Face Verification With Balanced Thresholds

Shuicheng Yan, *Member, IEEE*, Dong Xu, and Xiaoou Tang, *Senior Member, IEEE*

Abstract—The process of face verification is guided by a prelearned global threshold, which, however, is often inconsistent with class-specific optimal thresholds. It is, hence, beneficial to pursue a balance of the class-specific thresholds in the model-learning stage. In this paper, we present a new dimensionality reduction algorithm tailored to the verification task that ensures threshold balance. This is achieved by the following aspects. First, feasibility is guaranteed by employing an affine transformation matrix, instead of the conventional projection matrix, for dimensionality reduction, and, hence, we call the proposed algorithm *threshold balanced transformation* (TBT). Then, the affine transformation matrix, constrained as the product of an orthogonal matrix and a diagonal matrix, is optimized to improve the threshold balance and classification capability in an iterative manner. Unlike most algorithms for face verification which are directly transplanted from face identification literature, TBT is specifically designed for face verification and clarifies the intrinsic distinction between these two tasks. Experiments on three benchmark face databases demonstrate that TBT significantly outperforms the *state-of-the-art* subspace techniques for face verification.

Index Terms—Dimensionality reduction, face verification, subspace learning, threshold balance.

I. INTRODUCTION

IN biometrics, there exists a clear distinction between *identification* and *verification* tasks. For identification, a probe sample is presented and the system identifies which subject the sample belongs to by computing the similarities between the probe sample and all gallery samples in the database. For verification, a probe sample is presented and the system tries to find out whether the sample belongs to the subject he/she claims. Only the similarities between the probe sample and those from the gallery belonging to the claimed subject are compared, and the final decision is made based on a prelearned threshold. A large family of algorithms have been proposed for the face identification task [1], [7], [15], [16], [18], [20], [21], [24]. Also, many approaches [2], [6], [14] were presented for face verification; however, most of them [4], [9] are directly transplanted from the face identification literature and ignore the intrinsic distinction between these two tasks. Although there were some attempts [8], [17] to distinguish face verification from the identification task, extra postprocessing or online learning is required,

which requires an impractical expense in both computation and storage for online learning in real applications.

We observe that face verification in general is directly determined by a single prelearned global threshold; however, each specific class has its own specific optimal threshold, namely a client-specific threshold, for verification. In order to obtain maximum face verification accuracy, the specific optimal threshold for each class should be equal to each other; otherwise, the error rejection rate or error accept rate for a specific class will increase and consequently lead to degradation in global verification accuracy. However, algorithms transplanted from the face identification literature cannot guarantee a balance of the class specific thresholds.

Inspired by the above observations, we present a new subspace learning algorithm for face verification by balancing the class specific thresholds. As opposed to conventional subspace learning algorithms which applied the column-unitary and column-independent projection matrix for subspace learning, a more general affine transformation matrix is employed for dimensionality reduction in this work. We will show that this relaxation will allow us to balance the class specific thresholds as well as enhance the classification performance. Moreover, with the consideration that the right orthogonal matrix of singular value decomposition of a transformation matrix does not affect the similarity measure if based on Euclidean distance, the affine transformation matrix can be constrained to be the product of an orthogonal matrix and a diagonal matrix.

The affine transformation matrix is learned in an iterative manner. First, the transformation matrix T is initialized; specifically we may initialize it with a learned projection matrix from a conventional algorithm like Eigenfaces [20] or Fisherfaces [1], and then reset $T = U\Lambda$ provided that its singular value decomposition is $T = U\Lambda V'$. Then, the diagonal matrix is optimized to improve the balance of the class specific thresholds by solving a quadratic programming problem with linear constraints. Finally, T is optimized to pursue the compactness of the samples belonging to the same subject with the subject centers approximately fixed. The latter two steps iteratively optimize the threshold balance and classification capability until convergence. The major aspects of our proposed algorithm, referred to as *threshold balanced transformation* (TBT), are highlighted as follows.

1) TBT guarantees a balance of class specific thresholds without sacrificing any classification capability, which leads to better performance over traditional subspace learning algorithms for the verification task.
2) TBT is also of a potential advantage for face identification, since the learned affine transformation matrix can balance the margins between nearby classes; consequently, the distance between neighboring classes will be enlarged to a moderate scale.

S. Yan and X. Tang are with Microsoft Research Asia, Beijing 100080, China, and also with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: scyan@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

D. Xu is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: dongxu@ee.columbia.edu).

3) The approach of the *client-specific threshold* [5] is also a possible way to improve verification accuracy, but as analyzed later in this work, it is computationally impractical and not sufficiently robust. The core idea of TBT is to pursue a balance of class specific thresholds; its benefit to the verification task can be generalized to other algorithms for face verification, and it is a natural alternative to the idea of a *client-specific threshold* for the verification task.

