

A Fully Automated Face Recognition Using Transform-Invariant PCA

Antony Gnana Anusiya.S¹, Sujithra Jenifer.M², Muthumariammal.S³

Department of Information Technology, Francis Xavier Engineering College, Tirunelveli, India¹

Department of Information Technology, Francis Xavier Engineering College, Tirunelveli, India²

Assistant Professor, Department of Information Technology, Francis Xavier Engineering College, Tirunelveli, India³

Abstract: Transform-invariant PCA (TIPCA) technique goal is to accurately transform the human's natural face for analysis of training image. Normal alignment is different from TIPCA alignment which generates the favourable Eigen space. The objective of the Eigen space is to minimize the error of mean square error between the aligned images and their reconstructions. The FERET image validates the mutual promotion between image alignment and Eigen space representation. This makes the optimized coding and recognition performance. Experimental results also suggest many spatial and frequency domain methods can benefit from using the TIPCA-aligned faces, instead of the manually eye-aligned faces. Favourable accuracies against the state-of-the-art results on face coding and face recognition are reported.

Index Terms: Face alignment, face representation, face coding, face recognition, Eigen faces, principal component analysis

I. 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 covariance matrix of face images [1]. In this context, Turk and Pentland [2] developed a well-known Eigen faces method, where the Eigen faces 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 Eigen faces 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 Eigen face method possibly produces severely blurred components that mostly account for the transformations and ignore the more interesting and useful structure. To address this problem, Eigen face 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 centres. 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 Eigen face 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 Eigen space 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 Eigen space of the aligned image ensemble using PCA; and 2) aligning each image to the Eigen space using simultaneous inverse compositional algorithm. The resulting TI-Eigen space defines a unified coordinate system for various applications on face alignment, representation, and recognition.

II. STUDY OF TI-EIGEN SPACE

A. Eigen faces by PCA

Eigen faces 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 $\phi_j \in \mathbb{R}^d$

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

where μ is the average image, ϕ_1, \dots, ϕ_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 $a = (a_1, \dots, a_m)^T$ 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

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

In more detail, given a set of N example training images: I^i where $i = 1, 2, \dots, N$, the formulation of Eigen faces is based on a general principle that the mean square error between input patterns and their reconstructions is minimized. While undoubtedly successful in appearance based recognition, the theoretical foundation for the use of Eigen faces 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 Eigen face 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.

B. Transform-Invariant PCA

For the clearness of the formulation, we represent the (unaligned) training images $I^i(x)$ and the basis images $\phi_j(x)$ in the pixel form, where $x = (x, y)^T$ is a column vector containing the pixel coordinates, rather than the vector form. Let $W(x; p)$ denote the parameterized set of possible transformations, where $p = (p_1, \dots, p_n)^T$ 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.

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

Where the warp $W(x; p)$ takes the pixel x in the basis image $\phi_j(x)$ and maps it to the sub-pixel location $W(x; p)$ in the image I . Given a set of unaligned facial images $\{I^i\}_{i=1}^N$, we assume that the transformed images, denoted by $I^i(W(x; p^i))$, 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 Eigen faces are learned based on a modified principle that minimizes the mean square error between transformed patterns and their reconstructions.

$$\arg \min_{\mu, \phi_j} \frac{1}{N} \sum_{i=1}^N \left\{ \min_{p^i, a^i} \sum_x [e^i(x)]^2 \right\}, \quad (4)$$

$$e^i(x) = I^i(W(x; p^i)) - \left[\mu(x) + \sum_{j=1}^m a_j \phi_j(x) \right],$$

As the introduction of the transform parameter 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 $\{p^i, a^i\}$ in turn, assuming where necessary that estimates of the others are available. The training of TIPCA is initialized by the "coarse" Eigen space derived by applying standard PCA on the detected faces, and then starts to learn the transform invariant Eigen space by alternately conducting the two following steps:

Step 1) Eigen space based Alignment, i.e., fix $\mu, \{\phi_j\}_{j=1}^d$ and optimize $\{p^i\}_{i=1}^N$. Given m and $\{\phi_j\}_{j=1}^d$ that define a Eigen space, we use the simultaneous inverse compositional (SIC)

algorithm1 to optimize $\{p^i, a^i\}$ 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 p^i and the coding parameters a^i . Let q^i be the concatenated parameter vector of p^i and a^i , and the Jacobian (steepest descent) images of (4) is

