Multi-task Pose-Invariant Face Recognition

Changxing Ding, *Student Member, IEEE*, Chang Xu, and Dacheng Tao, *Fellow, IEEE*

***Abstract*— Face images captured in unconstrained environments usually contain significant pose variation, which dramatically degrades the performance of algorithms designed to recognize frontal faces. This paper proposes a novel face identification framework capable of handling the full range of pose variations within ±90° of yaw. The proposed framework first transforms the original pose-invariant face recognition problem into a partial frontal face recognition problem. A robust patch-based face representation scheme is then developed to represent the synthesized partial frontal faces. For each patch, a transformation dictionary is learnt under the proposed multi-task learning scheme. The transformation dictionary transforms the features of different poses into a discriminative subspace. Finally, face matching is performed at patch level rather than at the holistic level. Extensive and systematic experimentation on FERET, CMU-PIE, and Multi-PIE databases shows that the proposed method consistently outperforms single-task-based baselines as well as state-of-the-art methods for the pose problem. We further extend the proposed algorithm for the unconstrained face verification problem and achieve top-level performance on the challenging LFW data set.**

***Index Terms*— Pose-invariant face recognition, partial face recognition, multi-task learning.**

I. Introduction

FACE recognition has been one of the most active research topics in computer vision for more than three decades. With years of effort, promising results have been achieved for automatic face recognition, in both controlled [1] and uncontrolled environments [2], [3]. However, face recognition remains significantly affected by the wide variations of pose, illumination, and expression often encountered in real-world images. The pose problem in particular is still largely unsolved, as argued in a recent work [4]. In this paper, we mainly handle the identification problem of matching an arbitrary pose probe face with frontal gallery faces, which is the most common setting for both the research and application of pose-invariant face recognition (PIFR) [4]–[8]. At the end of the paper, we briefly extend the proposed approach to solve the unconstrained face verification problem [9].

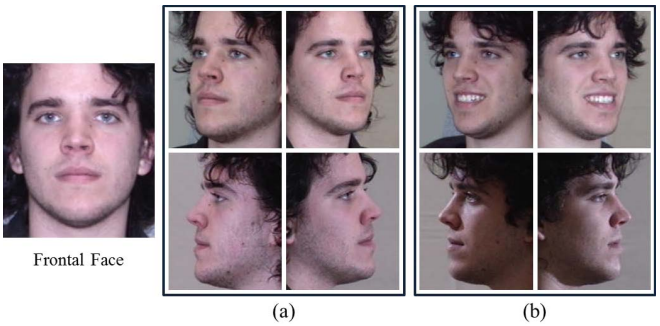


Fig. 1. (a) The rigid rotation of the head results in self-occlusion as well as nonlinear facial texture deformation. (b) The pose problem is combined with other factors, e.g., variations in expression and illumination, to affect face recognition.

Pose variation induces dramatic appearance change in the face image. Essentially, this is caused by the complex 3D geometrical structure of the human head. As shown in Fig. 1, the rigid rotation of the head results in self-occlusion, which means that some facial texture will be invisible with variations in pose. Even the shape and position in the image of visible facial texture vary nonlinearly from pose to pose. The pose problem is also usually combined with other factors, such as variations in illumination and expression, to affect the appearance of face images. In consequence, the extent of appearance change caused by pose variation is usually greater than that caused by differences in identity, and the performance of frontal face recognition algorithms degrades dramatically when the images to be matched feature different poses.

Directly matching faces in different poses is difficult. One intuitive solution is to conduct face synthesis so that the two facial images can be compared in the same pose. Most approaches following this idea are dedicated to recovering complete frontal faces from non-frontal faces [4], [8], [10]–[12]. However, synthesizing the entire frontal face from profile faces is difficult since most facial texture is invisible as a result of occlusion. Therefore, the aforementioned methods tend to constrain their recognition ability within ±45° of yaw variation.

Inspired by the observation that human beings can easily recognize profile faces without the need to elaborately recover the whole frontal face, we present a novel face representation approach that makes full use of just the facial texture that is occlusion-free. This representation approach is general in

Manuscript received August 7, 2014; revised October 27, 2014; accepted December 28, 2014. Date of publication January 12, 2015; date of current version January 30, 2015. This work was supported by the Australian Research Council under Grant DP-120103730, Grant DP-140102164, Grant FT-130101457, and Grant LP-140100569. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shiguang Shan. *Corresponding author: Dacheng Tao.*

C. Ding and D. Tao are with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology at Sydney, Sydney, NSW 2007, Australia (e-mail: changxing.ding@student.uts.edu.au; dacheng.tao@uts.edu.au).

C. Xu was with the Center of Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology at Sydney, Sydney, NSW 2007, Australia. He is now with the Key Laboratory of Machine Perception, School of Electronics Engineering and Computer Science, Ministry of Education, Peking University, Beijing 100871, China (e-mail: xuchang@pku.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2015.2390959

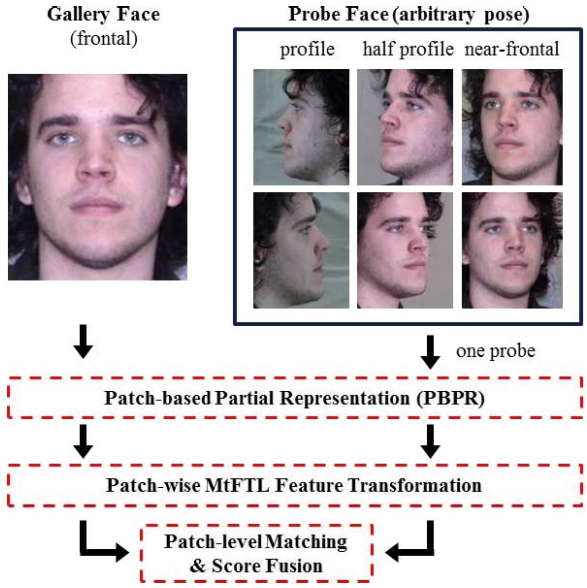


Fig. 2. Overview of the proposed PBPR-MtFTL framework for pose-invariant face recognition, as applied to the recognition of arbitrary pose probe faces.

nature, which means that it applies continuously to the full range of pose variations between $-90°$ and $+90°$ of yaw. Using simple pose normalization and pre-processing operations, our approach converts the original PIFR problem into a partial face recognition task [13], [14], in which the face is represented by the unoccluded facial parts in a patch-based fashion. The proposed face representation scheme is therefore named Patch-based Partial Representation (PBPR).

Feature transformation enhances recognition ability by transforming the features from the gallery and probe images to a common discriminative subspace. Based on this intuition, we propose a learning method called Multi-Task Feature Transformation Learning (MtFTL). By considering the correlation between the transformation matrices for different poses, MtFTL consistently achieves better performance than its single-task based counterparts. Its advantage is particularly evident when the size of training data is limited. The transformation matrices learnt by MtFTL are highly compact as a result of sharing most projection vectors across poses, which additionally reduces memory cost.

We term the entire proposed framework for tackling the pose problem PBPR-MtFTL. Under this framework, matching an arbitrary pose probe face and frontal gallery faces involves transformation of the extracted PBPR representation using the learnt MtFTL transformation dictionaries, followed by patch-level cosine distance computation and score fusion, as illustrated in Fig. 2. Extensive experiments on FERET, CMU-PIE, and Multi-PIE datasets indicate that superior performance is consistently achieved with PBPR-MtFTL.

The remainder of the paper is organized as follows: Section II briefly reviews related works for PIFR and multi-task learning. The proposed PBPR face representation scheme is illustrated in Section III. The multi-task feature transformation learning approach MtFTL is described in Section IV. Face matching using PBPR-MtFTL is introduced in Section V. Experimental results are presented in Section VI, leading to conclusions in Section VII.

II. Related Studies

A. Pose-Invariant Face Recognition

Many promising approaches have been proposed to tackle the pose challenge in face recognition [15]. These methods can be broadly classified into two categories: face image synthesis-based methods and synthesis-free methods.

Most previous works rely on face image synthesis, Face image synthesis can be accomplished with 2D or 3D techniques. Using 2D techniques, Ashraf *et al.* [16] learnt patch-wise warps between two images with the Lucas-Kanade optimization algorithm. Chai *et al.* [11] proposed the locally linear regression method for frontal face synthesis. Li *et al.* [17] proved that more accurate face synthesis can be achieved by imposing lasso or ridge regularization on the regression function. Ho *et al.* [8] proposed synthesizing the virtual frontal view using Markov Random Fields and a variant of belief propagation algorithm. Li *et al.* [18] proposed the Morphable Displacement Field method for frontal face synthesis and achieved pixel-level semantic correspondence between a face pair. Other 2D face synthesis methods can be found in related facial analysis topics, e.g., face hallucination [19], [20]. Methods based on 3D face models have also been introduced because pose variation is essentially caused by the 3D rigid transformation of the face. Blanz *et al.* [5] developed the 3D Morphable Model (3DMM) which can be used to fit a 2D face image in arbitrary poses. Face recognition can be conducted by rendering a specified view with the 3D model or directly matching the regression coefficents; however, the optimization of 3DMM is computationally expensive. In [21], an efficient 3D modeling method called Generic Elastic Models (GEM) was introduced. By assuming that the depth information between individuals is not highly discriminative, GEM estimates the 3D shape of a 2D frontal face by directly assigning generic depth information. Arbitrary pose face images can be rendered with GEM for matching; however, GEM can only estimate the 3D shape for frontal faces. Asthana *et al.* [22] proposed synthesizing the frontal face image by aligning and mapping the texture of a 2D face image to a 3D generic shape model. This method can handle pose variations within $\pm 45°$ of yaw.