The rest of the paper is organized as follows. In Section II, we discuss the face identification and verification tasks, which directly motivate the TBT algorithm for face verification, as described in Section III. Then, in Section IV, we report experimental results that compare face verification and identification techniques. Finally, we conclude this paper with a discussion of future work in Section V.

II. IDENTIFICATION VERSUS VERIFICATION TASK

In the notation of this work, matrices are represented as uppercase letters, such as $P, T, \ldots$; italic lowercase letters such as $x, y, z, \ldots$ denote vectors, and normal lower-case letters such as $a, b, c, \ldots$ denote scalar quantities. For a classical learning problem, the training sample set is represented as $\{x_i \in \mathbb{R}^d, i = 1, \ldots, N\}$, where N is the total number of samples. The sample x_i belongs to the subject indexed as $c_i \in \{1, 2, \ldots, N_c\}$, where N_c is the subject number. There are n_c samples belonging to the cth subject in the training set.

The key distinction between the identification and verification tasks can be described by the questions answered by the systems. For identification, the question "*Who is this?*" is asked, and the system determines the subject identity by comparing the similarities or distances between the provided image x and *all gallery images* in the database, i.e.,

$$i^* = \arg \min_{i=1,\ldots,N} D(x, x_i) \tag{1}$$

and then the subject is identified to be c_{i*}. Here, $D(x, x_i)$ is the distance between the new image x and the sample x_i. For some specific applications, a threshold is needed to decide whether the probe object belongs to the sample subject set.

For verification, the question "*Is this person whom he/she claims to be?*" is asked and the system attempts to verify the provided identity of c by computing the similarities or distances between the presented image x and the *gallery images of the claimed subject*. A decision is made based on a prelearned threshold θ, that is

$$\begin{cases} \text{Accept}, & \text{if } \min_{c_i=c} D(x, x_i) < \theta \\ \text{Reject}, & \text{else.} \end{cases} \tag{2}$$

Face verification is fundamentally different from the face identification task. However, most previous algorithms for face verification have ignored the intrinsic characteristics of the face verification task and were directly transplanted from the face identification literature. Hence, these algorithms are not necessarily optimal for face verification, and there exists additional information that can be applied to further improve face verification accuracy.

In face verification, commonly only one global threshold is used for determining the subject identity; however, each class
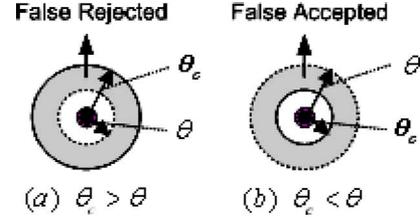

Fig. 1. Examples in a 2-D space to illustrate why threshold balance is necessary for optimal verification: (a) when the class specific optimal threshold θ_c is larger than the global threshold θ, the samples in the darker area will be wrongly rejected, and (b) when the class specific optimal threshold is smaller than the global one, the samples in the darker area will be falsely accepted. Note that the solid-line circles denote the gallery samples belonging to the cth class according to the class specific threshold θ_c, while the dashed circles denote the data area classified as the cth class according to the global threshold θ.

has its own specific optimal threshold for face verification. Let θ denote the global optimal threshold and θ_c denote the class specific optimal threshold of the cth class. As demonstrated in Fig. 1, when the class specific optimal threshold θ_c is inconsistent with the global optimal threshold θ, the *error rejection rate* or *error acceptance rate* will increase, and, consequently, the verification accuracy will be degraded; therefore, it is desirable to design a face verification algorithm that can provide balanced class-specific optimal thresholds for different classes. In the following, we introduce our solution to this problem.

III. THRESHOLD BALANCED AFFINE TRANSFORMATION FOR FACE VERIFICATION

A large number of algorithms have been proposed for the face verification task; aside from techniques based on hidden Markov models (HMMs), the Bayesian method, support vector machines (SVMs), and neural networks (NNs) [13], [14], subspace learning techniques such as linear discriminant analysis (LDA) are the most popular. Motivated by the benefit to face verification of balancing class specific thresholds, in this work we develop a new subspace learning algorithm that can extract effective features for the verification task and at the same time ensure the balance of the class specific thresholds.

In general, for a subspace learning algorithm, a matrix is learned to transform the original high-dimensional feature to a low-dimensional one. For ease of understanding, we clarify the difference between a *projection matrix* and an *affine transformation matrix*. A projection matrix consists of linearly independent and unitary column vectors, while an affine transformation matrix has no unitary constraint on its column vectors.

