Pattern Recognition 77 (2018) 426–437

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

From one to many: Pose-Aware Metric Learning for single-sample face recognition

Weihong Deng*, Jiani Hu, Zhongjun Wu, Jun Guo

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 10 September 2016
Revised 26 September 2017
Accepted 16 October 2017
Available online 17 October 2017

Keywords:
Face recognition
Single sample per person
Metric learning
3D generic elastic model
Face re-rendering
3D face construction

ABSTRACT

Pose and illumination variations are very challenging for face recognition with a single sample per person (SSPP). In this paper, we address this issue by a Pose-Aware Metric Learning (PAML) approach. Our primary idea is "*from one to many*": Synthesizing many images of sufficient pose and illumination variability from the single training image, based on which metric learning approach is applied to reduce these "synthesized" variations at each quantified pose. For this purpose, given a single frontal training image, a multi-depth generic elastic model and an extended generic elastic model are developed to synthesize facial images of the target pose with varying 3D shape (depth) and illumination variations respectively. To reduce these "synthesized" variability, Pose-Aware Metric spaces are separately learnt by linear regression analysis at each quantized pose, and pose-invariant recognition is performed in the corresponding metric space. By preserving the detailed texture and reducing the shape variability, the PAML method achieves an 100% accuracy on the Multi-PIE database under the test setting across poses, which is significantly better than the traditional methods that use a large generic image ensemble to learn the cross-pose transformations. On the more challenging setting across both poses and illuminations, PAML outperforms the recent deep learning approaches by over 10% accuracy.

1. Introduction

We consider the pose-invariant face recognition problem with a single training sample per person. This single-sample face recognition (SSFR) problem is one of the major challenges in many real-world applications on law enforcement and homeland security [1]. Theoretically, it is an extreme case of the small sample size problem that deteriorates conventional pattern recognition techniques. As the supervised learning techniques are not applicable without intraclass information, unsupervised techniques, which find the low-dimensional embedding of the gallery data by ICA [2], PCA [3] or their variants [4–6], have been widely applied. However, these methods are suitable only for face representation and effective only for the recognition under constraint variations. Invariant features (e.g. Gabor feature [7,8] and local binary patterns [9]) are effective to increase the robustness to the lighting and expression changes. Unfortunately, since they discard all information about the 3D layout of the face, these feature descriptors are deficient to counteract the unobserved pose variations.

Pose variation is widely regarded as a major challenge in the automatic face recognition application. We envision the typical applications where enrollment of subjects is through frontal images with neutral light and expression (i.e., typical enrollment images for most applications), but test images come from real-world unconstrained scenarios with various poses and illuminations. Since the face images under variable pose reside in a highly nonlinear subspace, conventional subspace learning [8,10], manifold learning [11], sparse representation [12–14] and metric learning [15,16] methods designed for SSFR can not achieve satisfactory performance. Previous SSFR approaches across pose differences mostly rely on a (external) generic training set with multiple samples per person (MSPP) of similar viewpoints to the test samples [17,18]. Recently, deep learning techniques are used to learn the cross-pose transformation for the unconstrained face based on an external multi-view image ensemble [19]. Unfortunately, the performance of these methods depends heavily on the representativeness of the generic training set, though they achieve state-of-the-art performance on the Multi-PIE database [20] with the same training and testing viewpoints. In contrast, the 3D model based methods [21,22] is more flexible, since they do not rely on the similarity of the training and testing viewpoints.

In this paper, we aim to address the pose-invariant SSFR problem by exploring the discriminative information in the gallery images, by the extensions of 3D generic elastic model [23,24]. In general, our primary idea is "*from one to many*": Synthesizing many

* Corresponding author.
E-mail addresses: whdeng@bupt.edu.cn, cvpr_dwh@126.com (W. Deng).

https://doi.org/10.1016/j.patcog.2017.10.020
0031-3203/

images of the pose and illumination variability from one single (frontal) gallery image, based on which the metric learning approach further reduces the "synthesized" variations at each quantified pose. Following this idea, the contributions of paper are as follows.