The synthesis-free approaches aim to extract pose-robust features or to transform features from different poses into a shared subspace. For example, [12], [23] proposed learning pose-insensitive face representations with neural networks. Wright *et al.* [24] represented the face image as a histogram of quantized patch descriptors which are expanded with their spatial locations. Yi *et al.* [25] proposed the extraction of pose adaptive features by transforming Gabor filers according to the pose of the face image. Of the feature transformation-based methods, Li *et al.* [26] proposed employing Canonical Correlation Analysis to project image patches under two different views to a shared subspace, where the patch similarity can be measured by the correlation of their projections. Similar methods incorporate Partial Least Squares (PLS) [27] and

Generalized Multiview Analysis [28]. Recently, Kan *et al.* [29] propose an efficient algorithm that simultaneously learns pose-specific transformations to a common discriminative space. This approach enables convenient matching between two faces with arbitrary poses. Other methods to learn robust representations or feature transformations can be found in related areas, e.g., scene classification [30] and image denoising [31].

The two categories of method described above are closely related. For example, it is shown in [12] and [23] that pose-invariant features can be utilized for frontal-face reconstruction. It is also demonstrated that the coefficients of regression models for face synthesis [4], [5], [11], [17] can be regarded as pose-insensitive features for face matching. Our work is related to both categories, and there is clear novelty: First, the proposed PBPR method represents arbitrary pose face images from the perspective of partial face recognition. Second, the MtFTL approach learns compact feature transformation for various poses based on the principle of multi-task learning, which is novel for PIFR. Third, the PBPR-MtFTL framework continuously tackles the full range of pose variations from $-90°$ to $+90°$ of yaw and obtains strong performance. In comparison, the recognition ability of existing methods is typically restricted to a range of $\pm 45°$ of yaw [4], [32].

B. Multi-Task Learning

Multi-task learning (MTL) is a machine learning technique that learns several tasks simultaneously for better performance by capturing the intrinsic correlation between different tasks. MTL implicitly increases the sample size and improves the generalization ability for each task; hence, it is especially beneficial when the training data for the tasks is small.

While MTL has been widely applied to computer vision tasks, e.g., visual tracking [33], action recognition [34]–[36], and face recognition [37], [38], it is new for PIFR. Existing approaches for PIFR ignore the correlation between the feature transformations of different poses [26], [28]. To the best of our knowledge, MTL for PIFR is only briefly mentioned in [39] but no detailed information is provided, and multi-view reconstruction is targeted rather than feature transformation learning. Nevertheless, MTL provides a principled way for us to model the correlation between poses if we view the learning of feature transformation for each pose as a task. MtFTL is arguably the first MTL approach that jointly learns feature transformations for different poses and is shown to profit from the latent inter-pose correlations.

III. Face Representation for the Pose Problem

Existing face representation methods tend to extract fixed-length features from face images, with the underlying assumption that all facial components are visible in the image [13]. However, as shown in Fig. 2, this hypothesis does not hold for a profile face where there is severe self-occlusion. In this section, we propose the flexible PBPR face representation scheme, where the length of face representation is related to the pose of the face; for example, a frontal face image will have larger face representation than a profile face image. This is reasonable, since the profile face provides less information

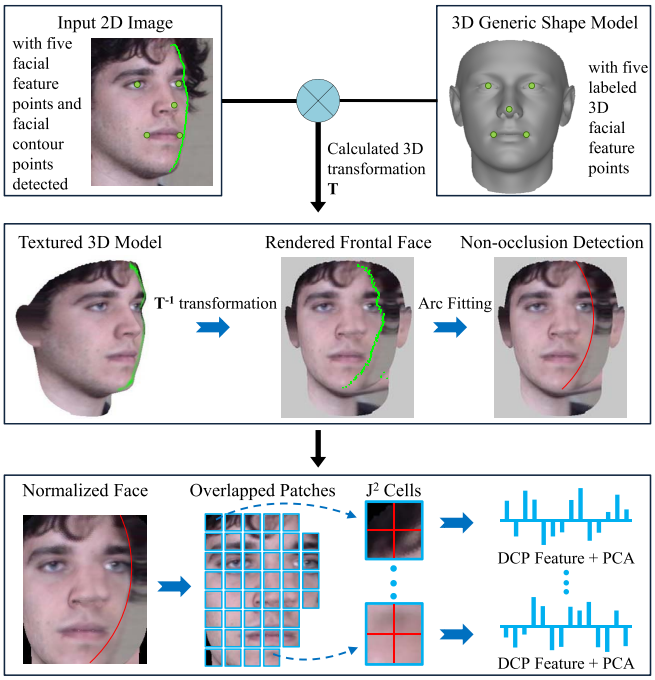


Fig. 3. Overview of the proposed PBPR face representation method. PBPR is applied to arbitrary pose face images. The final PBPR representation is a set of patch-level DCP features after dimension reduction by PCA.

for recognition. As shown in Fig. 3, PBPR is essentially composed of three steps: face pose normalization, unoccluded facial texture detection, and patch-wise feature extraction. In this section, we describe the three main components in detail.

A. Face Pose Normalization

A standard 3D method is adopted for face pose normalization [22]. The five most stable facial feature points, i.e., the centers of both eyes, the tip of the nose, and the two mouth corners, are first detected automatically or manually. For profile faces (as shown in Fig. 2), the coordinates of the occluded facial feature points are estimated. Using the orthographic projection model [40] and the detected five facial feature points, a 3D generic shape model is aligned to the 2D face image.[1] The 2D face image is then back-projected to the 3D model, and a frontal face image is rendered with the textured 3D model.

Previous works rely on dense facial feature points, e.g., 68 points in [22] and 79 points in [4], for accurate pose normalization. However, detecting dense facial feature points for profile faces is difficult due to the severe self-occlusion of the face, which in turn restricts the range of poses that these methods can handle. In stead, only the five most stable facial feature points are utilized for pose normalization in this paper. This greatly facilitates the realization of a fully automatic face recognition system and extends the range of poses that can be processed. Although using sparse facial feature points will result in larger normalization error, we highlight the

[1] In this step, we roughly estimate the pose of the 2D face image by the method described in [41].

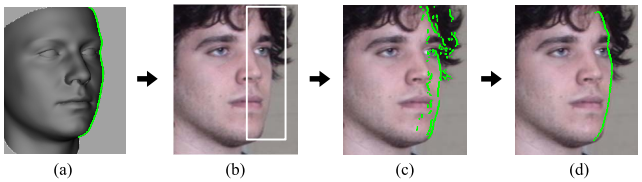


Fig. 4. Illustration of facial contour detection. (a) The 3D generic shape model is projected to the 2D plane and its facial contour is detected; (b) the region containing the facial contour of the 2D face image is estimated; (c) candidate facial contour points; (d) facial contour obtained by point set registration.

power of the proposed PBPR-MtFTL framework given its low normalization requirements.

B. Unoccluded Facial Texture Detection

Pose normalization corrects the deformation of facial texture resulting from pose variations, but it cannot recover the texture lost by occlusion. Rather than trying to synthesize the occluded texture to obtain a complete frontal face [4], we propose to make full use of the unoccluded texture only. This is inspired by the observation that human beings can easily recognize profile faces without the need to recover the whole frontal face. As shown in Fig. 3, the main boundary between the occluded and unoccluded facial texture is the facial contour. Therefore, facial contour detection is the key to identifying the occluded facial texture.

Although there are off-the-shelf face alignment tools for facial contour detection, they return only sparse facial contour points and may not be reliable enough to severe occlusion, expression, and pose variations. We propose a much simpler but effective method that makes use of the 3D generic shape model. After aligning the 3D model and the 2D face image, it can be projected to the 2D image plane roughly in the pose of the 2D face. As shown in Fig. 4(a), the contour of the 3D model can be easily detected. Based on the contour of the 3D model, the facial contour search of the 2D face can be constrained within a certain region, as illustrated in Fig. 4(b). The edge points are then detected in this region by the Canny operator [42]. To reduce imposters, only the edge points with horizontal gradient directions are saved, with the prior that facial contour extends in a vertical direction. Lastly, the facial contour is obtained by a point sets registration algorithm called Coherent Point Drift (CPD) [43]. Briefly, CPD iteratively aligns the facial contour highlighted in Fig. 4(a) to the edge point set shown in Fig. 4(c) with affine transformations. The imposter contour points in Fig. 4(c) can gradually be detected and ignored. The obtained facial contour is shown in Fig. 4(d). More detection examples on unconstrained face images in the LFW dataset [9] are shown in Fig. 5.

Alongside the projection of facial texture in the first step, the detected facial contour points are first projected to the 3D model and then projected to the rendered frontal face image. Since the head is approximately an ellipsoid, the facial contour points in the frontal view are fitted with an arc. As shown in Fig. 3, the arc effectively separates the unoccluded and occluded texture in the rendered frontal image.

Fig. 5. Examples of facial contour detection for unconstrained face images in the LFW dataset.

In the following subsection, face representation is built using only the detected unoccluded facial texture.

C. Patch-Based Face Representation

The area of the unoccluded facial texture in the rendered frontal view varies with pose change, with demonstrable fluctuation in the amount of effective information available for face recognition. In light of this observation, a variable-length face representation method is proposed.

As illustrated in Fig. 3, the normalized face image is first divided into $M \times N$ overlapped patches. The severity of occlusion for each patch is then evaluated based on the detected boundary between the occluded and unoccluded facial texture. If more than 80% of pixels in one patch fall into the unoccluded region, then it is designated as an unoccluded patch; otherwise, the patch is ignored due to the large area of occlusion. Next, each of the unoccluded patches is split into $J \times J$ cells. A state-of-the-art local descriptor called Dual-Cross Patterns (DCP) [2] is employed for feature extraction. The concatenated DCP histogram feature from the J^2 cells forms the raw feature of the patch. Following [2], elements in the DCP histogram are normalized by square root. Lastly, Principal Component Analysis (PCA) is applied to each patch to project its feature into a subspace with dimension D, by which the noise is suppressed.