In the following, we first study the advantage of the affine transformation matrix over the projection matrix for the face verification task.

A. Affine Transformation Matrix for Dimensionality Reduction

Most previous subspace learning algorithms assume that there exists a projection matrix to map the original high-dimensional feature to the desired low-dimensional one. For example, LDA seeks the directions which maximize the interclass scatter and at the same time minimize the intraclass scatter. Let the projection matrix be $P \in \mathbb{R}^{d \times m}$, where m is the dimension of the low-dimensional feature space, then the solution matrix P optimizes the following objective function:

$$P^* = \arg\max_P \left\{ \frac{\sum_{c=1}^{N_c} n_c \|P'\overline{x}_c - P'\overline{x}\|^2}{\sum_{i=1}^{N} \|P'x_i - P'\overline{x}_{c_i}\|^2} = \frac{Tr(P'S_bP)}{Tr(P'S_wP)} \right\}$$

$$S_b = \sum_{c=1}^{N_c} n_c(\overline{x}_c - \overline{x})(\overline{x}_c - \overline{x})', \quad S_w = \sum_{i=1}^{N}(x_i - \overline{x}_{c_i})(x_i - \overline{x}_{c_i})' \tag{3}$$

where $\overline{x}_c$ is the mean of the cth class and $\overline{x}$ is the mean of all samples; $Tr(\cdot)$ denotes the trace of a matrix and P' is the transpose of matrix P. Commonly there is no closed form solution for (3), and instead this problem is often transformed to maximizing the objective function $Tr[(P'S_wP)^{-1}(P'S_bP)]$ and is solved using the generalized eigenvalue decomposition method

$$S_b p_i = \lambda_i S_w p_i, \quad i = 1, 2, \ldots, m \tag{4}$$

where p_i is the eigenvector corresponding to the ith largest eigenvalue λ_i of the generalized eigenvalue decomposition problem; meanwhile p_i constitutes the ith column vector of the projection matrix P.

The learned projection matrices from traditional algorithms like LDA cannot guarantee that the class specific optimal thresholds are balanced. Affine transformation matrices, however, have the potential to balance these thresholds. Intuitively, as shown in Fig. 2, although the class specific thresholds for different classes based on the initial low-dimensional representations are unbalanced, they can be easily balanced by properly scaling the vertical coordinate. In theory, provided that we have obtained a low-dimensional feature space with balanced class-specific thresholds for the first c classes, for the $(c+1)$th class, it is possible to find another projection direction perpendicular to the space spanned by the first c class centers. This new projection direction then only affects the metrics related to the $(c + 1)$th class; hence, we can properly scale the projection direction to obtain the final feature space with balanced class-specific thresholds for all of the $(c + 1)$ classes. This analysis motivates us to achieve a balance of class specific thresholds by utilizing an affine transformation matrix instead of a projection matrix for dimensionality reduction. Moreover, the distance between nearby classes will be balanced and the narrow margins will be enlarged to a moderate scale; thus, the identification capability can also be improved based on the threshold balanced representations.

Let the affine transformation matrix be $T \in \mathbb{R}^{d \times m}$, $m < d$, and its singular value decomposition is

$$T = \widetilde{U}\widetilde{\Lambda}\widetilde{V}' \tag{5}$$

where the matrix $\widetilde{U} \in \mathbb{R}^{d \times d}$ and $\widetilde{V} \in \mathbb{R}^{m \times m}$ are both orthogonal matrices, and $\widetilde{\Lambda} = [\text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_m\}, 0, \ldots, 0]$ is a diagonal matrix. The similarity measure based on Euclidean distance is invariant to the matrix $\widetilde{V}$; therefore, the affine transformation matrix T can be simply constrained without sacrificing any accuracy to be

$$T = U\Lambda. \tag{6}$$

(a) Representations from projection matrix

(b) Representations from affine tranformation matrix
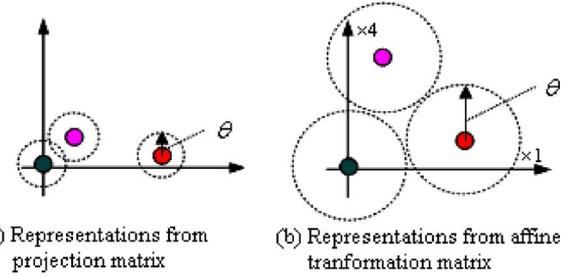
Fig. 2. Toy problem to show that the affine transformation matrix is potentially superior to the projection matrix for the verification task: (a) sample distribution from a projection matrix and (b) sample distribution from an affine transformation matrix by reweighing the vertical coordinate. Note that the solid circles denote three different gallery samples belonging to three different classes; and the radius of the dashed line circles are the learned global optimal thresholds from the projection matrix in (a) and the affine transformation matrix in (b), respectively. The symbol "× 4" denotes the vertical coordinate in (b) is that of (a) multiplied by 4. It shows that although the class specific thresholds are unbalanced in (a), after properly reweighing the features in (b), the threshold balance is significantly improved.