Firstly, we develop a multi-depth 3D generic elastic model (MD-GEM) with variable depth to characterize the uncertainty of the 3D shape [25], when rendering faces of different poses from a single frontal image. To address this problem caused by the depth ambiguity of the face, we make a linear assumption on the depth channel of 3D generic elastic model with a single parameter, instead of a single settled depth map of the conventional GEM [24].

Secondly, we propose an extended generic elastic model (E-GEM) [26] that couples the 3D generic elastic model with the quotient image [27] technique to synthesize faces under different poses and illumination conditions from a single frontal images. Specifically, we develop a shape-free alignment for the quotient image to achieve better face re-rendering results. This adaptive quotient image (AQI) is then used to generate the texture surface of GEM to render varying lightings.

Thirdly, inspired by the "divide and conquer" strategy, we address the pose-invariant face recognition problem by Pose-Aware Metric Learning (PAML) using the synthesized training images of each quantized pose separately. For each quantized pose, PAML applies linear regression analysis technique [16] to transform the synthesized training samples of one subject into a single point of the metric space. In the recognition stage, we first estimates the pose of probe face and then applies the corresponding pose-specific metric to perform classification.

By the virtue of GEM, the proposed PAML method can take advantage of the full texture details of the gallery image under arbitrary poses, which is essential for the highly accurate recognition. Extensive experiments on the Multi-PIE database [20] demonstrates that our method is a superior SSFR solution for variable pose, without using training set of external subjects to learn the pose-invariant transformation. Specifically, on the test setting under variable pose, the PAML method, based on the LBP descriptor of the synthesized images via MD-GEM, achieves 100% accuracy on the MPIE database. On the test setting across both poses and illuminations, PAML, based on the LBP descriptor of the synthesized images via E-GEM, outperforms the recent deep learning methods by over 10% accuracy. Moreover, PAML does not rely on any external data for training, while existing methods use a large generic image ensemble of hundreds of people to learn the pose variations.

It should be noted that this paper is an extended work of our previous conference papers [25,26]. In this paper, we present the comprehensive related works, more technical details on MD-GEM [25] and E-GEM [26], and a combined face synthesis algorithm (Algorithm 1). Moreover, we also integrate them into the proposed PAML framework to obtain much better performance than our previous work. In particular, we demonstrate the 100% accuracy of MD-GEM on the additional pose-invariant recognition experiments of MPIE database.

Algorithm 1 Image synthesis via MD-GEM and E-GEM.

Input: A frontal facial image, a 3D generic model with depth parameter α, a bootstrap set of images with a unified shape, target pose angles

Output: The ensemble of synthesized images at target poses

1: Locate the 77 feature points by off-the-shelf face alignment method or manual labeling.
2: Compute dense correspondence between the input image and 3D-GEM according to the delaunay triangulation of feature points, and then allocate the depth according to the parameter α.
3: Transform the input face to the unified shape of the bootstrap set. Solve Q and light coefficient x_j according to Eq. 4, and then stimulate each illumination by setting lighting coefficient l_j. Finally, transform the re-rendered facial images back to the original shape.
4: Map the texture from the re-rendered images with various illuminations to the 3D-GEMs according to the dense correspondence obtained in Step 2.
5: At each target pose, render the synthesized images with the 3D-GEMs of input depth (α) and lighting (l_j) parameters.

Algorithm 2 Pose-Aware Metric Learning.

Input: The training set with a frontal training image per person, quantized pose angles

Output: The transformation matrix $W^{(p)}$ for each quantized pose

1: Synthesize a predefined number of images at all quantized poses for each frontal gallery image using Algorithm 1.
2: Align all the synthesized images by the predefined landmarks. Extract feature vectors of the aligned faces.
3: At each quantized pose, compute the transformation matrix $W^{(p)}$ in Eq. 7 using the feature matrix of the synthesized images at that pose.