The set of patch-level DCP features following PCA processing from all unoccluded patches forms the representation of the face image. Note that this representation method is general in nature, meaning that it applies to faces with arbitrary poses. This is a valuable property, because we do not need to apply different algorithms to frontal and non-frontal faces, unlike some existing approaches [8].

IV. Multi-Task Feature Transformation Learning

The previous section has introduced the PBPR face representation scheme, whereby face recognition can be

Fig. 6. Pose normalization for non-frontal images. The boundary between unoccluded and occluded facial texture is detected by the method illustrated in Fig. 3. (a) $-90° \leq yaw \leq -45°$; (b) $-30° \leq yaw \leq +30°$; (c) $+45° \leq yaw \leq +90°$. The image quality is degraded with the increase in value of the yaw angles, and the amount of unoccluded facial texture for recognition decreases.

accomplished by directly matching corresponding patch features of two face images. In this section, we further propose the MtFTL approach for learning transformation dictionaries, which enable the patch features of a frontal face and a non-frontal face to be transformed into a common discriminative space to enhance recognition ability. The learning process is patch-wise, which means a separate transformation dictionary is learnt by MtFTL for each patch. Consequently, we obtain $M \times N$ transformations dictionaries. Details of the MtFTL approach are illustrated below.

A. Feature Transformation Learning

Three aspects are considered in the design of feature transformation learning. First, as shown in Fig. 6, the normalized images from different poses are of different image quality, therefore there will be differences in the transformations for different poses. Second, a strong correlation exists between the feature transformations for different poses, since they essentially process the data of the same subjects. Third, the amount of training data might be limited in real scenarios, because collecting multi-pose face images tends to be difficult. Ideally, the shared knowledge from different poses should be leveraged for robust transformation learning.

These considerations call for a multi-task strategy for feature transformation learning, in which the learning for each pose type is regarded as a task. Therefore, we propose the MtFTL approach which takes into consideration both the correlation and difference between tasks. Instead of learning a separate transformation matrix for each task [44], MtFTL learns a common transformation dictionary for all the tasks. Differences between the tasks are reflected by the selection of different projection vectors in the transformation dictionary. Hence, MtFTL learns more compact feature transformations than previous approaches [44].

Before presenting the formulation of the proposed model, several necessary notations are introduced. Let P be the number of tasks, i.e., the number of pose types that are available in the training set for the current patch. The set $\{(X_t, Y_t) : 1 \leq t \leq P\}$ stores the training data composed of intra-personal and inter-personal patch pairs. $X_t \in \mathbb{R}^{D \times N_{tp}}$ and $Y_t \in \mathbb{R}^{D \times N_{tn}}$, where N_{tp} and N_{tn} are the number of intra-personal and inter-personal patch pairs for the tth task, respectively. The nth column of X_t is denoted as x_t^n and $x_t^n = x_{tn} - x_{0n}$, where $\{x_{tn}, x_{0n}\}$ is one intra-personal patch pair between the pose type t and the frontal pose. Similarly, y_t^n is the nth column of Y_t and $y_t^n = y_{tn} - y_{0n}$, where $\{y_{tn}, y_{0n}\}$ is one inter-personal patch pair between the pose type t and the frontal pose. We learn the transformation dictionary $U \in \mathbb{R}^{D \times D}$ shared by all tasks, and the set of vectors $\alpha_t \in \{0, 1\}^D, 1 \leq t \leq P$. α_t selects projection vectors for task t from the shared dictionary U. Let $A \in \mathbb{R}^{D \times P}$ be the matrix of stacked vectors α_t. In addition, $A_t \in \mathbb{R}^{D \times D}$ is a diagonal matrix expanded by α_t, denoted as $A_t = diag(\alpha_t)$.

The loss function for task t is denoted as $T_{U,t}$. It is based on the principle that the margin between intra-personal patch pairs and inter-personal patch pairs should be as large as possible.

$$T_{U,t}(\alpha_t) = \frac{1}{N_{tp}} \|A_t U^T X_t\|_F^2 - \frac{\lambda}{N_{tn}} \|A_t U^T Y_t\|_F^2, \quad (1)$$

where λ is a regularization parameter that weights the intra-personal and inter-personal terms. We then formulate the multi-task learning algorithm as the optimization problem:

$$\min_{U,A} \ \frac{1}{P} \sum_{t=1}^{P} T_{U,t}(\alpha_t) \quad s.t. \ \ U^T U = I. \quad (2)$$

For simplicity, the above problem is relaxed to

$$\min_{U,A} \frac{1}{P} \sum_{t=1}^{P} T_{U,t}(\alpha_t) + \mu \|U^T U - I\|_F^2, \quad (3)$$

where μ is another regularization parameter. We set the number of non-zero elements in α_t as d, and aim to optimize α_t to select d most discriminative projection vectors from U for the tth task. The optimal value of μ, d, and λ is estimated through cross validation.

Note that the transformation dictionary U is learnt jointly for all P tasks, which enables knowledge sharing between the tasks. We show in the experiment section that knowledge sharing is especially important when the amount of training data for the tasks is limited.

B. Iterative Optimization Algorithm

The optimization problem (3) of MtFTL is convex in U for fixed A and in A for fixed U. Therefore, we solve this problem by alternately optimizing U and A. The final learning algorithm is summarized in Algorithm 1. The main optimization procedure can be outlined in two steps.

Algorithm 1 Multi-Task Feature Transformation Learning

Input: data $\{(X_t, Y_t) : 1 \le t \le P\}$, parameters λ, μ, d, ϵ
Output: U, A
Initialize $U^{(0)} := rand(d, d)$, $A^{(0)} := \mathbf{0}$, $s := 0$;
repeat
  $s := s + 1$;
  for $t = 1$ **to** P **do**
    Let $\alpha_t^{(s)} = \arg\min_{\alpha_t} T_{U,t}(\alpha_t)$: Eq. (5)
  end
  Let $U^{(s)} = \arg\min_{U} T_A(U)$: Eq. (7)
  (via the LBFGS algorithm)
  $\Delta U = U^{(s)} - U^{(s-1)}$
  $\Delta A = A^{(s)} - A^{(s-1)}$
until $\|\Delta U\|_F^2 < \epsilon$ *and* $\|\Delta A\|_F^2 < \epsilon$;

Learning A: With fixed U, the optimization problems for each task decouple. For the tth task, the optimal α_t is obtained:

$$\min_{\alpha_t \in \mathbb{R}^D} T_{U,t}(\alpha_t), \tag{4}$$

$$\begin{aligned} T_{U,t}(\alpha_t) &= tr\left(\frac{1}{N_{tp}} A_t U^T X_t X_t^T U A_t - \frac{\lambda}{N_{tn}} A_t U^T Y_t Y_t^T U A_t\right) \\ &= tr\left\{A_t\left(\frac{1}{N_{tp}} U^T X_t X_t^T U - \frac{\lambda}{N_{tn}} U^T Y_t Y_t^T U\right) A_t\right\} \\ &= \alpha_t^T B_t \alpha_t, \end{aligned} \tag{5}$$

where $tr(\cdot)$ represents the trace of a matrix; and B_t is a diagonal matrix by directly copying the diagonal elements from the matrix $\frac{1}{N_{tp}} U^T X_t X_t^T U - \frac{\lambda}{N_{tn}} U^T Y_t Y_t^T U$. Since the role of α_t is to select d most discriminative projection vectors for the tth pose type, the elements in α_t that correspond to d smallest diagonal elements in B_t are set as 1 while the other elements in α_t are set as 0.

Learning U: The shared transformation dictionary U couples all the tasks. In this step, U is updated efficiently via the limited-memory BFGS (LBFGS) algorithm. The choice of LBFGS algorithm here is due to both its high efficiency and low memory requirement.

While A is fixed, the optimization problem (3) reduces to

$$\min_{U \in \mathbb{R}^{D \times D}} T_A(U), \tag{6}$$

$$\begin{aligned} T_A(U) = \frac{1}{P} \sum_{t=1}^{P} &\left(\frac{1}{N_{tp}} \|A_t U^T X_t\|_F^2 - \frac{\lambda}{N_{tn}} \|A_t U^T Y_t\|_F^2\right) \\ &+ \mu \|U^T U - I\|_F^2. \end{aligned} \tag{7}$$

The derivative of $T_A(U)$ with respect to U is

$$\begin{aligned} \frac{\partial T_A(U)}{\partial U} = \frac{2}{P} \sum_{t=1}^{P} &\left(\frac{1}{N_{tp}} X_t X_t^T U A_t A_t - \frac{\lambda}{N_{tn}} Y_t Y_t^T U A_t A_t\right) \\ &+ 4\mu\left(U U^T U - U\right). \end{aligned} \tag{8}$$

With the provided formula for calculating $T_A(U)$ and $\frac{\partial T_A(U)}{\partial U}$, the optimization problem can be readily solved with the LBFGS algorithm [45].

Initialization: The transformation matrix U is simply initialized with a random matrix whose elements are drawn from the standard uniform distribution on the open interval (0, 1). In the experiment section, we show that even the randomly initialized U achieves promising performance.

Stopping criterion: The iterative optimization process stops when the Frobenius norms of both ΔU and ΔA are below $\epsilon = 10^{-3}$, where ΔU and ΔA are the difference matrices between two successive iterations for U and A, respectively.

C. Theoretical Analysis

In this subsection, we study the robustness and generalization error of the proposed MtFTL algorithm. The detailed proof can be found in the Appendix. All through the theoretical analysis, we consider the loss function for face patch feature x_θ at non-frontal pose θ as

$$\ell(A, U, x, \theta) = \|A_\theta U^T (x_\theta - x_0)\|, \tag{9}$$

whose maximum value is assumed to be B.