Here, matrix U is the first m columns of $\widetilde{U}$, with diagonal matrix $\Lambda = \text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_m\}$.

B. Procedure to Learn Threshold Balanced Affine Transformation Matrix

The affine transformation matrix has been shown to have the potential to present balanced class-specific thresholds for the face verification task. In this section, we introduce how to learn an affine transformation matrix effective for both balancing thresholds and improving classification capability.[1] The whole learning procedure consists of three steps: initialization, threshold balance optimization, and classification capability optimization. For the learning procedure, the objectives are two-fold: high classification power and good threshold balance. These two objectives lead to good verification capability. The detailed procedure to iteratively improve one objective while sufficiently retaining the other is as follows.

1) Initialization: The affine transformation matrix T can be initialized arbitrarily and is only required to be columnly independent. To speed up the algorithm, we can utilize the traditional subspace learning algorithm principal component analysis (PCA)+LDA [1] to initialize the matrix T. First, the original image feature is projected to a proper low-dimensional space with matrix P_{pca} derived from PCA, and then LDA as in (3) is applied to further reduce the feature dimension with the projection matrix P_{lda}. Finally, the transformation matrix T is initialized as

$$P_{\text{pca}} P_{\text{lda}} = \widetilde{U}\widetilde{\Lambda}\widetilde{V}'; \quad T = U\Lambda. \tag{7}$$

The relations between $\widetilde{U}, \widetilde{\Lambda}, U$, and $\Lambda = \text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_m\}$ are the same as in (5) and (6).

2) Threshold Balance Optimization: For each class, the class specific optimal threshold is approximately characterized by the distances between the class center and its nearest class centers. Thus, to ensure that the class specific thresholds are balanced,

---

[1]Note that DLDA [23] also utilizes an affine transformation matrix for dimensionality reduction, but it cannot guarantee a low-dimensional representation with balanced thresholds for verification.

it is sufficient to balance the average distances from each class center to its k-nearest class centers. In this paper, we experimentally set $k = 3$.

To this end, for the cth class, we first find its k-nearest class centers, and the corresponding class indices are denoted as the set S_c^k. Then, the objective function to be minimized for threshold balance is as follows:

$$\sum_{c \neq c^*} \left(\sum_{o \in S_c^k} \|T'\overline{x}_c - T'\overline{x}_o\|^2 - \sum_{o \in S_{c^*}^k} \|T'\overline{x}_{c^*} - T'\overline{x}_o\|^2 \right)^2. \tag{8}$$

A direct solution of this objective function is not possible. As analyzed in Section III-A, however, the threshold balance can be substantially improved by reweighing the projection vectors. Moreover, when pursuing threshold balance, we need to sufficiently retain the classification capability. Therefore, we optimize the objective function (8) by keeping the matrix U fixed since it encodes the most discriminating information of the original features, and the diagonal matrix is optimized to pursue the threshold balance. Then, the objective function (8) is simplified to

$$\sum_{c \neq c^*} \left(\sum_{o \in S_c^k} \|\Lambda \overline{x}_c^U - \Lambda \overline{x}_o^U\|^2 - \sum_{o \in S_{c^*}^k} \|\Lambda \overline{x}_{c^*}^U - \Lambda \overline{x}_o^U\|^2 \right)^2 \tag{9}$$

where $\overline{x}_c^U = U'\overline{x}_c$. To avoid the trivial solution $\Lambda = 0$, we impose the constraint that $\sum_i^m \gamma_i^2 = 1$; meanwhile, in order to ensure that all the selected features contribute to final verification, we impose another constraint that $\gamma_i > \epsilon > 0$. Then, the objective function is constrained by

$$\sum_{i=1}^m \gamma_i^2 = 1; \quad \gamma_i > \epsilon > 0, \quad i = 1, 2, \ldots, m. \tag{10}$$

Here, $\epsilon (= 0.1/\sqrt{m}$ in this work) is the minimal value for each parameter, which ensures that each parameter will be above zero.