2. Background

Many interesting improvements on face recognition have been reported in the literature to handle pose variation. Robust feature descriptors are expected to be more robust than pixel intensity to counteract the appearance change caused by poses. LGBP [28] is a high-dimensional face descriptor which first convolves the images by a family of Gabor kernels followed by LBP coding of the filtered images. LE+LDA [29] method encodes the micro-structures of the face by a new learning-based encoding method, which can automatically achieve very good tradeoff between discriminative power and invariance. CRBM+LDA [30] learns the local descriptor by local convolutional restricted Boltzmann machines, which exploits the global structure and maintains the robustness to small misalignments.

Several statistical learning methods are proposed by leveraging the correlation among features across poses. CCA [31] aims at projecting the images of different poses onto a common feature space where the correlation between them are maximized. PLS method [32] attempts to project samples from two poses to a common latent subspace, with one pose as regressor and the other pose as response. GMA method [33] is a generalized multi-view analysis method attempting to project the images of all poses to a discriminative common space, where pose variations are minimized. Li et al. [18] represented a test face as a linear combination of training images and utilized the linear regression coefficients as features for face recognition.

3D model based methods provide straightforward solutions for pose-invariant recognition. The 3D morphable model [34] is built using PCA on 3D facial shapes and textures acquired from a laser scanner. The learned 3D face model is reconstructed by fitting the model to the input 2D image. Pose-invariant recognition can be performed by transforming posed face images to the frontal view or comparing the reconstruction coefficients of PCA. However, the PCA subspace may not be accurate enough to characterize the detailed textures of test faces. Besides 3DMM, several 3D model based methods are proposed to rotate the non-frontal face to the frontal one. Different from 3DMM, VAAM method [35] proposes a fully automatic 3D pose normalization method, which can synthesize a frontal view by aligning an average 3D model to the input non-frontal face based on the view-based AAM. We recently

improve this work by recovering the lighting condition of the occluded face part [21]. MDF method [18] generates a virtual image at the pose of the gallery image for the probe image through the Morphable Displacement Field, and then matches the synthesized face with the gallery faces. StackFlow method [36] warps a non-frontal face image to the fontal one progressively through one or more correspondences between them at the patch level.

Recently, deep learning approaches have achieved premier accuracy on recognition by learning cross-pose transformation. DAE method [37] learns pose-robust features by modeling the complex non-linear transform from the non-frontal face images to frontal ones through a single deep auto-encoder. Further, SPAE method [38] proposes to learn the non-linear transformation from the non-frontal faces to the frontal faces in a progressive way, in which each stack learns the transformation across a small angle in a supervised manner. The face identity-preserving (FIP) features [39] are learned by a deep network that combines the feature extraction layers and the reconstruction layers. The former layers encode a face image into the pose-invariant FIP features, while the latter transforms them to an image in the canonical view. RL+LDA method [39] further improves the performance by applying local descriptors and LDA to the frontal reconstructed images. Recently, a multi-view perceptron (MVP) is proposed to untangle the identity and pose by using random hidden neurons. CPF method [19] is a recent work which can rotate the input face to a target-pose face image with a multi-task deep neural network.

In the past, it is believed that the true 3D shape must be estimated faithfully by 3D Morphable model [34] for the pose-invariant recognition task. Unfortunately, the reconstruction process is unstable for the single image with natural lighting. 3D generic elastic model (GEM) is introduced in [23,24] as a low computational but efficient 3D modeling method. The underlying assumption is that face depth information does not dramatically change among individuals as long as the corresponding 2D face feature points are aligned. Although this assumption seems strong, experimental results show that coarse face shape is good enough to achieve reasonable results. In this work, we improve the basic GEM in order to address the SSFR problem across compound variations of poses and illuminations. An advantage of GEM is the capability to preserve original 2D texture in the synthesized images, so that the local texture can be accurately discriminated. Pose-Aware Metric Learning is developed to address the recognition problem of each quantized pose separately, because a single linear subspace cannot characterize the images under variable pose.

3. GEM extensions with shape and illumination variability

To address the problem on image synthesis **from a single frontal training face**, this section describes two extensions of 3D generic elastic model: (1) multi-depth GEM (MD-GEM) with shape (depth) variations and (2) extended GEM (E-GEM) with illumination variations. MD-GEM improves the GEM model by transferring a single depth map to variable ones with a "depth" parameter. E-GEM combines the 3D generic elastic model (pose synthesis) and the quotient image (lighting synthesis) [27] together to simultaneously address the lighting and pose synthesis .