1) Robustness Analysis: If two corresponding face patch features of two images are from the same subject, then their associated losses are close. This property is formalized as "robustness" in [46], and the precise definition is given below:

Definition 1: An algorithm $\mathcal{A}$ is $(K, \epsilon(\cdot))$ robust, for $K \in \mathbb{N}$ and $\epsilon(\cdot) : \mathcal{Z} \to \mathbb{R}$, if the sample $\mathcal{Z}$ can be partitioned into K disjoint sets, denoted as $\{C_i\}_{i=1}^K$, so that the following holds for all $\mathbf{s} \in \mathcal{Z}$, given the loss function $\ell(\mathcal{A}_s, z)$ of the algorithm $\mathcal{A}_s$ trained on s:

$$\begin{aligned} &\forall s \in \mathbf{s}, \quad \forall z \in \mathcal{Z}, \quad \forall i = 1, \cdots, K : \\ &if\ s, z \in C_i, \quad then\ |\ell(\mathcal{A}_s, s) - \ell(\mathcal{A}_s, z)| \le \epsilon(\mathbf{s}). \end{aligned}$$

Given two face patch features s and z of the same subject from different poses θ_s and θ_z, if $\|s - z\| \le \gamma$ and $|\theta_s - \theta_z| \le \Delta_\theta$, we suggest that these two face patch features are close. We assume that for one subject, the difference between any of its non-frontal face patch feature x_θ and its frontal face patch feature x_0 can be bounded by $\|x_\theta - x_0\| \le \gamma_0$. Since matrix A_θ at pose θ is a sparse diagonal matrix that has d non-zero elements, we have $\|A_\theta\| \le \sqrt{d}$. Also, we restrict $\|A_{\theta_1} - A_{\theta_2}\| \le \mathbf{\Omega}_{\Delta_\theta}$ for any two matrices A_{θ_1} and A_{θ_2} at different poses. A face recognition algorithm is said to be robust if the corresponding face patch features from images of the same subject have close losses. This robustness can be measured by the following theorem.

Theorem 1: Example z is in space $\mathcal{Z} \subset \mathbb{R}^D$, which can be partitioned into K disjoint sets and denoted as $\{C_{i=1}\}_{i=1}^K$. Given the algorithm $\mathcal{A}$ $\{A, U : z \to \mathbb{R}^d\}$, we have for any $\mathbf{s} \subset \mathcal{Z}$,

$$\begin{aligned} &|\ell(\mathcal{A}_s, z) - \ell(\mathcal{A}_s, s)| \le \sqrt{d}\gamma + \mathbf{\Omega}_{\Delta_\theta}\gamma_0 \\ &\forall i, \quad j = 1, \cdots, K \ : s \in C_i \text{ and } z \in C_j. \end{aligned}$$

Hence $\mathcal{A}$ is $(K, \sqrt{d}\gamma + \mathbf{\Omega}_{\Delta_\theta}\gamma_0)$-robust.

Robustness is a fundamental property which ensures that a learning algorithm performs well. Since the sparse diagonal matrices A_s and A_z, which select projection vectors from the transformation dictionary U, are learned from face patches of different poses, they cannot be identical. According to Theorem 1, it is instructive to suggest that the robustness of the algorithm will be improved for the face patches at close poses if their feature transformations

have more shared elements, that is, encouraging Ω_{Δ_θ} to be small.

2) Generalization Analysis: Based on the robustness analysis, we show a PAC generalization bound for the algorithm, i.e., the difference between the expected error $\mathcal{L}(\mathcal{A}_s)$ and the empirical error $\mathcal{L}_{emp}(\mathcal{A}_s)$. We begin by presenting a concentration inequality [47] that helps to derive the bound.

Proposition 1: Let $(|N_1|, \cdots, |N_K|)$ *be an IID multinomial random variable with parameters n and* $(\beta(C_1), \cdots, \beta(C_K))$. *By the Breteganolle-Huber-Carol inequality we have* $Pr\{\sum_{i=1}^{K} |\frac{N_i}{n} - \beta(C_i)| \geq \zeta\} \leq 2^K \exp(\frac{-n\zeta^2}{2})$, *hence with probability at least* $1-\delta$,

$$\sum_{i=1}^{K} |\frac{N_i}{n} - \beta(C_i)| \leq \sqrt{\frac{2K \ln 2 + 2\ln(1/\delta)}{n}}.$$

The generalization error bound is presented in the following theorem.

Theorem 2: If the algorithm $\mathcal{A}$ *is* $(K, \epsilon(\cdot))$*-robust and the training sample* $\mathbf{s}$ *is composed of n examples* $\{s_i\}_{i=1}^n$, *which are generated from* β, *then for any* $\delta > 0$, *with the probability at least* $1-\delta$ *we have,*

$$|\mathcal{L}(\mathcal{A}_s) - \mathcal{L}_{emp}(\mathcal{A}_s)| \leq \epsilon(s) + B\sqrt{\frac{2K \ln 2 + 2\ln(1/\delta)}{n}}.$$

By combining the results of Theorem 1 and Theorem 2, we can easily illustrate the generalization error of the proposed algorithm. Exploiting the shared information of face patches from different poses can strengthen the robustness of the algorithm and then improve the generalization error.

V. Face Matching With PBPR-MtFTL

In this section, the face matching problem is addressed based on the proposed PBPR-MtFTL framework. It is assumed that $\left\{(U^i, A^i) : 1 \leq i \leq MN\right\}$, i.e., the set of patch-wise transformation dictionaries and selection matrices, has been learnt by MtFTL.

Suppose we are matching a probe face image x_t of pose type t to a frontal gallery face image x_0. It is also assumed that there are K unoccluded patches for x_t. Without loss of generality, we denote the sets of features for the K patches as $\{x_{t1}, x_{t2}, \cdots, x_{tK}\}$ and $\{x_{01}, x_{02}, \cdots, x_{0K}\}$ for x_t and x_0, respectively. First, the features of each patch pair $\{(x_{tk}, x_{0k}) : 1 \leq k \leq K\}$ are projected into the discriminative space using the learnt U^k and A^k.

$$\begin{aligned} \hat{x}_{tk} &= A_t^k (U^k)^T x_{tk} \\ \hat{x}_{0k} &= A_t^k (U^k)^T x_{0k}, \end{aligned} \tag{10}$$

where A_t^k is the diagonal matrix expanded by the tth column of A^k. Then, the cosine metric is utilized to calculate the similarity of each patch pair and the similarity scores of all K patch pairs are fused by the sum rule.

$$s(x_t, x_0) = \frac{1}{K} \sum_{k=1}^{K} \frac{\hat{x}_{tk}^T \hat{x}_{0k}}{\|\hat{x}_{tk}\| \|\hat{x}_{0k}\|}, \tag{11}$$

where s is the similarity score between the probe image x_t and the gallery image x_0. Lastly, the nearest neighbor (NN) classifier is adopted for face identification.

Although the adopted matching scheme is simple compared to existing methods [17], [26], it is still expected that the proposed PBPR-MtFTL framework will achieve stronger performance, since the recognition ability of PBPR-MtFTL has been enhanced by exploiting the correlation between poses.

VI. Experimental Evaluation

In this section, extensive experiments are conduct to present the effectiveness of PBPR-MtFTL. We mainly conduct identification experiments on the three most popular databases for the pose problem, i.e., CMU-PIE [48], FERET [49], and Multi-PIE [50]. These experiments are to recognize a subject across pose variations with a single enrolled frontal face image. At the end of this section, we slightly modify the proposed framework to deal with the unconstrained face verification problem, and conduct experiments on the challenging LFW dataset [9].

The CMU-PIE [48] and FERET [49] datasets incorporate multi-pose images of 68 and 200 subjects, respectively. For the two databases, we adopt the same protocols as previous works [7], [18] that exclude both illumination and expression variations [51]. The Multi-PIE [50] database contains images of 337 subjects, each of which is captured in up to four recording sessions. Images in each session cover 15 view points and 20 illumination conditions. As there is no unified protocol for the pose problem on Multi-PIE, we adopt the three most popular protocols in the literature [17], [22], [39].

Eight sets of experiments are conducted. First, the performance of PBPR-MtFTL is briefly compared with previous works for PIFR on CMU-PIE and FERET. Next, the MtFTL approach is compared with its single-task baselines on Multi-PIE to justify the significance of MTL for the pose problem. Then, considering that the pose problem is often combined with other factors, we evaluate the performance of PBPR-MtFTL in three different settings, i.e., combined variations of pose and illumination, combined variations of pose and recording session, and combined variations of pose, illumination, and recording session. We also test the sensitivity of PBPR-MtFTL to the value of model parameters and face alignment errors. Lastly, we slightly modify the proposed approach to deal with the unconstrained face verification problem and present experimental results on the LFW database.

All images in this paper are normalized as follows. The mean shape of the Basel Face Model (BFM) [52] is adopted as the 3D generic shape model. The five facial feature points are manually labeled in the first six experiments and automatically detected in the last two experiments.[2] After the pose normalization step described in Section III, the face images are cropped and resized to 156×130 pixels, as shown in Fig. 6. The patch size $M \times N$ is set at 26×24 pixels, with 50% overlap between nearby patches. The number of cells $J \times J$ within each patch is set at 2×2. For the first seven experiments, images are further photometrically normalized using a simple operator [1], with the two parameters σ_1 and σ_2

[2]Coordinates of manually labeled facial feature points for Multi-PIE are provided by Zhu et al. [12]; we use an off-the-shelf tool [53] for automatic facial feature point detection.