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

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

$$\Delta q = -[\sum_x J^T(x) J(x)]^{-1} \sum_x J^T(x) e^i(x) \quad (6)$$

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

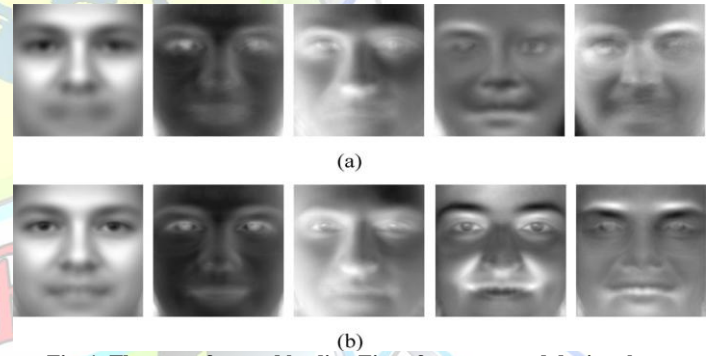


Fig. 1. The mean face and leading Eigen face computed during the learning process of TIPCA. (a) those of the initialization. (b) those of the second iteration. Interestingly, the alternating optimization seems to "deblur" the basis images, which suggests that the alignment step is effective to reduce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

Step 2) Eigen space Update, i.e., fixes $\{p^i\}_{i=1}^N$ and optimize $\mu, \{\phi_j\}_{j=1}^d$. If p^i is known, we can compute the transform $W(x; p^i)$ 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 $I^i(W(x; p^i))$, stack it as a vector, and then perform PCA on these vectors, update m to be the mean vector of the aligned ensemble, and $\phi_j, j = 1, \dots, d$ to be the eigenvectors of the covariance matrix with the d largest Eigen value. 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.

C. Complexity Control

The iterative “image alignment–Eigen space update” procedure guarantees that the MSE can be reduced to be a local minimum. 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 Eigen space for alignment.

An 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 Eigen space cannot characterize enough variation of the image appearance that ensures the alignment algorithm to be converged. On the other hand, high dimensional Eigen space 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” Eigen space at initial iterations, which ensures the convergence 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 Eigen space exclude the blur components, and thus allows the next alignment step to select higher dimensional “deblurred” Eigen space, which in turn benefit precise alignment. In summary, as the coarse-to-fine Eigen space is used for alignment, the mutual promotion of alignment and Eigen face 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 Eigen face and without resorting to more complex models, such as active appearance model [10] and morph able 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. This geometric information (the relationship between facial features) is 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.

III. APPLICATIONS OF TIPCA

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 Eigen face. Taking the former as the final result, TIPCA can be regarded as a approach to batch image alignment. More importantly, the set of TI-Eigen face, which define a TI-Eigen space, provides an invariant appearance model leading to broad applications. This section details how the TI-Eigen space can be applied to align, encode, and recognize the unseen images.

A. TIPCA-Based Image Alignment

Image alignment aims to align a facial image, typically the out-put of the face detector, to the transform-invariant Eigen space defined by the Eigen face corresponding to the top m Eigen value. 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 p and the appearance parameter a by solving following optimization problem:

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

The complexity of the alignment algorithm increases dramatically with a large m , but, fortunately, low dimensional TI-Eigen space, 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-Eigen space to make them comparable within an unified coordinate.

B. TIPCA-Based Fully Automatic Face Recognition

By combining the TIPCA-based image alignment and feature extraction, a fully automatic Eigen face based recognition algorithm can be readily figured out, as illustrated in Fig. 2. In the training stage, a TI-Eigen space 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 are first extracted. Finally, the nearest neighbour classifier is used for classification. In our experiments, the distances between two arbitrary feature vectors, $a^i, a^j \in \mathbb{R}^d$ used in our experiment are defined as follows:



Fig. 2. 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-Eigen space

$$\begin{aligned} d_{Ed} &= (a^i - a^j)^T (a^i - a^j) \\ d_{Md} &= (a^i - a^j)^T \Sigma^{-1} (a^i - a^j) \\ d_{Wc} &= \frac{(a^i - a^j)^T \Sigma^{-1} (a^i - a^j)}{\|a^i\| \|a^j\|} \end{aligned} \quad (8)$$

where $\Sigma \in IR^{d \times d}$ is the covariance matrix of the training data. For the de-correlated principal components, S is diagonal and the diagonal elements are the (Eigen values) variance of the corresponding components. Ed , Md , Wc defines the Euclidean distance, Mahalanobis distance, whitened cosine distance, respectively.