Due to the space limits, some related techniques, such as dense correspondence and quotient image, may only be briefly introduced, and the readers can refer to Prabhu and co-workers [24,27,40] for the details of the background knowledge.

3.1. Multi-depth GEM with a single control parameter [25]

3D generic elastic model provides a practical method for generating the 3D model from a single frontal face according to a generic 3D facial depth map [24]. In this model, 2D (u, v) information needs to be extracted and depth information can be recovered by morphing a depth-map based on the 2D facial observations. Given an input face and a generic 3D model, there are two stages to recover the 3D model of the input face: *dense correspondence* and *depth allocation* [24]. The dense correspondence step finds out the pixel-wise correspondence from the input 2D image to the 3D reference model, and the depth allocation step simply allocates the depth information from the reference model to the corresponding location of 2D image. In the following, we describe the procedures of each step.

Dense correspondence: Firstly, sparse facial landmarks are detected and each face (I) is partitioned into triangular polygons (P) by Delaunay Triangulation. Similarly, the generic depth-model $(\bar{D})$ is partitioned into a mesh (M) from the predefined landmark points. The corresponding points of input image and generic depth map are registered by means of the landmarks as illustrated in Fig. 1. The point density increases simultaneously with loop subdivision, which can be considered as a process of establishing dense correspondence between the input mesh and the depth model. In our implementation, 77 fiducial facial landmarks of the input frontal face are automatically detected by SIFT feature based STASM [41]. About 17500 dense vertices and 35000 triangles are obtained after 4 times of loop subdivisions. The procedures are similar to those used in [24,40], which are illustrated in Fig. 1.

Depth Allocation: The prior depth-model function (F), sampled at the spatial locations of mesh (M), is warped onto the input triangle mesh (P) by a piece-wise affine transformation for the purpose of depth estimation. In this manner, each point in the input image has an exact corresponding point in the depth-model. Based on this correspondence, we allocate the depth information for the 3D shape of the input face. Finally, the texture information of the input image $I(P(u, v))$, sampled at the spatial locations of $P(u, v)$, is mapped onto the 3D shape. The variables for reconstructing 3D face model can be represented by:

$$S = (u, v, z = F(M(\tilde{u}, \tilde{v}))) \tag{1}$$

$$T = I(P(u, v)) = (R_{u,v,z}, G_{u,v,z}, B_{u,v,z}) \tag{2}$$

where $(\tilde{u}, \tilde{v})$ in M is the registered points (u, v) in image P, S and T are the shape and texture of the 3D model for the input face.

Different from the GEM with a single depth-map in GEM [24], we attempt to generate multiple realistic depth-maps of the human face for a single frontal image. The new model is called multi-depth GEM (MD-GEM). Specifically, we assume that *the depth map of a specific face is a linearly stretched or flatten version of the average depth map of large populations*. The multi-depth GEM offers the depth information with the function $F(M(\tilde{u}, \tilde{v}))$ of the spatial location $(\tilde{u}, \tilde{v})$ of the mesh M. Referring to the depth map in [24], the depth-map function $F(M(\tilde{u}, \tilde{v}))$ is defined as a linear function of a depth parameter α as follows.

$$F(M(\tilde{u}, \tilde{v})) = z_{ref} + \alpha(\bar{D}(M(\tilde{u}, \tilde{v}))) - z_{ref}) \tag{3}$$

where z_{ref} is the depth value of a fixed reference point (We empirically select the corner of the nose).