Let the vector z represent $z = [z_1, z_2, \ldots, z_m]'$ where $z_i = \gamma_i^2$, and let y_c^o be an m-dimensional vector with $y_c^o(j) = (\overline{x}_c^U(j) - \overline{x}_o^U(j))^2, j = 1, 2, \ldots, m$. Then, the objective function (9) can be expressed as

$$\sum_{c \neq c^*} \left(\sum_{o \in S_c^k} \|\Lambda \overline{x}_c^U - \Lambda \overline{x}_o^U\|^2 - \sum_{o \in S_{c^*}^k} \|\Lambda \overline{x}_{c^*}^U - \Lambda \overline{x}_o^U\|^2 \right)^2$$

$$= \sum_{c \neq c^*} \left(\sum_{o \in S_c^k} z'y_c^o - \sum_{o \in S_{c^*}^k} z'y_{c^*}^o \right)^2 = z'Az$$

where

$$A = \sum_{c \neq c^*} \left(\sum_{o \in S_c^k} y_c^o - \sum_{o \in S_{c^*}^k} y_{c^*}^o \right) \left(\sum_{o \in S_c^k} y_c^o - \sum_{o \in S_{c^*}^k} y_{c^*}^o \right)'.$$

Then, the optimization problem defined in (9) and (10) is changed to a constrained quadratic optimization problem

$$\min_z z'Az, \quad s.t. \sum_{i=1}^m z_i = 1, z_i \geq \epsilon^2, i = 1, 2, \ldots, m. \tag{11}$$

This problem can be easily solved by using the optimization toolbox in Matlab (such as with the function *quadprog* in Matlab 7.0).

3) Classification Capability Optimization: To further improve the classification capability while sufficiently maintaining balanced thresholds, we propose to pursue an affine transformation matrix that leads to compact distributions of class samples around their class centers while approximately retaining the threshold balance property. Then, the objective function to be optimized for classification is simplified to be the denominator of the objective function in (3). Denote the transformed class centers from the transformation matrix T obtained in Section III-B2 as

$$m_c = \frac{1}{n_c} \sum_{c_i = c} T'x_i. \tag{12}$$

Then, the corresponding objective function is

$$T^* = \arg\min_T \left\{ F(T) = Tr(T'S_wT) = \sum_{i=1}^N \|T'x_i - m_{c_i}\|^2 \right\}. \tag{13}$$

Denoting $X = [x_1, x_2, \ldots, x_N]$ and $M = [m_{c_1}, m_{c_2}, \ldots, m_{c_N}]$, the optimal affine transformation matrix T can be obtained by setting the derivative of the objective function to zero

$$\frac{\partial F(T)}{\partial T} = XX'T - XM' = 0. \tag{14}$$

Therefore, the optimal T is

$$T = (XX')^{-1}XM'. \tag{15}$$

When the matrix XX' is not of full rank, the inverse matrix can be replaced with the pseudo-inverse matrix. Finally, the transformation matrix T is reset to be the product of a columnly orthogonal matrix and a diagonal matrix by using singular value decomposition as in (5) and (6).

After the initialization step, the second and third steps are iterated to enhance the threshold balance and classification capability until reaching convergence or a user-defined number of iterations. Due to the rotation invariance of the transformation matrix, the algorithm is considered to converge if $\|T_tT_t' - T_{t-1}T_{t-1}'\| < \varepsilon$ (set to be 0.1 in this work), where T_t denotes the derived transformation matrix from the tth step. Finally, the learned affine transformation matrix T is used for dimensionality reduction and face verification as in (2) with distance function

$$D(x, x_i) = \|T'x - T'x_i\|. \tag{16}$$
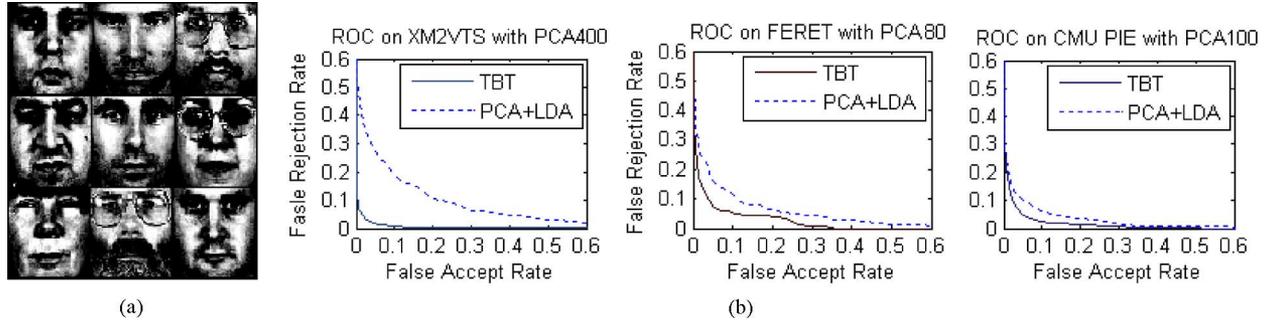
Fig. 3. (a) Sample images after image preprocessing. (b) The ROC curves of TBT compared with PCA+LDA on databases XM2VTS, FERET, and CMU PIE. Note that we omit the values in the range (0.6, 1.0) for both false rejection rate and false acceptance rate for greater clarity of the graphs.