Beyond the direct matching of the Eigen face codes, TIPCA can benefit various recognition methods via precise image alignment. By automatically learning the TI-Eigen face from the training ensemble, and aligning both the gallery and the probe images to a unified Eigen space defined by TI-Eigen face, 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 Eigen face based approaches, and makes them operated in a fully automatic way. Some applications will be demonstrated in the following experiment section.

IV. 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 test bed 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.

To evaluate the generalization ability of TIPCA, in each iteration, we align the probe images using the identical Eigen space that used in the alignment step, then reconstruct the aligned probe images by the first 100 Eigen face (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 decreases as the algorithm iterates. These results indicate the TIPCA has improved generalization ability to represent facial images than traditional Eigen face based approaches.

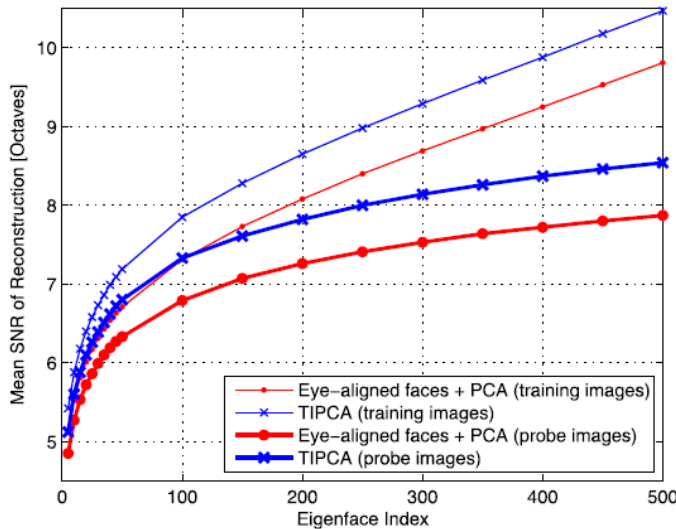


Fig. 3. The average signal-to-noise ratio as a function of number of Eigen face used for reconstruction.

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-Eigen face. The reconstructed images become clearer as the number of Eigen face is increased. For comparison, Fig. 6a shows the PCA based reconstruction on the same (eye-aligned) probe image, where the same number of Eigen face, 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 reasons: 1) the traditional Eigen face 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 Eigen space. 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.

A. 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, Eigen face are constructed from a train-ing set of face images and particular probe faces are recognized by comparing the principal components (Eigen face 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 reduce performance [17]. Our empirical results validated that optimal recognition performance is achieved with a dimensionality roughly 150, using nearest-neighbour 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 Eigen space 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

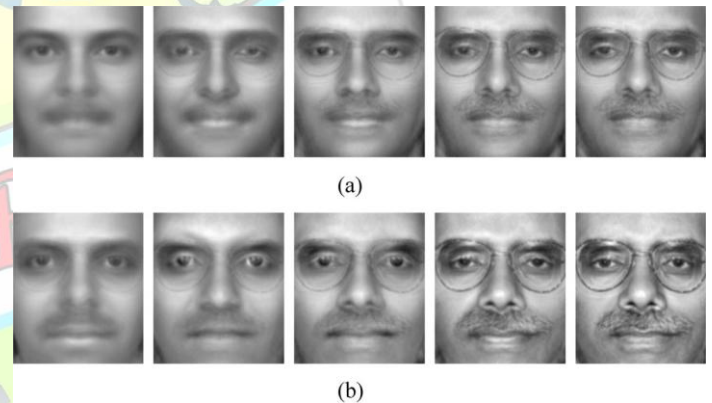


Fig. 4. Some reconstructed images based on (a) PCA and (b) TIPCA with the dimension $d=20; 40; 60; 80; 100$. Note that reconstructed region of PCA is manually defined by two eye centres, but that of TIPCA is automatically selected from the detected face.