Fig. 2(a) shows six MD-GEMs with the varying depth parameter α that decides the depth of facial surface. The setting $\alpha = 1$ equals to the conventional GEM based on the average depth map [24]. The setting $\alpha > 1$ or $\alpha < 1$ generates linearly stretched or flattened depth map of average face in z axis, respectively. Fig. 2(b–d) show six different generic depth maps with $\alpha = \{1.1, 1.0, 0.9, 0.8, 0.7, 0.6\}$, and one can see from the figure that natural face surfaces can be generated by controlling a single parameter α. Example rendered images of varying α-depth maps are shown in Fig. 2, which suggests that the plausible 3D structures of

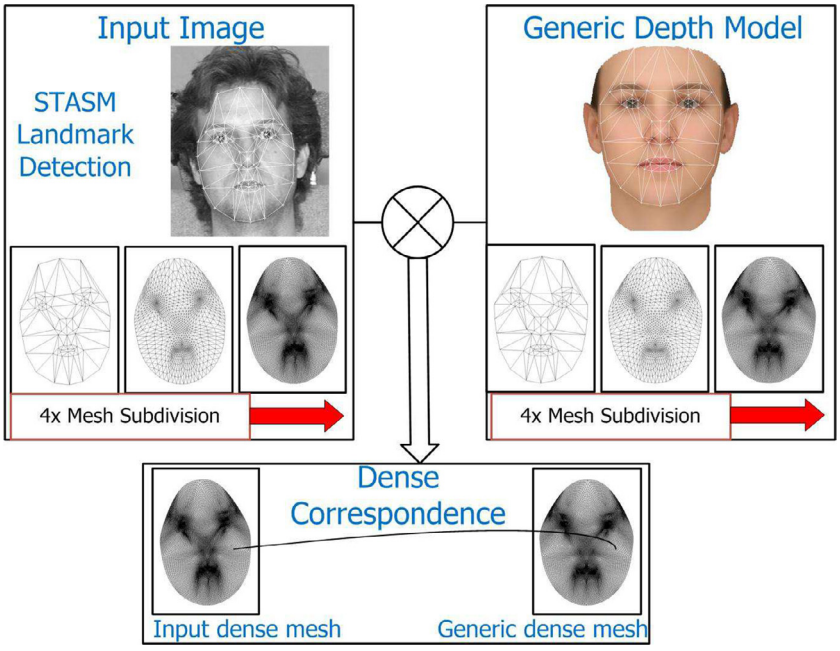

Fig. 1. The dense correspondence establishes the pixel-wise correspondence from the input 2D image to the 3D reference model. It is determined by the loop subdivision of the triangular polygons. Specifically, based on the detected shape (landmarks), both the face image and generic model are partitioned into a mesh of triangular polygons. After registering points between the input image and the generic depth-map, one can increase the point density simultaneously using Loop subdivision. Dense correspondence between the input mesh and the depth-model is established after 4 times of loop subdivisions.