TABLE I
MODEL PARAMETERS ESTIMATED ON THE VALIDATION SUBSETS FOR DIFFERENT DATABASES

Database	D	μ	d	λ
CMU-PIE	300	0.1	200	0.5
FERET	200	0.1	50	0.5
Multi-PIE[1]	300	0.1	200	0.5
Multi-PIE[2]	300	0.1	200	0.5
Multi-PIE[3]	300	0.1	200	0.8
LFW	600	0.05	200	0.8

set at 1.4 and 2.0. The two parameters R_{in} and R_{ex} for DCP are set at [3, 7]. For the last experiment, we omit the photometric normalization step since it slightly degrades the performance of PBPR-MtFTL on the View 1 data of LFW. Three-scale DCP features are extracted and concatenated, with the parameters set at [2, 4], [4, 8], and [6, 12], respectively.

For each identification experiment, the subjects in the training data are randomly divided into two subsets, one for model training and the other for validation. The two subsets are of equal size. The optimal values of model parameters μ, d, and λ are estimated on the validation subset and applied to the test data. For simplicity, the value of μ across the models of all patches keeps consistent and this applies to d and λ. The random division of the training data is repeated five times, and the mean rank-1 identification rates on the test data are reported. The estimated model parameters for different databases are tabulated in Table I, where the superscripts "$*^1$", "$*^2$", and "$*^3$" stand for one of the three protocols adopted for the Multi-PIE database, respectively.

A. Comparison on CMU-PIE and FERET

All 68 CMU-PIE subjects with neutral expression and normal illumination at 11 different poses are employed. Note that pose type C31 and C25 are with hybrid yaw and pitch variations. The 68 frontal images are utilized as gallery images and all the rest are used as probes. Following previous works [18], [57], we train the MtFTL model with randomly selected 50 subjects in Multi-PIE database since there are only 68 subjects in CMU-PIE. For FERET, all 200 subjects at 9 different poses are incorporated. Images of the first 100 subjects consist the training data and the rest 100 subjects are used for testing.

As shown in Table II and III, the proposed PBPR-MtFTL approach outperforms the other methods. But the advantage of PBPR-MtFTL is not well exhibited, because the performance of existing methods has nearly reached the saturation point on the two databases. Therefore, we focus on the larger and more challenging Multi-PIE database in the following experiments.

B. Comparison With Single-Task Baselines

In this experiment, we aim to justify the importance of MTL for the pose problem. The proposed MtFTL algorithm is compared with three single-task baselines: (a) the Linear Discriminative Analysis (LDA) approach, which learns a single LDA model for all poses; (b) the single-task feature transformation learning (StFTL) approach, which learns a single feature transformation for all poses. StFTL is equal to the Discriminative Locality Alignment (DLA) model [58]; (c) the multiple independent feature transformation learning (MiFTL) approach, which independently learns a DLA model for each pose. Unlike LDA and StFTL, MtFTL and MiFTL learn pose specific feature transformations. The main difference between MtFTL and MiFTL is that MiFTL learns the transformation for each pose independently, while MtFTL learns compact transformations simultaneously and benefits from the correlation of different poses.

The protocol defined in [17] is employed. This protocol covers 249 subjects in Session 1, in which images with neutral expression under 20 illumination conditions are involved. The first 100 subjects (Subject ID 001 to 100) are used for training and the remaining 149 subjects (Subject ID 101 to 250) are used for testing. The gallery set is composed of 149 frontal images (Pose ID 051) with the illumination ID 07. The probe sets cover 20 illumination conditions of the same subjects. In Fig. 7, we present the performance of the four algorithms with varied size of training data on the three most challenging poses. The number of subjects (S) utilized for model learning is gradually increased from 20 to the maximum number 50.

It is shown in Fig. 7 that MtFTL consistently outperforms the baselines under all settings. Specifically, MtFTL significantly outperforms StFTL and LDA, which means that learning pose specific feature transformations is necessary. MtFTL outperforms MiFTL while learning much more compact transformations. This proves that MTL is helpful for enhancing the ability to recognize non-frontal faces. The advantage of MtFTL is more evident when the amount of training data is limited, which indicates that knowledge sharing among related tasks is important for better generalization ability.

C. Recognition Across Pose and Illumination

In this subsection, the performance of the PBPR-MtFTL framework is compared with existing algorithms under the setting of the combined variations of pose and illumination. The adopted protocol is the same as the previous experiment [17]. The experimental results are shown in Table IV and Fig. 8. In general, face recognition across combined variations of pose and illumination is a difficult problem. However, it is clear that the proposed method outperforms existing approaches [12], [17] with a large margin, even though only half training data is employed to train the MtFTL model.

It is worth noting that the algorithm proposed in [17] also employs photometric normalization, and that all three approaches employ manually labeled facial feature points. Notably, we employ exactly the same facial feature point coordinates as the method followed in [12].

D. Recognition Across Pose and Recording Session

This experiment is to test the performance of algorithms under the combined variations of pose and recording session. The protocol described in [22] is followed. This protocol covers all 337 subjects across the four recording sessions.

TABLE II

PERFORMANCE COMPARISON WITH STATE-OF-THE-ART PIFR METHODS ON CMU-PIE

	C31	C34	C14	C11	C29	C05	C37	C02	C22	C25
yaw angle	$-47°$	$-66°$	$-47°$	$-32°$	$-17°$	$+16°$	$+31°$	$+44°$	$+62°$	$+44°$
pitch angle	$+11°$	$0°$	$0°$	$0°$	$0°$	$0°$	$0°$	$0°$	$0°$	$+11°$
SA-EGFC [18]	-	91.00	100.0	100.0	100.0	100.0	100.0	100.0	96.00	-
Multi-scale MRFs [54]	88.00	83.00	91.00	89.00	91.00	98.00	100.0	95.00	79.00	95.00
MRFs+DAISY [55]	100.0	95.00	100.0	100.0	100.0	100.0	100.0	100.0	92.00	100.0
PBPR-MtFTL	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.53	100.0

TABLE III

PERFORMANCE COMPARISON WITH STATE-OF-THE-ART PIFR METHODS ON FERET

Methods	Gallery Size	bi $-65°$	bh $-45°$	bg $-30°$	bf $-15°$	be $+15°$	bd $+30°$	bc $+45°$	bb $+65°$	Mean
VAAM [22]	200	-	90.50	98.00	98.50	97.50	97.00	91.90	-	-
MRFs+MLBP [56]	200	92.00	98.50	99.50	100.0	99.50	99.00	99.50	91.00	97.38
RR+Gabor [17]	100	78.00	91.00	96.00	96.00	98.00	99.00	96.00	87.00	92.63
SA-EGFC [18]	100	98.00	100.0	100.0	100.0	100.0	100.0	100.0	98.00	99.50
PBPR-MtFTL	100	100.0	99.00	100.0	100.0	100.0	100.0	100.0	98.00	99.63

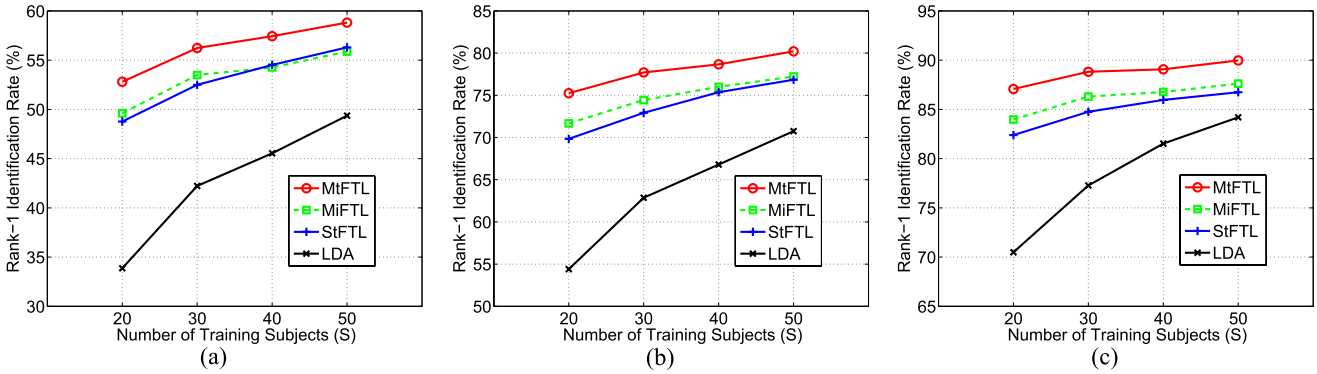


Fig. 7. Performance comparison of MtFTL and the three single-task baselines on the Multi-PIE database with varying numbers of training subjects. (a) yaw $= \pm 90°$; (b) yaw $= \pm 75°$; (c) yaw $= \pm 60°$.

TABLE IV

RANK-1 IDENTIFICATION RATES ON COMBINED VARIATIONS OF POSE AND ILLUMINATION ON MULTI-PIE

Methods	090 $-60°$	080 $-45°$	130 $-30°$	140 $-15°$	050 $+15°$	041 $+30°$	190 $+45°$	200 $+60°$	mean
RR+Gabor [17]	50.64	65.30	70.97	81.07	77.21	73.69	58.12	45.97	65.37
RL+LDA [12]	-	67.10	74.60	86.10	83.30	75.30	61.80	-	-
PBPR-MtFTL	**90.86**	**97.91**	**99.41**	**99.05**	**99.94**	**99.23**	**98.21**	**87.75**	**96.55**

Only images with neutral expression and frontal illumination are employed. Images of the first 200 subjects (Subject ID 001 to 200) are used for training, and images of the remaining 137 subjects (Subject ID 201 to 346) are employed for testing. The frontal images from the earliest recording sessions for the testing subjects are collected as the gallery set (137 images in total). The non-frontal images of the testing subjects construct fourteen probe sets. The comparisons between our approach and the state-of-the-art methods are presented in Table V and Fig. 9. We observe that:

1) In general, the performance of all the algorithms is good when the pose value of the probe images is small. While high performance is achieved by all methods on the probe sets 130, 140, 050, and 041, our method achieves perfect identification rates on all four probe sets.
2) There is a substantial drop in performance for existing methods on the probe sets 080 and 190, where the yaw angles are $\pm 45°$. PBPR-MtFTL performs significantly better than the other methods on both probe sets, indicating that it is more robust to large pose variations.
3) While most existing methods can only handle yaw angle variations within $[-45°, +45°]$, the proposed method can tackle the full range of yaw angle variation. Fig. 9 shows that high performance is achieved even when the yaw angle approaches $\pm 75°$.