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 Eigen face 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.

By observing the finally optimized performance with TI-Eigen space learned from 23 iterations, we find that the imaginary part which are squared and added to obtain an estimate of energy at a particular location and frequency band. The response of each filter is down sampled 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.

Leave-Out Test

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-Eigen space. However, in the large-scale face recognition/retrieval applications, it is difficult to collect the entire 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 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-Eigen space 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/retrieval applications. It is possible to build a subject-independent TI-Eigen space by which generic facial images can be efficiently and precisely aligned for accurate recognition.

B. Sparse Representation Face Recognition

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

performance differences using different Eigen space dimensions for alignment is not significant. This suggest that TIPCA can be applied in an efficient way using low dimensional Eigen space for alignment, while keeping highly accurate recognition performance.

TABLE 2

Comparative FERET dup2 Recognition Rates with/without the Involved subjects for TIPCA training

TIPCA #l=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 3

Comparative FERET Recognition 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

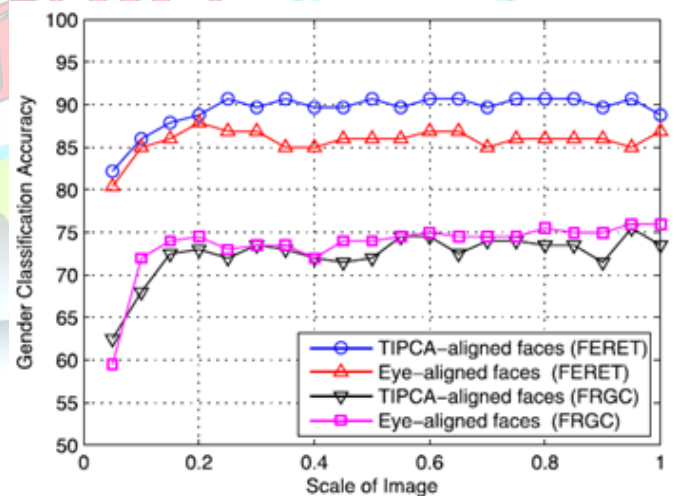


Fig. 5. The gender classification rate of SVM as a function of image resolution using three alignment method

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-Eigen space ($\#I = 23$, $\#D = 40$) 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.

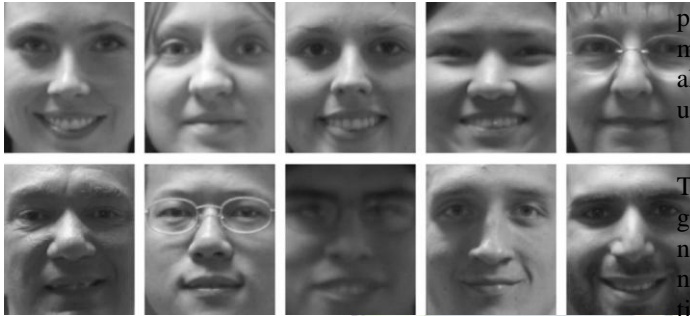


Fig. 6. The TIPCA-aligned facial images of FRGC database for gender classification test. Note that the TI-Eigen space learned from FERET database is used for alignment.

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].

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

C. 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 labelling), and scalable to a huge number of training images by parallelized training. Because over 99 per-cent 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 Eigen space, and duplicates the updated Eigen space on other machines.

The applications of TIPCA are efficient. Because TIPCA builds a unified TI-Eigen space 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 Eigen space used. As enumerated in Table 5, alignment with 20 dimensional Eigen space takes only 0.24 seconds, but the time increases to 3.25 seconds if 100 dimensional Eigen space is used, using our C++ implementation. Fortunately, our automatic alignment method can surpass manually eye-alignment with 20 dimensional TI-Eigen space, and thus the computational cost is acceptable for most applications, even for some real-time applications.

V. CONCLUSION

The experiments suggest a number of conclusions:

- 1) The proposed TIPCA technique is effective to automatically learn a set of Eigen face 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 Eigen face 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-Eigen space can define a subject-independent coordinate for face alignment. Provided that the



numbers 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 centres 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.

REFERENCES

1. L. Sirovich and M. Kirby, "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, "Eigen face 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 Eigen feature 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.