C. Discussion

In this section, we discuss some aspects of the face verification task in relation to our proposed algorithm for learning a TBT matrix.

1) Why Not Directly Train a Classifier for Each Class to Avoid the Threshold Balance Issue: It is obviously beneficial to learn a class-specific classifier if we have enough samples for each class. However, it is often the case that the available number of samples for each class is small, and the learning problem for each class will be extremely unbalanced; thus, such an approach is generally not robust enough. Our proposed algorithm is a better choice which satisfies the requirements of both threshold balance and algorithmic feasibility.

2) What are the Differences Between the Approaches of "Client-Specific Threshold" and "Threshold Balance": There have been many algorithms [5] that use the client-specific threshold for verification, yet they need to learn different thresholds for different subjects, which is computationally impractical; moreover, as discussed in the above question, the solution is not robust due to the imbalance between the numbers of positive and negative samples. Hence, although "client-specific threshold" is a possible way to improve verification accuracy, our proposed method of "threshold balance" has the advantages of feasibility. Moreover, it is much easier for a user to tune a threshold for computing the false acceptance rate with a given false rejection rate compared to the method of "client-specific threshold."

3) Complexity and Convergence Analysis: For TBT, the complexities of both initialization and classification capability optimization steps are $O(d^3)$ when the sample number and feature number is comparable; and for the threshold balance optimization step, it is very fast since $m \ll d$ in general. Therefore, the complexity of the whole algorithm is $T_{max}O(d^3)$ where T_{max} is the iteration number. The latter two steps optimize the transformation matrix iteratively, and in our experiments, we find that commonly the solution will converge after about $T_{max} = 5$ loops.

IV. EXPERIMENTS

In this section, we present a set of face verification and identification experiments on the popular face databases XM2VTS [10], FERET [11], [12], and CMU PIE [19] to demonstrate the effectiveness of our proposed algorithm in comparison with the most popular subspace learning algorithms for face verification, namely Eigenfaces [20] and Fisherfaces/(PCA+LDA) [1]. We also implement a simplified noniterative version of TBT without the third step, denoted as S-TBT in the experiments. The performance is systematically evaluated on different combinations of PCA dimensions and LDA dimensions to compare with PCA+LDA. In all the experiments, the histogram equilibrium method is applied for image preprocessing, some samples of which are displayed in Fig. 3(a).

A. Data Preparation

There are standard protocols for face recognition on the XM2VTS [10] and FERET [11] databases. These two protocols are both designed for open-set evaluation, which means that the testing set may contain subjects that do not appear in the training set. Our proposed TBT and S-TBT are specifically developed for closed-set face verification due to the expectation of threshold balance. Hence, in this work, we design new experimental configurations, instead of the standard ones in [10] and [11], for algorithmic comparison and evaluation. The details are as follows.

The XM2VTS database contains 295 persons and each person has four frontal face images taken in four different sessions. All the images are aligned by fixing the locations of the two eyes and normalizing the image size to $64 * 64$ pixels. We use $295 * 3$ images from the first three sessions for model training; the gallery set consists of the 295 images from the first session; and the probe set is composed of the 295 images from the fourth session.

For the FERET database, we select the seventy subjects with six different images for each subject. Three images of each subject are randomly selected for model training, and one of the three images is randomly selected for the gallery set. Another three images of each subject are used as the probe set. All the images are aligned by fixing the locations of the two eyes and normalizing the size to $56 * 46$ pixels. The data set is randomly partitioned into gallery and probe sets. Three images of each person are randomly selected for training; one of the three training images is randomly selected for each subject as the gallery set; and the remaining three images are used for the probe set.

The CMU PIE (Pose, Illumination, and Expression) database contains more than 40,000 facial images of 68 persons. The images were acquired over different poses, under variable illumination conditions and with different facial expressions. In our

TABLE I
FACE VERIFICATION EER (%) AND IDENTIFICATION ACCURACY (%)
OF EIGENFACES (EI.), PCA+LDA (P+L), TBT, AND
S-TBT (S-T) ON THE XM2VTS DATABASE

Task	Task-1: verification				Task-2: identification			
Algorithm	Ei.	P+L	TBT	S-T	Ei.	P+L	TBT	S-T
P590L294	9.7	9.8	**3.0**	3.6	68.6	79.2	**97.0**	**97.0**
P400L294	9.7	15.3	**2.7**	5.1	68.5	69.2	**96.3**	90.2
P200L150	8.6	8.5	**3.4**	5.3	68.8	86.4	**92.5**	87.1