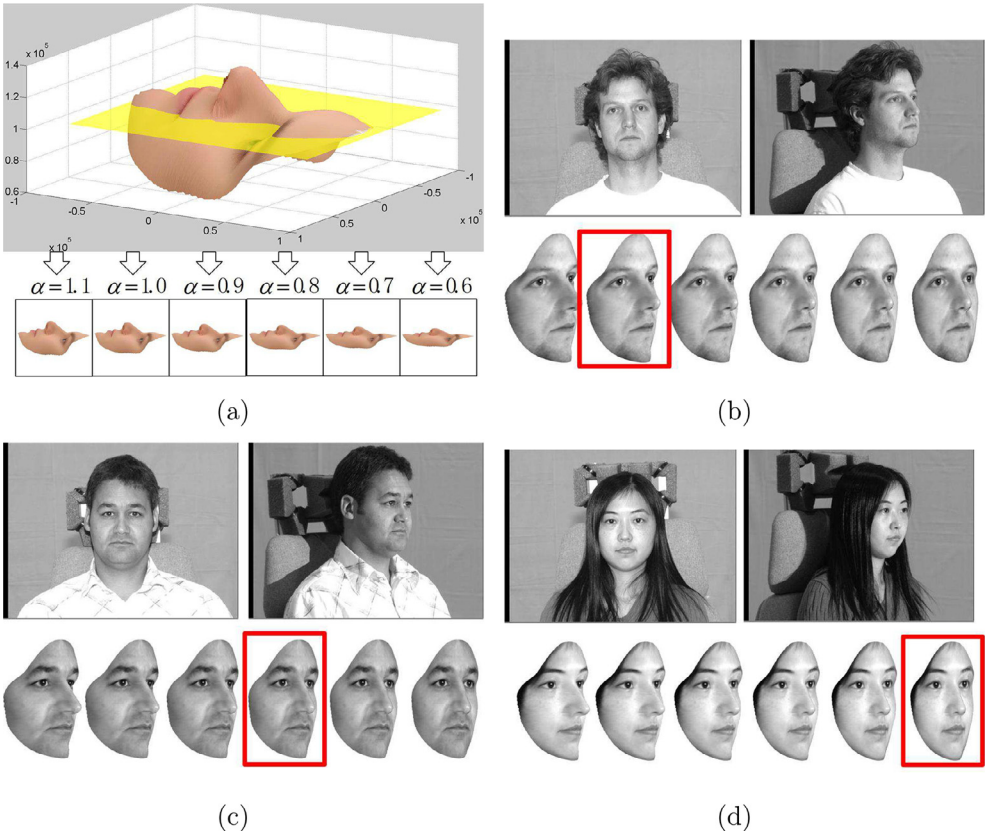

Fig. 2. We design an α depth function to generalize the generic elastic model to describe variable face surfaces. (a) Illustration of the α depth function that decides the depth of the generic depth-map. When $\alpha < 1$, the generic depth-map becomes flatter. Otherwise, when $\alpha > 1$, the generic depth-map becomes deeper. Given a single frontal image, it is difficult to infer the accurate 3D model for each face, but our α-depth function can derive several realistic 3D model with certain values of α. (b–d) show example images of three faces from the frontal view and the yaw angle of 45°. (b) $\alpha = 1.0$ is most realistic, (c) $\alpha = 0.8$ is most realistic, (d) $\alpha = 0.6$ is most realistic.

Table 1
Comparative recognition rates under single-session testing of the Multi-PIE database. "✓" indicates the method needs to learn from the additional 100 subjects of the Multi-PIE; " - ", otherwise, indicates the method is database independent and may generalize well on complex environment.

Methods	−45°	−30°	−15°	+15°	+30°	+45°	Avg	Train
VAAM [35]	74.1	91.0	95.7	95.7	89.5	74.8	86.8	-
MDF [18]	78.7	94.0	99.0	98.7	92.2	81.8	90.7	✓
PLS [32]	51.1	76.9	88.3	88.3	78.5	56.5	73.3	✓
CCA [31]	53.3	74.2	90.0	90.0	85.5	48.2	73.5	✓
GMA [33]	75.0	74.5	82.7	92.6	87.5	65.2	79.6	✓
DAE [37]	69.9	81.2	91.0	91.9	86.5	74.3	82.5	✓
SPAE [38]	84.9	92.6	96.3	95.7	94.3	84.4	91.4	✓
RR [18]	97.0	97.0	**100**	**100**	97.0	92.0	96.8	✓
RL+LDA [39]	97.8	98.6	**100**	**100**	98.6	98.4	98.4	✓
PAML (ours)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	-

Table 2
Comparative recognition rates under multi-sessions testing of Multi-PIE database. "✓" indicates the method needs to learn from the additional 200 subjects of the MultiPIE; " - ", otherwise, indicates the method is database independent and may generalize well on complex environment.

Methods	−45°	−30°	−15°	+15°	+30°	+45°	Avg	Train
LGBP [28]	37.7	62.5	77	83	59.2	36.1	59.3	-
VAAM [35]	74.1	91	95.7	95.7	89.5	74.8	86.9	-
MDF [18]	93	98.7	99.7	99.7	98.3	93.6	97.2	✓
LE+LDA [29]	86.9	95.5	99.9	99.7	95.5	81.8	93.2	✓
CRBM+LDA [30]	80.3	90.5	94.9	96.4	88.3	75.2	87.6	✓
FIP+LDA [39]	93.4	95.6	**100**	98.5	96.4	89.8	95.6	✓
RL+LDA [39]	95.6	98.5	**100**	99.3	98.5	97.8	98.3	✓
MVP [46]	93.4	**100**	**100**	**100**	99.3	95.6	98.1	✓
PAML (ours)	**98.3**	99.3	**100**	**100**	**100**	**98.3**	**99.3**	-

real facial images can be characterized by the simple α-depth function. Each face has a certain α that is realistic to the corresponding 3D shape. Although one can not infer the proper α from a single frontal image, we attempt to limit the variability by mapping the synthetic images of varying α together into the feature space for accurate recognition.

