.3 / **44.9**	**81.2** / 77.8	**82.9** / 81.6	78.6 / **80.8**
#D=50	41.0 / **43.6**	78.6 / **79.5**	**83.3** / 81.6	79.5 / **80.8**
#D=70	42.3 / **42.7**	**79.9** / 78.2	**82.9** / 81.6	81.2 / **81.6**
#D=100	**42.3** / 40.6	79.1 / 79.1	**82.9** / 80.8	**81.2** / 79.9

TABLE 4
Comparative FERET Recozgnition Rates on Differently Aligned Faces Using SRC

Alignment	fb	fc	dup1	dup2
Eye-aligned faces+SRC	83.2	74.2	46.1	30.8
Detected faces+SRC	73.5	38.7	34.5	33.3
DSRC [23]	**95.2**	28.4	46.1	20.3
TIPCA-aligned faces+SRC	86.0	**80.9**	**60.9**	**50.0**

Table 3 reports the comparative FERET dup2 recognition rates with/without the involved subjects for TIPCA training. It is somehow surprising that the recognition accuracies of the TIPCA-aligned faces are almost equivalent whether the recognized subjects are involved in the training set or not. In the cases where only 20 dimensional TI-eigenspace is used for alignment, the "leave-out" recognition accuracies are even slightly better than those of previous experiments. This extraordinary generalization ability to align and represent unseen subjects endows TIPCA the practical usefulness in the large-scale face recognition/retrivial applications. It is possible to build a subject-independent TI-eigenspace by which generic facial images can be efficiently and precisely aligned for accurate recognition.

4.5 Face Recognition via Sparse Representation

Sparse representation-based classification [21], [22] is a face recognition breakthrough in recent years. To solve the misalignment problem in SRC, a deformable sparse recovery and classification (DSRC) [23] have used tools from sparse representation to address the alignment problem given sufficient number of gallery images per subject. In contrast, TIPCA builds an unified appearance model for aligning all gallery subjects, regardless of the sample size per subject, which might be a good alternative for DSRC in the undersampled situation. Therefore, it is interesting to combine TIPCA-based alignment and SRC-based recognition,[4] and compare its performance with DSRC.

For comparison purpose, we also apply SRC to the eye-aligned faces and the detected faces. For a fair comparison, all the aligned faces are all downsampled to 75×65 to be compatible with those used in [23]. The face recognition performance of SRC using the four alignment methods is tabulated in Table 1, which shows that the best performance on three of the four probe sets is achieved using TIPCA-based faces. DSRC performs better than TIPCA+SRC only when expression variation (fb set) is presented. In contrast, using TIPCA-aligned faces (#I=23, #D=20) achieves substantially improved accuracy (about 6 to 18 percent) than other alignment methods on the fc, dup1, and dup2 probe sets. This suggests that TIPCA constructs an unified appearance model that is more robust against the complex variations of the facial appearance. Additionally, by comparing Table 4 with Table 1, we find that SRC performs better than the eigenface based approaches with various distance measures. This indicates that the sparsity assumption of SRC indeed provides some superiority over traditional appearance based approaches. In particular, with the precise alignment by TIPCA, the superiority of SRC becomes more notable.

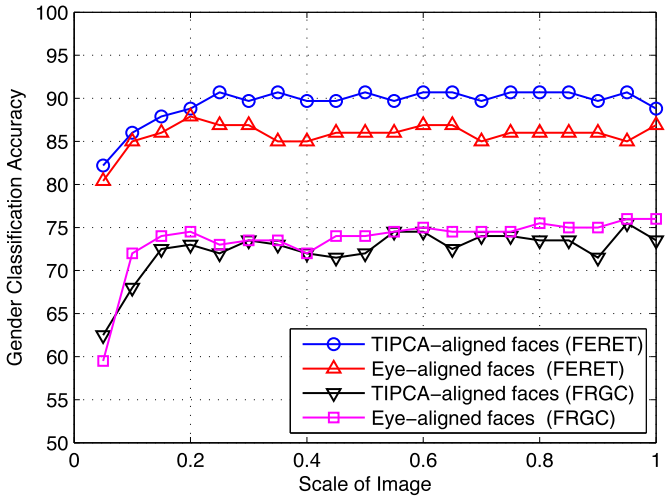

Fig. 7. The gender classification rate of SVM as a function of image resolution using three image alignment methods. The full resolution of the image is 150×130 at the scale of 1.