TABLE II
FACE VERIFICATION EER (%) AND IDENTIFICATION ACCURACY (%)
OF EIGENFACES, PCA+LDA, TBT, AND S-TBT ON FERET

Task	Task-1: verification				Task-2: identification			
Algorithm	Ei.	P+L	TBT	S-T	Ei.	P+L	TBT	S-T
P140L69	20.2	5.2	**4.6**	5.1	46.2	87.6	**91.0**	90.0
P120L69	19.9	5.2	**5.0**	6.1	45.7	87.1	**90.0**	84.3
P80 L69	19.3	10.7	**6.5**	11.7	45.2	71.0	**82.4**	63.8

TABLE III
FACE VERIFICATION EER (%) AND IDENTIFICATION ACCURACY (%)
OF EIGENFACES, PCA+LDA, TBT, AND S-TBT ON CMU PIE

Task	Task-1: verification				Task-2: identification			
Algorithm	Ei.	P+L	TBT	S-T	Ei.	P+L	TBT	S-T
P315L62	35.3	4.8	**4.4**	4.6	39.9	89.3	91.5	**91.7**
P200L62	35.3	5.6	**4.1**	5.1	39.9	88.4	**90.0**	89.3
P100L62	35.3	7.9	**5.0**	7.2	39.9	85.6	**88.6**	85.7

experiment, five near frontal poses (C27, C05, C29, C09, and C07) and illumination indexed as 08, 10, 11, and 13 are used. 63 of the 68 persons are used for data incompleteness experiments for the remaining five persons. Thus, each person has twenty images and all the images are aligned by fixing the locations of the two eyes and normalizing the image size to $64 * 64$ pixels. The data set is randomly partitioned into the gallery and probe sets. Six images of each person are randomly selected for model training. Different from the experiments on XM2VTS, we use the mean image of the training samples as the gallery image for each subject, and the remaining fourteen images are used for the probe set.

B. Face Verification and Identification

We systematically evaluate the performance of our proposed TBT algorithm in both face verification and identification tasks. Three sets of experiments on the above three databases are conducted. For face verification, the equal error rate (EER), i.e., the threshold with equal false acceptance rate (FAR) and false rejection rate (FRR) is used to test the verification accuracies of the different algorithms. For face identification, based on the learned projection matrix or affine transformation matrix, we map the image vector from the original feature space to the low-dimensional one, and then the nearest neighbor method is used for final classification. For all experiments, the PCA dimension and final dimension are the same for PCA+LDA, TBT, and S-TBT.

Tables I–III list detailed comparison results of Eigenfaces, PCA+LDA, TBT and S-TBT, and Fig. 3(b) plots the receiver operating characteristic (ROC) curves of the TBT algorithm compared to those of the PCA+LDA algorithm for the three databases. For each database, the algorithms were compared in three different configurations with PCA dimensions as $N - N_c$ and two other smaller numbers respectively. For simplicity, the experiments are denoted as PmLn where m is the dimension of PCA step and n is the dimension of LDA step.

C. Discussion

From the results listed in Tables I–III and Fig. 3(b), we can observe a number of interesting points.

1) For the face verification task, our proposed TBT algorithm consistently outperforms Eigenfaces and the PCA+LDA algorithm in all dimension combinations; and the EER is reduced up to more than 60%, in many cases, and the average EER reduction rate is more than 20%. As shown in Fig. 3, for any given false acceptance rate, the false rejection rate of TBT is consistently lower than that of PCA+LDA.
2) For the face identification task, TBT is comparable with PCA+LDA; and in most cases, TBT is significantly superior to the PCA+LDA algorithm, especially when the PCA dimension is not well selected.
3) For the face verification task, TBT is relatively stable to variations in PCA dimension, while the PCA+LDA algorithm appears to be very sensitive to the PCA dimension variation, which again validates the effectiveness of the threshold balance strategy.
4) S-TBT is comparable with PCA+LDA for both verification and identification tasks, and the extra third step of TBT can further improve algorithmic performance in most cases, especially for face verification.
5) In this work, we did not systematically compare TBT with many other algorithms such as SVM and the work in [21], since the focus of this work is on subspace learning algorithms; the core idea of TBT to pursue balanced class-specific thresholds can be easily generalized to other algorithms for face verification.