TABLE V
RANK-1 IDENTIFICATION RATES ON COMBINED VARIATIONS OF POSE AND RECORDING SESSION ON MULTI-PIE

Methods	Alignment	080 $-45°$	130 $-30°$	140 $-15°$	050 $+15°$	041 $+30°$	190 $+45°$	Mean
VAAM [22]	Auto	74.10	91.00	95.70	95.70	89.50	74.80	86.80
SA-EGFC [18]	Manual	93.00	98.70	99.70	99.70	98.30	93.60	97.17
MRFs [8]	N/A	86.30	89.70	91.70	91.00	89.00	85.70	88.90
RL+LDA [12]	Manual	95.60	98.50	**100.0**	99.30	98.50	97.80	98.28
SPAE [32]	Auto	84.90	92.60	96.30	95.70	94.30	84.40	91.37
MDF-PM [57]	Manual	90.00	94.30	95.30	94.70	93.70	87.70	92.62
MVP+LDA [39]	Manual	93.40	**100.0**	**100.0**	**100.0**	99.30	95.60	98.05
PBPR-MtFTL	Manual	**98.67**	**100.0**	**100.0**	**100.0**	**100.0**	**98.33**	**99.50**

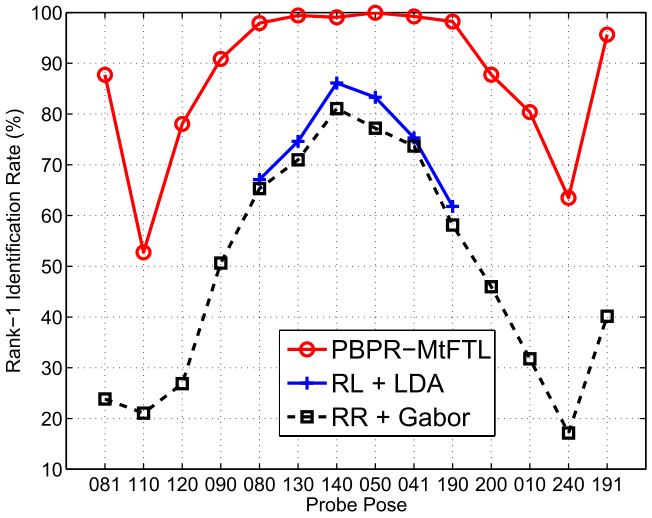


Fig. 8. Performance comparison on combined variations of pose and illumination. The probe sets 081 and 191 are with hybrid yaw and pitch variations. The other probe sets contain only yaw variations from $-90°$ to $+90°$.

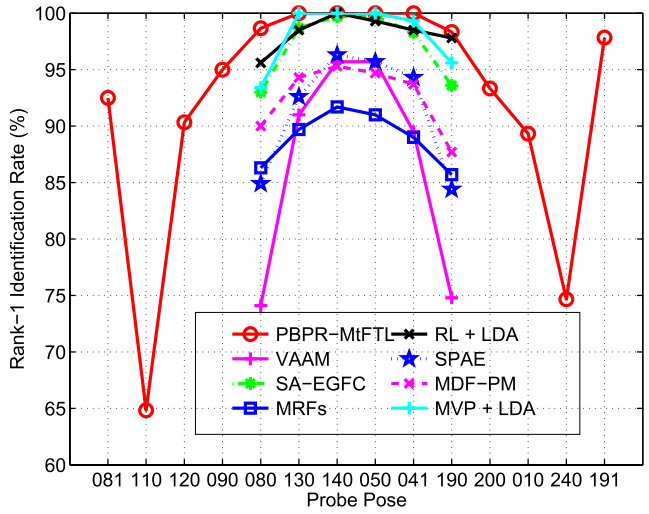


Fig. 9. Performance comparison of different methods on combined variations of pose and recording session.

E. Recognition Across Pose, Illumination, and Recording Session

To examine the robustness of the proposed algorithm under more challenging conditions, a new protocol specified in [39] is employed. This protocol extends the original protocol designed in [22] by incorporating all 20 illumination types, while the other settings remain the same. Therefore, the gallery set is exactly the same as [22], while the number of probe images is 20 times more than that in [22]. The performance of the proposed method, compared with the state-of-the-art approaches, is presented in Table VI and Fig. 10. All methods in Table VI employ manually labeled facial feature points. We make the following observations:

1) PBPR-MtFTL significantly outperforms the other three approaches across all probe sets. This result is consistent with those observed in the previous two experiments.
2) Among the four approaches, PBPR-MtFTL is the only one that can handle full range of pose variations, and its performance degrades gracefully across wide pose variations including $\pm 60°$.

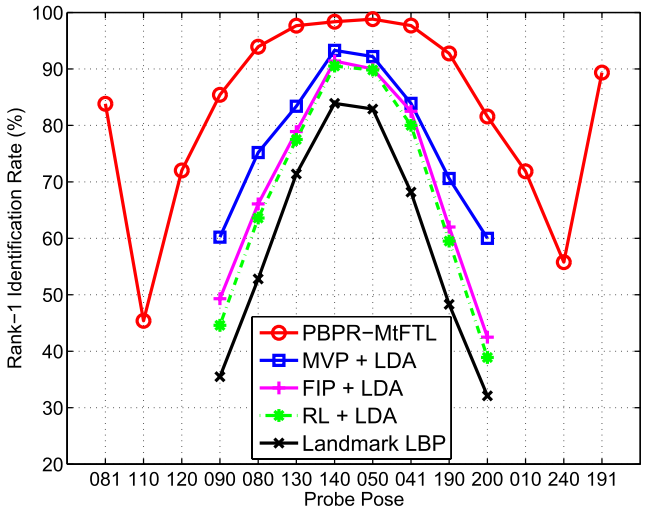


Fig. 10. Performance comparison of different methods on combined variations of pose, illumination, and recording session.

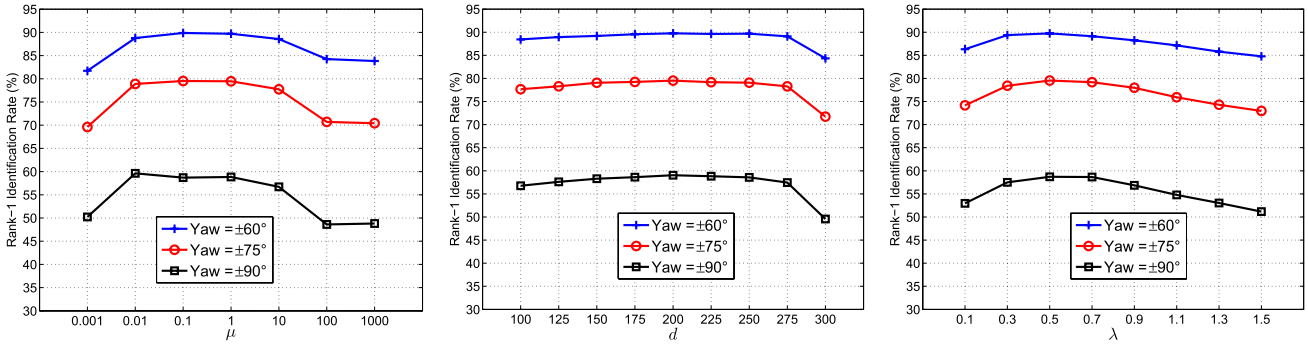
F. Parameter Evaluation for MtFTL

In the above experiments, the optimal value of model parameters μ, d, and λ is estimated on the validation subsets. In this experiment, the impact of their value on the performance of MtFTL is investigated. The same protocol [17] as the second experiment is followed. Also, the rank-1 identification rates on the three most challenging poses are reported.

TABLE VI

RANK-1 IDENTIFICATION RATES ON COMBINED VARIATIONS OF POSE, ILLUMINATION, AND RECORDING SESSION ON MULTI-PIE

Methods	090 $-60°$	080 $-45°$	130 $-30°$	140 $-15°$	050 $+15°$	041 $+30°$	190 $+45°$	200 $+60°$	Mean
Landmark LBP+LDA [59], [39]	35.50	52.80	71.40	83.90	82.90	68.20	48.30	32.10	59.39
FIP+LDA [12], [39]	49.30	66.10	78.90	91.40	90.00	82.50	62.00	42.50	70.34
RL+LDA [12], [39]	44.60	63.60	77.50	90.50	89.80	80.00	59.50	38.90	68.05
MVP+LDA [39]	60.20	75.20	83.40	93.30	92.20	83.90	70.60	60.00	77.35
PBPR-MtFTL	**85.41**	**93.93**	**97.66**	**98.36**	**98.81**	**97.68**	**92.74**	**81.58**	**93.27**



(a) (b) (c)

Fig. 11. Influence of the parameters μ, d, and λ to the performance of MtFTL. (a) evaluation against the value of μ while d and λ are set at 200 and 0.5, respectively; (b) evaluation against the value of d while μ and λ are set at 0.1 and 0.5, respectively; (c) evaluation against the value of λ while μ and d are set at 0.1 and 200, respectively.