3.2. Extended GEM with illumination variability [26]

In order to synthesize the illumination changes over 3D generic elastic model, we develop an Extended GEM (E-GEM) that overlays the texture of the quotient images [27] on the surface of the 3D generic model to characterize the compound variation of poses and illuminations. Quotient Image method [27] is a classical technique of face re-lighting. As a class of object, human face is considered as Lambertian surface with a reflection function: $\rho(u, v)n(u, v)^T s$, where $0 \le \rho(u, v) \le 1$ is the surface reflectance (gray-level) associated with point u, v in the image, $n(u, v)$ is the normal direction of the surface associated with point u, v in the image, and s is the light source direction (point light source) and whose magnitude is the light source intensity. In [27], the assumption on *Ideal Class of Object*, i.e., objects that have the same shape but differ in surface albedo, is defined. Under this assumption, the *Quotient Image* $Q_y(u, v)$ of face y against face a is defined: $Q_y(u, v) = \frac{\rho_y(u,v)}{\rho_a(u,v)}$, where u, v range over the image. Thus, Q_y depends only on the relative surface texture information and is independent of illumination.

The derivation of quotient image is based on a bootstrap set (reference set) consisting of L (L is small) faces under M unknown independent illumination (totally $M \times L$ images). Given this boostrap set, the quotient image Q_y of an input image $Y(u, v)$ can be computed as

$$Q_{y(u,v)} = \frac{Y(u, v)}{\sum_{j=1}^{M} \bar{A}_j(u, v) c_j}, \tag{4}$$

where $\bar{A}_j(u, v)$ is the average image under illumination j of the bootstrap set and c_j can be derived from the bootstrap set images and the input image $Y(u, v)$. See [27] for more details of deriving c_j. According to the quotient image Q_y computed by Eq. (4), the image space created by the input face, under all possible illuminations, is spanned by

$$y_s = Q_y \otimes \left(\sum_j \bar{A}_j l_j \right) \tag{5}$$

where $\otimes$ denotes the Cartesian product(pixel by pixel multiplication). In our experiments, we utilize the "one-hot" coefficient $l_j = 1$ for $j = 1, \ldots, M$ to simulate each illumination on the input face respectively.

The basic assumption of quotient image is that all involved objects share the same shape. Although human faces share similar global shape, they still have non-negligible local individual shape variations. The commonly used face alignment process in [27] (global affine transformation) can not well satisfy the definition of *Ideal Class of Object*. The assumption on the ideal class could be better satisfied if we perform piecewise correspondence between images (say frontal images) of the class. The piecewise correspondence of triangulations compensates for the shape variation and leaves only the texture variation. In our implementation, we first warp all images into a generic shape by piecewise affine transformations, shown in Fig. 3, and then apply the quotient image technique to the shape-free images. After quotient image is derived and the re-lighting images are rendered, we warp the relighting images back to the original shape. Example re-rendered results are demonstrated in Fig. 4. Since this procedure makes the shapes of the images adapted to the basic assumption of the quotient image, we called it adaptive quotient image (AQI).