V. CONCLUSION AND FUTURE WORK

The work presented in this paper has given insights into the essential distinction between face verification and face identification tasks, and proposed a novel subspace learning algorithm, called TBT, tailored to the face verification task. It learns an affine transformation matrix, instead of a projection matrix as done conventionally, to pursue a balance of the class-specific thresholds. Extensive experiments demonstrate that our proposed algorithm brings encouraging accuracy improvements over the traditional algorithms for both face verification and identification tasks, and also exhibits enhanced algorithmic stability to PCA dimension variation. A possible extension of this work is to pursue an optimal affine transformation matrix, instead of projection matrix, for the general classification problem, and to apply the technique for refining previous subspace learning algorithms based on projection matrices.

REFERENCES

[1] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[2] F. Cardinaux, C. Sanderson, and S. Bengio, "Face verification using adapted generative models," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 825–830.

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

[4] K. Jonsson, J. Matas, J. Kittler, and Y. Li, "Learning support vectors for face verification and recognition," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 208–213.

[5] J. Kittler, Y. Li, and J. Matas, "Face verification using client specific Fisher faces," in *The Statistics of Directions, Shapes and Images*, J. Kent and R. Aykroyd, Eds. Surrey, U.K.: Univ. Surrey, 2000, pp. 63–66.

[6] J. Kittler, R. Ghaderi, T. Windeatt, and G. Matas, "Face verification using error correcting output codes," in *Proc. Computer Vision and Pattern Recognition Conf.*, 2001, pp. 755–760.

[7] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[8] H. Liu, C. Su, Y. Chiang, and Y. Hung, "Personalized face verification system using owner-specific cluster-dependent LDA-subspace," in *Proc. Int. Conf. Pattern Recognition*, 2004, pp. 344–347.

[9] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," in *Proc. Computer Vision and Pattern Recognition Conf.*, 2004, pp. 855–861.

[10] J. Luettin and G. Maitre, "Evaluation protocol for the extended M2VTS database (XM2VTS)," presented at the DMI for Perceptual Artificial Intelligence, 1998.

[11] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput. J.*, vol. 16, no. 5, pp. 295–306, 1998.

[12] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 1090–1104, Sep. 2000.

[13] J. Matas *et al.*, "Comparison of face verification results on the XM2VTS database," in *Proc. Inf. Conf. Pattern Recognition*, 2000, vol. 4, pp. 858–863.

[14] K. Messer *et al.*, "Face authentication competition on the BANCA database," in *Proc. Inf. Conf. Biometric Authentication (ICBA)*, Hong Kong, 2004, pp. 8–15.

[15] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, pp. 1771–1782, 2000.

[16] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.

[17] M. Sadeghi and J. Kittler, "Decision making in the LDA space: Generalised gradient direction metric," in *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 248–253.

[18] A. Shashua and A. Levin, "Linear image coding for regression and classification using the tensor-rank principle," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Dec. 2001, vol. 2, pp. 42–49.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[20] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[21] M. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis for image ensembles," in *Proc. Computer Vision and Pattern Recognition Conf.*, Jun. 2003, vol. 2, pp. 93–99.

[22] L. Wiskott, J. Fellous, N. Krger, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[23] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data with application to face recognition," *Pattern Recognit.*, vol. 34, pp. 2067–2070, 2001.

[24] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," in *Proc. IEEE. Conf. Computer Vision Pattern Recognition*, Santa Barbara, CA, Jun. 1998, pp. 164–169.
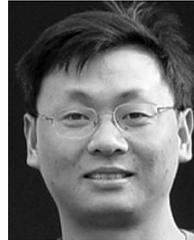
Shuicheng Yan (M'06) received the B.S. and Ph.D. degrees from the Applied Mathematics Department, School of Mathematical Sciences, Peking University, China, in 1999 and 2004, respectively.

His research interests include computer vision and machine learning.

Dong Xu received the B.S. and Ph.D. degrees from the Electronic Engineering and Information Science Department, University of Science and Technology of China, in 2001 and 2005, respectively.

His research interests include pattern recognition, computer vision, and machine learning.

Xiaoou Tang (S'93–M'96–SM'02) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is a Professor and Director of the Multimedia Laboratory in the Department of Information Engineering, The Chinese University of Hong Kong. He is also the Group Manager of the Visual Computing Group, Microsoft Research Asia, Beijing.

His research interests include computer vision, pattern recognition, and video processing.

Dr. Tang is a Local Chair of the IEEE International Conference on Computer Vision (ICCV) 2005, an Area Chair of ICCV'07, a Program Chair of ICCV'09, and a General Chair of the ICCV International Workshop on Analysis and Modeling of Faces and Gestures 2005. He is a Guest Editor of the Special Issue on Underwater Image and Video Processing for the IEEE JOURNAL OF OCEANIC ENGINEERING and the Special Issue on Image- and Video-Based Biometrics for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He is an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.