The performance of the MtFTL approach for different value of μ, d, and λ is shown in Fig. 11. Their optimal value is around 0.1, 200, and 0.5, respectively. The experimental results also indicate that the performance of MtFTL is robust to the fluctuation of parameter value.

G. Performance in the Fully-Automatic Mode

The performance of the presented PBPR-MtFTL framework is related to the accuracy of the facial feature detection and pose estimation algorithms. The previous experiments are semi-automatic (SA), i.e., the facial feature points are labeled manually and it is assumed that the probe image poses are known. In this experiment, the PBPR-MtFTL framework is run in the fully-automatic (FA) mode. We leave the manually labeled facial feature points for the gallery and training images intact. This is reasonable since the labeling work could be conducted offline. For all probe images, the five facial feature points are automatically detected. Since existing face alignment tools cannot reliably detect facial feature points for profile or half-profile faces, we limit the yaw range of the probe images to within $\pm 45°$ in this experiment. For pose estimation, we compare the unoccluded region of each probe image with those of a set of training images whose poses are known. The pose of the probe image is assigned to that of the training image whose unoccluded region is the most similar.

The same protocol [17] as used in the second experiment is adopted. The performance of PBPR-MtFTL in the SA and FA modes is compared in Fig. 12. There is a minor drop in performance under the FA mode. In fact, the performance drop is mainly caused by the failure of face detection, whose failure rates on the six probe sets are 3.66%, 1.95%, 1.21%, 1.51%, 1.98%, and 3.72%, respectively. Besides, when considered along with the results shown in Fig. 8, the performance of PBPR-MtFTL in the FA mode is still considerably better than the state-of-the-art methods.

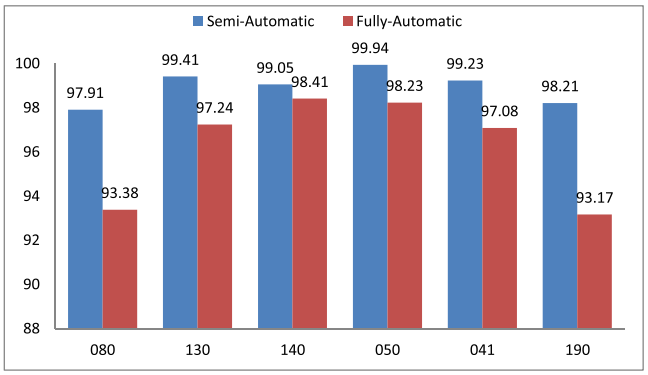


Fig. 12. Performance comparison of the proposed PBPR-MtFTL framework in the SA and FA modes. In the FA mode, both facial feature point detection and pose estimation are completely automatic. Note that the identification error in the FA mode incorporates the failure in face detection.

H. Extension to Unconstrained Face Verification

In this experiment, we slightly modify the proposed approach to tackle the unconstrained face verification problem, and present experimental results on the LFW database [9].

In Section IV, we assume that the tth task of MtFTL is to learn the feature transformation between the tth non-frontal pose type and the frontal pose. As shown in Fig. 13, image pairs defined in LFW may contain no frontal pose image. Therefore, we add tasks in the model that learn the feature transformation between every possible pair of poses.

Fig. 13. Many image pairs defined in LFW contain no frontal faces. The first line shows the first images in the image pairs, while the second line shows the second images in the image pairs.

TABLE VII
PERFORMANCE COMPARISON ON LFW WITH STATE-OF-THE-ART METHODS BASED ON SINGLE FACE REPRESENTATION

Methods	Accuracy ± Error(%)
MOSS + ITML [62]	85.17 ± 0.61
Sub-SML [63]	87.15 ± 0.56
OCLBP + LDA [64]	88.75 ± 0.59
FVF + Mah. Metric [61]	88.85 ± 2.35
PBPR + MtFTL	**91.78 ± 0.58**
LBP + PLDA [65]	87.33 ± 0.55
High-dim LBP + Joint Bayesian [59]	**93.18 ± 1.07**
PBPR + PLDA	**92.95 ± 0.37**

Another characteristic of LFW is that it has very few profile face images. To make sure that each pose type incorporates sufficient training data, we quantize the pose space into three types, i.e., left profile (LP, $yaw < -10°$), frontal pose (FP, $-10° \leq yaw \leq +10°$), and right profile (RP, $yaw > +10°$). Therefore, there are six tasks in total, i.e., LP-LP, LP-FP, LP-RP, FP-FP, FP-RP, and RP-RP. Besides, the weight of each task in Eq. 3 is modified to be proportional to the number of training samples in each task, since the training samples are far from balanced among the tasks.

For each pair of faces in LFW, both images are normalized as described in Section III. Occlusion detection is conducted for non-frontal face images. Features are extracted only from the patches that are un-occluded in both images. The features are transformed with the learnt MtFTL model, as described in Section V. Similarity scores between all un-occluded patch pairs are averaged as the similarity score of the face image pair. Since MtFTL explicitly employs the image labels, the proposed method falls in the paradigm of "Unrestricted, Label-Free Outside Data" [9]. We conduct performance comparison with the state-of-the-arts in Table VII. To promote performance, most existing methods designed for the LFW challenge fuse multiple face representations, e.g., employing several descriptors [60] or mirroring the face image [61]. In comparison, this work is not targeted at the LFW challenge and we employ only a single face representation. For fair comparison, we report the best performance of all approaches in Table VII achieved with a single face representation.[3]

[3]The performance of [61] is obtained using the code and data released by the authors, while the performance of the other approaches is directly cited from the original papers.

The first five approaches in Table VII adopt metric learning based classifiers, and PBPR-MtFTL achieves significantly better performance than the other approaches. Recently, generative model based classifiers have been introduced to the LFW challenge. We then replace the MtFTL model with the Probabilistic Linear Discriminative Model (PLDA) [66]. The dimension of the PLDA subspace is set at 100. With generative model based classifiers, the high-dim LBP approach [59] achieves a slightly higher accuracy than our approach. However, this approach relies on dense facial feature detection. We emphasize here that only the 5 most stable facial feature points are required by our method. This makes our algorithm easier to use in practical applications.

VII. CONCLUSION

Face recognition across pose is a challenging task because of the significant appearance change caused by pose variations. We handle this problem from two aspects. First, we propose the PBPR face representation scheme that makes use of the unoccluded face textures only. PBPR can be applied to face images in arbitrary pose, which is a great advantage over existing methods. Second, we present the MtFTL model for learning compact feature transformations by utilizing the correlation between poses. Clear advantage is shown compared to single-task based methods. To the best of our knowledge, this is the first time that MTL has been formally applied to the PIFR problem. As the proposed PBPR-MtFTL framework effectively utilizes all the unoccluded face texture and the correlation between different poses, very encouraging results for face identification in all three popular multi-pose databases are achieved. We also slightly modify the proposed approach to tackle the unconstrained face verification problem, and achieve top level performance on the challenging LFW database.

APPENDIX A
PROOF OF THEOREM 1

Proof: We can partition $\mathcal{Z}$ into K disjoint sets, so that if two face patch features s and z are close, then

$$\|s - z\| \leq \gamma \quad \text{and} \quad |\theta_s - \theta_z| \leq \Delta_\theta. \tag{12}$$

By arranging the loss functions so that the first loss is always larger than the second one, we therefore have

$$\begin{aligned}
&|\ell(A_s, U, s) - \ell(A_z, U, z)| \\
&= \|A_s U^T (s - x_0)\| - \|A_z U^T (z - x_0)\| \\
&= \|A_s U^T (s + z - z - x_0)\| - \|A_z U^T (z - x_0)\| \\
&\leq \|A_s U^T (s - z)\| + \|A_s U^T (z - x_0)\| - \|A_z U^T (z - x_0)\| \\
&\leq \|A_s U^T (s - z)\| + \|(A_s - A_z + A_z) U^T (z - x_0)\| \\
&\quad - \|A_z U^T (z - x_0)\| \\
&\leq \|A_s U^T (s - z)\| + \|(A_s - A_z) U^T (z - x_0)\| \\
&\leq \|A_s\| \|s - z\| + \|A_s - A_z\| \|z - x_0\| \\
&\leq \sqrt{d}\gamma + \Omega_{\Delta_\theta} \gamma_0,
\end{aligned}$$

which completes the proof. ■

APPENDIX B
PROOF OF THEOREM 2

Proof: Let N_i be the set of index of points of $\mathbf{s}$ that fall into C_i. $(|N_1|, \cdots, |N_K|)$ is an IID random variable with parameters n and $(\beta(C_1), \cdots, \beta(C_K))$. We have

$$
\begin{aligned}
&\left|\mathcal{L}(\mathcal{A}_s) - \mathcal{L}_{emp}(\mathcal{A}_s)\right| \\
&= \left|\sum_{i=1}^{K} E_{z\sim\beta}(\ell(\mathcal{A}_s, z)|z \in C_i)\beta(C_i) - \frac{1}{n}\sum_{i=1}^{n}\ell(\mathcal{A}_s, s_i)\right| \\
&\leq \left|\sum_{i=1}^{K} E_{z\sim\beta}(\ell(\mathcal{A}_s, z)|z \in C_i)\frac{N_j}{n} - \frac{1}{n}\sum_{i=1}^{n}\ell(\mathcal{A}_s, s_i)\right| \\
&\quad + \left|\sum_{i=1}^{K} E_{z\sim\beta}(\ell(\mathcal{A}_s, z)|z \in C_i)\beta(C_i)\right. \\
&\qquad \left. - \sum_{i=1}^{K} E_{z\sim\beta}(\ell(\mathcal{A}_s, z)|z \in C_i)\frac{N_j}{n}\right| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^{K}\sum_{j\in N_i}\max_{z\in C_i}|\ell(\mathcal{A}_s, s_j) - \ell(\mathcal{A}_s, z)|\right| \\
&\quad + \left|\max_{z\in\mathcal{Z}}|\ell(\mathcal{A}_s, z)|\sum_{i=1}^{K}|\frac{|N_i|}{n} - \beta(C_i)|\right| \\
&\leq \epsilon(s) + B\sum_{i=1}^{K}|\frac{N_i}{n} - \beta(C_i)| \\
&\leq \epsilon(s) + B\sqrt{\frac{2K\ln 2 + 2\ln(1/\delta)}{n}}.
\end{aligned}
$$

The first inequality is due to the triangle inequality, and the second inequality is because of $\sum_{i=1}^{K}\beta(C_i) = 1$ and $\sum_{i=1}^{K}\frac{N_j}{n} = 1$. Finally, the last inequality is the application of Proposition 1. ∎

ACKNOWLEDGMENT

The authors would like to thank the associate editor Prof. Shiguang Shan and the three anonymous reviewers for their careful reading and valuable remarks, which have contributed to improving the quality of the paper. The authors would also like to thank Prof. Thomas Vetter in the Department of Computer Science, the University of Basel to provide the BFM model.