4.6 Preliminary Results on Gender Classification

Besides the identify recognition, it is also useful to extract categorical information from faces, such as gender or ethnicity. Makinen and Raisamo presented a systematic study on gender classification with automatically detected and aligned faces [5]. One of the findings was that current automatic face alignment methods, such as AAM [24] and integral projection [25], perform worse than the manually located eye-based alignment for gender classification. Therefore, we evaluate whether the TIPCA based alignment can be better than manually located eye-based alignment for gender classification. We follow the training and testing partitions[5] of Makinen and Raisamo [5]. There are 304 training images (152 males/152 females) and 107 test images (60 males/47 females). We compare the performance of eye-aligned faces and TIPCA-aligned faces (#I = 23, #D = 40). Further, as the gender classification is related to image resolutions, we resize the aligned images by factors from 0.05 to 1, with an interval of 0.05, to better examine the quality of image alignments. Classification is performed by support vector machine,[6] which is widely regarded as the best gender classifier, and the resulting accuracies are illustrated in Fig. 7. TIPCA based alignment improves manually located eye-based alignment for gender classification by an accuracy of 1-5 percent under varying image resolutions.

To evaluate the generalization ability against the uncontrolled lighting condition, we further test the gender classification accuracy on 200 images (100 males/100 females, one image per subject) from the FRGC uncontrolled image sets. Note that both the TI-eigenspace (#I = 23, #D = 40) and the SVM model are learned from the FERET database. Fig. 7 shows that cross-database classification performance on FRGC database drops to 70–75 percent.

4. The Homotopy method is applied to solve the ℓ^1-minimization problem with the regularization parameter $\lambda = 0.003$. The source code was downloaded at http://www.users.ece.gatech.edu/~sasif/homotopy/.

5. The lists of the training and test images were downloaded from http://www.sis.uta.fi/~em55910/data sets/.

6. The LIBSVM implementation [26] of linear C-SVM is applied, and ten-fold validation on the training set are used to select the optimal parameter C from $\{1, 2, \ldots, 20\}$.

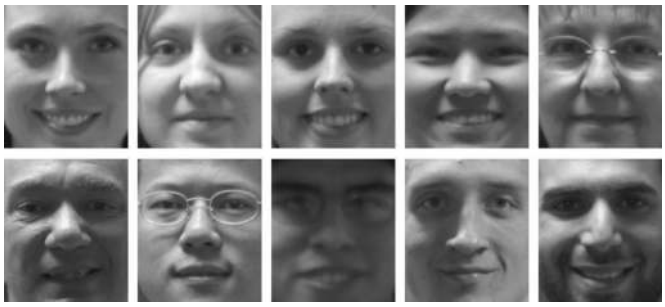

Fig. 8. The TIPCA-aligned facial images of FRGC database for gender classification test. Note that the TI-eigenspace learned from FERET database is used for alignment.

Although the training set of does not contain uncontrolled illuminations, TIPCA aligns the FRGC uncontrolled images precisely, as shown in Fig. 8. As a result, the accuracies on TIPCA-aligned faces are comparable to those on the eye-aligned faces at all resolutions. It should be noted that the main concern of our work is face recognition, and these preliminary results are only aimed to show some other potential applications of TIPCA. To design the dedicated protocol for gender classification, we refer the reader to the recent work of Grosso et al. [27].

4.7 Computational Issues

Table 5 enumerates the CPU time of the training process (with different numbers of iterations) of TIPCA on the 1,196 images using our C++ implementation on a PC with Quad Core 2.80 GHz Pentium CPU and 4 GB memory. In particular, training with 23 iterations takes about 8.74 hours. Although this training process is relatively slow, it is offline, fully automatic (avoid tedious manual labeling), and scalable to a huge number of training images by parallelized training. Because over 99 percent computational cost focuses on the alignment step which is independent for each training image, one can easily implement training parallelism by distributing the alignment step to multiple machines. At each iteration, one central machine collects the aligned faces for updating eigenspace, and duplicates the updated eigenspace on other machines.

The applications of TIPCA are efficient. Because TIPCA builds a unified TI-eigenspace for aligning both gallery and probe images, the alignment time per image is not related to the number of images per gallery subject or the number of subjects involved in the system. The alignment time per image depends only on the dimension of eigenspace used. As enumerated in Table 5, alignment with 20 dimensional eigenspace takes only 0.24 seconds, but the time increases to 3.25 seconds if 100 dimensional eigenspace is used, using our C++ implementation. Fortunately, our automatic alignment method can surpass manually eye-alignment with 20 dimensional TI-eigenspace, and thus the computational cost is acceptable for most applications, even for some real-time applications.

TABLE 5
The CPU Time Spent for TIPCA Training and Alignment

# Iterations	Offline TIPCA training time	# Dimension	Alignment time per test image
1	0.09 hours	20	**0.24 seconds**
5	0.56 hours	30	0.43 seconds
9	1.32 hours	40	0.68 seconds
13	2.42 hours	50	0.96 seconds
17	4.09 hours	70	1.65 seconds
23	8.74 hours	100	3.25 seconds

5 Conclusion

The experiments suggest a number of conclusions:

1) The proposed TIPCA technique is effective to automatically learn a set of eigenfaces that characterizes intrinsic structure of the faces from a large set of training images with various in-plane transformations. By removing the transform-related components, the MSE between the TIPCA-aligned images and their reconstructions is about 30 percent lower than that of the manually eye-aligned images.
2) There is a close relationship among alignment, representation, and recognition: Image alignment and eigenface representation mutually promote each other, which can eventually improve the image reconstruction and recognition performance.
3) State-of-the-art invariant descriptors and classification methods can benefit from using the TIPCA-aligned faces, instead of the eye-aligned faces, in the applications such as face recognition and gender classification.
4) The TI-eigenspace can define a subject-independent coordinate for face alignment. Provided that the number of training images are sufficiently large, TIPCA provides equivalently precise alignment for the images from seen (training) and unseen subjects.
5) A considerable amount of transform-related components exist in the eye-aligned face ensemble, even though the eye centers are manually located. The relatively high MSE, low SNR, low face recognition/gender classification accuracies suggests that the eye-aligned faces are far from optimal for face processing. Although these eye-aligned faces have been used by almost all the current studies on face coding, recognition, and classification as the *ground-truth alignment*, TIPCA based alignment can improve its performance to a large extent.

We should point out that TIPCA is shown to be effective only for the frontal faces with in-plane transformation. Current algorithm is likely to break down under out-of-plane pose changes, and so new transformation models are needed to support the algorithms presented in this paper. We are currently investigating the possibility of aligning and representing the 3D face using the methodology of TIPCA.

Acknowledgments

The authors would like to thank the anonymous reviewers for their thoughtful and constructive remarks that are helpful to improve the quality of this paper. This work was partially sponsored by National Natural Science Foundation of China (NSFC) under Grant No. 61375031, No. 61005025, No. 61002051, and No. 61273217. This work was also supported by the Fundamental Research Funds for the Central Universities, Beijing Higher Education Young Elite Teacher Project, and the Program for New Century Excellent Talents in University.

References

[1] L. Sirovich and M. Kiby, "Low-Dimensional Procedure for the Characterization of Human Face," *J. Optical Soc. of Am. A*, vol. 4, no. 03, pp. 519-524, 1987.
[2] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
[3] I. Craw, N. Costen, T. Kato, and S. Akamatsu, "How Should We Represent Faces for Automatic Recognition?" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 725-736, Aug. 1999.
[4] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward Practical Smile Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2106-2111, Nov. 2009.

[5] E. Makinen and R. Raisamo, "Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 541-547, Mar. 2008.
[6] J. Bekios-Calfa, J. Buenaposada, and L. Baumela, "Revisiting Linear Discriminant Techniques in Gender Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 858-864, Apr. 2011.
[7] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of Mis-Alignment in Face Recognition: Problem and a Novel Mis-Alignment Learning Solution," *Proc. IEEE Sixth Int'l Conf. Automatic Face and Gesture Recognition*, pp. 314-320, 2004.
[8] W. Deng, J. Guo, J. Hu, and H. Zhang, "Comment on '100% Accuracy in Automatic Face Recognition,'," *Science*, vol. 321, no. 5891, p. 912, 2008.
[9] H. Schweitzer, "Optimal Eigenfeature Selection by Optimal Image Registration," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 1999.
[10] S. Baker, I. Matthews, and J. Schneider, "Automatic Construction of Active Appearance Models as an Image Coding Problem," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1380-1384, Oct. 2004.
[11] S. Baker, R. Gross, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 3," Technical Report CMU-RI-TR-03-35, Robotics Inst., Carnegie Mellon Univ., 2003.
[12] T. Vetter, M. Jones, and T. Poggio, "A Bootstrapping Algorithm for Learning Linear Models of Object Classes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 40-46, 1997.
[13] R. Gross, I. Matthews, and S. Baker, "Generic vs. Person Specific Active Appearance Models," *Image and Vision Computing*, vol. 23, no. 12, pp. 1080-1093, 2005.
[14] P.J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
[15] D. Bolme, J. Ross Beveridge, M. Teixeira, and B. Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure," *Proc. Third Int'l Conf. Computer Vision Systems*, pp. 304-313, 2003.
[16] P. Viola and M. Jones, "Rapid Object Detection Using A Boosted Cascade of Simple Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. I-511-I-518, 2001.
[17] M. Meytlis and L. Sirovich, "On the Dimensionality of Face Space," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1262-1267, July 2007.
[18] T. Ahonen, A. Hadid, and M. Pietikinen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.
[19] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893, 2005.
[20] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467-476, Apr. 2002.
[21] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, Feb. 2009.
[22] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864-1870, Sept. 2012.
[23] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372-386, Feb. 2012.
[24] T. Cootes, G. Edwards, and C. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681-685, June 2001.
[25] Z. Zhou and X. Geng, "Projection Functions for Eye Detection," *Pattern Recognition*, vol. 37, no. 5, pp. 1049-1056, 2004.
[26] C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1-27:27, 2011.
[27] E. Grosso, A. Lagorio, L. Pulina, and M. Tistarelli, "Understanding Critical Factors in Appearance-Based Gender Categorization," *Proc. European Conf. Computer Vision Workshops and Demonstrations*, pp. 280-289, 2012.