REFERENCES

[1] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[2] C. Ding, J. Choi, D. Tao, and L. S. Davis. (2014). "Multi-directional multi-level dual-cross patterns for robust face recognition." [Online]. Available: http://arxiv.org/abs/1401.5311

[3] M. Gunther *et al.*, "The 2013 face recognition evaluation in mobile environment," in *Proc. IEEE/IAPR Int. Conf. Biometrics*, Jun. 2013, pp. 1–7.

[4] R. Abiantun, U. Prabhu, and M. Savvides, "Sparse feature extraction for pose-tolerant face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2061–2073, Oct. 2014.

[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[6] S. J. D. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 970–984, Jun. 2008.

[7] S. R. Arashloo and J. Kittler, "Energy normalization for pose-invariant face recognition based on MRF model image matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1274–1280, Jun. 2011.

[8] H. T. Ho and R. Chellappa, "Pose-invariant face recognition using Markov random fields," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1573–1584, Apr. 2013.

[9] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.

[10] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 454–461.

[11] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[12] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 113–120.

[13] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1193–1205, May 2013.

[14] R. Weng, J. Lu, J. Hu, G. Yang, and Y.-P. Tan, "Robust feature set matching for partial face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 601–608.

[15] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognit.*, vol. 42, no. 11, pp. 2876–2896, 2009.

[16] A. B. Ashraf, S. Lucey, and T. Chen, "Learning patch correspondences for improved viewpoint invariant face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[17] A. Li, S. Shan, and W. Gao, "Coupled bias–variance tradeoff for cross-pose face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 305–315, Jan. 2012.

[18] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan, "Morphable displacement field based image matching for face recognition across pose," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 102–115.

[19] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, 2014.

[20] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, Sep. 2013.

[21] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3D generic elastic models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1952–1961, Oct. 2011.

[22] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 937–944.

[23] Y. Zhang, M. Shao, E. K. Wong, and Y. Fu, "Random faces guided sparse many-to-one encoder for pose-invariant face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2416–2423.

[24] J. Wright and G. Hua, "Implicit elastic matching with random projections for pose-variant face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1502–1509.

[25] D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3539–3545.

[26] A. Li, S. Shan, X. Chen, and W. Gao, "Maximizing intra-individual correlations for face recognition across pose differences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 605–611.

[27] A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 593–600.

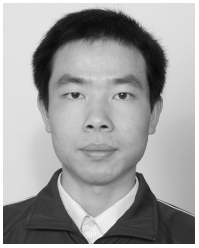
[28] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2160–2167.

[29] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 808–821.

[30] L. Zhang, X. Zhen, and L. Shao, "Learning object-to-class kernels for scene classification," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3241–3253, Aug. 2014.

[31] R. Yan, L. Shao, and Y. Liu, "Nonlocal hierarchical dictionary learning using wavelets for image denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4689–4698, Dec. 2013.

[32] M. Kan, S. Shan, H. Chang, and X. Chen, "Stacked progressive auto-encoders (SPAE) for face recognition across poses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1883–1890.
[33] Z. Hong, X. Mei, D. Prokhorov, and D. Tao, "Tracking via robust multi-task multi-view joint sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 649–656.
[34] B. Mahasseni and S. Todorovic, "Latent multitask learning for view-invariant action recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3128–3135.
[35] L. Liu, L. Shao, and P. Rockett, "Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition," *Pattern Recognit.*, vol. 46, no. 7, pp. 1810–1818, 2013.
[36] F. Zhu and L. Shao, "Weakly-supervised cross-domain dictionary learning for visual recognition," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 42–59, 2014.
[37] D. Masip, Á. Lapedriza, and J. Vitrià, "Multitask learning: An application to incremental face recognition," in *Proc. Int. Conf. Comput. Vis. Appl.*, 2008, pp. 585–590.
[38] D. Masip, Á. Lapedriza, and J. Vitrià, "Multitask learning applied to face recognition," in *Proc. 1st Spanish Workshop Biometrics*, 2007, pp. 1–8.
[39] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Multi-view perceptron: A deep model for learning face identity and view representations," in *Advances in Neural Information Processing Systems 27*. Red Hook, NY, USA: Curran Associates, 2014, pp. 217–225.
[40] Z.-L. Sun, K.-M. Lam, and Q.-W. Gao, "Depth estimation of face images using the nonlinear least-squares model," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 17–30, Jan. 2013.
[41] A. M. Bruckstein, R. J. Holt, T. S. Huang, and A. N. Netravali, "Optimum fiducials under weak perspective projection," *Int. J. Comput. Vis.*, vol. 35, no. 3, pp. 223–244, 1999.
[42] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[43] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[44] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3656–3670, Aug. 2014.
[45] M. Schmidt. *The minFunc Package*. [Online]. Available: http://www.cs.ubc.ca/schmidtm/Software/minFunc.html, accessed Jul. 2014.
[46] H. Xu and S. Mannor, "Robustness and generalization," *Mach. Learn.*, vol. 86, no. 3, pp. 391–423, 2012.
[47] A. W. van der Vaart and J. A. Wellner, *Weak Convergence*. New York, NY, USA: Springer-Verlag, 1996.
[48] 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.
[49] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[50] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[51] L. Liu, L. Shao, and X. Li, "Evolutionary compact embedding for large-scale image classification," *Inf. Sci.*, pp. 1–15, Jul. 2014.
[52] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. AVSS*, Sep. 2009, pp. 296–301.
[53] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3476–3483.
[54] S. R. Arashloo, J. Kittler, and W. J. Christmas, "Pose-invariant face recognition by matching on multi-resolution MRFs linked by super-coupling transform," *Comput. Vis. Image Understand.*, vol. 115, no. 7, pp. 1073–1083, 2011.
[55] S. R. Arashloo and J. Kittler, "Fast pose invariant face recognition using super coupled multiresolution Markov random fields on a GPU," *Pattern Recognit. Lett.*, vol. 48, pp. 49–59, Oct. 2014.
[56] S. R. Arashloo and J. Kittler, "Efficient processing of MRFs for unconstrained-pose face recognition," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl., Syst.*, Sep./Oct. 2013, pp. 1–8.
[57] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan, "Maximal likelihood correspondence estimation for face recognition across pose," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4587–4600, Oct. 2014.
[58] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.
[59] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3025–3032.
[60] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3499–3506.
[61] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proc. Brit. Mach. Vis. Conf.*, 2013, vol. 1, no. 2, pp. 1–12.
[62] Y. Taigman, L. Wolf, and T. Hassner, "Multiple one-shots for utilizing class label information," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–12.
[63] Q. Cao, Y. Ying, and P. Li, "Similarity metric learning for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2408–2415.
[64] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1960–1967.
[65] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S. J. D. Prince, "Probabilistic models for inference about identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 144–157, Jan. 2012.
[66] S. J. D. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.



Changxing Ding received the B.E. degree in automation from the Shandong University of Science and Technology, Qingdao, China, and the M.E. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology at Sydney, Sydney, NSW, Australia. His research interests include computer vision, machine learning, and in particular, on face recognition.

Chang Xu received the B.E. degree from Tianjin University, Tianjin, China, in 2011. He is currently pursuing the Ph.D. degree with the Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China. He was a Research Intern with the Knowledge Mining Group, Microsoft Research Asia, Beijing, and a Research Assistant with the Center of Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology at Sydney, Sydney, NSW, Australia. He received the best student paper award in ACM ICIMCS 2013. His research interests lie primarily in machine learning, multimedia search, and computer vision.

Dacheng Tao (F'15) iis currently a Professor of Computer Science with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology at Sydney, Sydney, NSW, Australia. He mainly applies statistics and mathematics to data analytics, and his research interests spread across computer vision, data science, image processing, machine learning, neural networks, and video surveillance. His research results have expounded in 1 monograph and over 100 publications at prestigious journals and prominent conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the *Journal of Machine Learning Research*, the *International Journal of Computer Vision*, the Workshop on Neural Information Processing Systems, the International Conference on Machine Learning, the Conference on Computer Vision and Pattern Recognition, the International Conference on Computer Vision, the European Conference on Computer Vision, the International Conference on Artificial Intelligence and Statistics, and the International Conference on Data Mining, and the ACM Conference on Knowledge Discovery and Data Mining, with several best paper awards, such as the Best Theory/Algorithm Paper Runner-Up Award in the IEEE ICDM'07, the best student paper award in the IEEE ICDM'13, and the 2014 ICDM 10-Year Highest Paper Award.