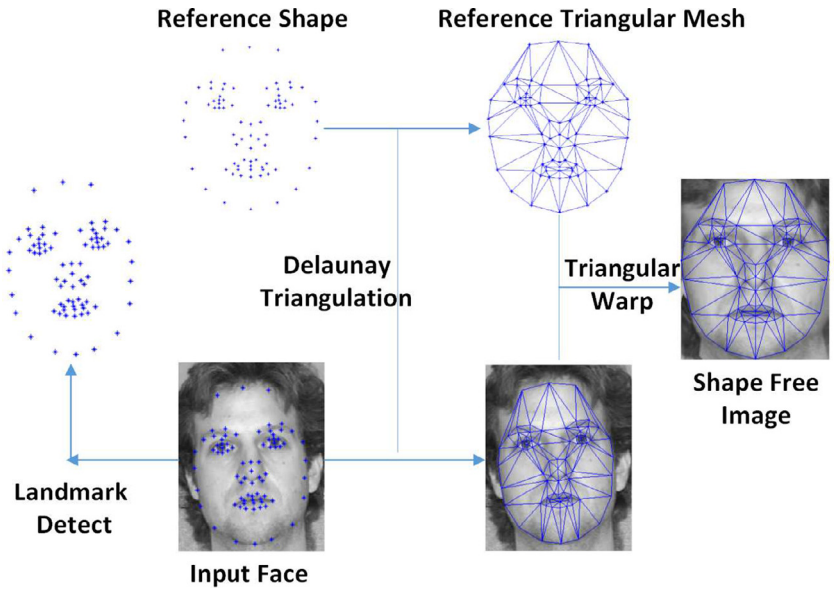

Fig. 3. The procedure to generate the "shape-free" image, which transforms the image by the piecewise affine warp according to the delaunary triangulation. All images after this procedure are assumed to have the same shape. The proposed adaptive quotient image technique maps all images to the "shape-free" image before the derivation of the quotient image.

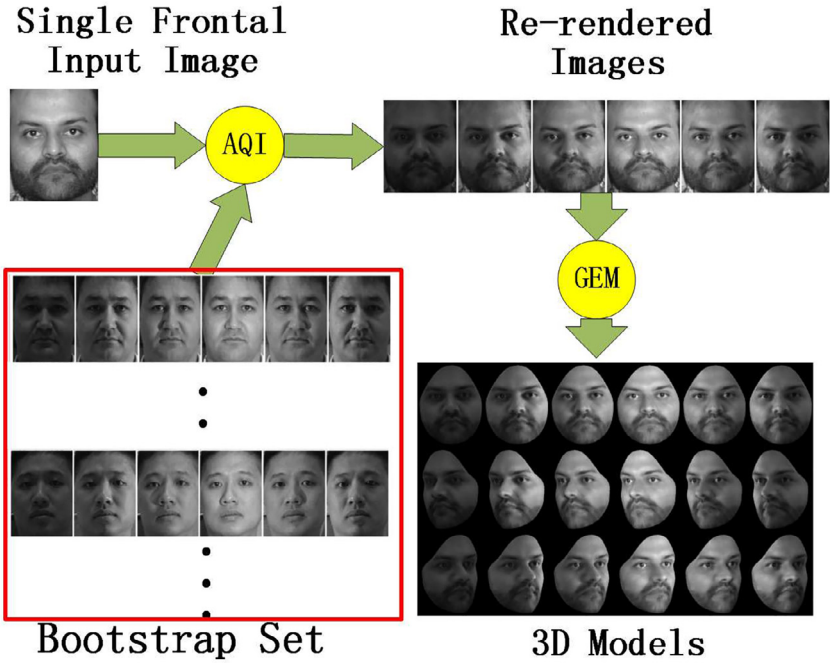

Fig. 4. Visual illustration of overall pipeline of Extended GEM based image rendering with adaptive quotient image. The 3D Models generated by GEM are rendered at yaw $0^{\circ}, -30^{\circ}, +30^{\circ}$ from top to bottom.

Based on the same assumption of the generic face shape, GEM and AQI are two complementary models that characterize the pose and illumination variations of human face respectively. Using the same delaunary triangulation based on the same landmarks, it is natural to perform the dense correspondence between the AQI and the 3D generic model. With such a dense correspondence, the texture of AQI can naturally be mapped to 3D surface of GEM to characterize the variable illumination. The pipeline of image augmentation of Extended GEM is summarized in Fig. 4. With Adaptive Quotient Image, we first virtually re-render the input image under variable lighting, and then construct corresponding 3D models from the re-lighting images by 3D GEM. These 3D models are rendered at different views to synthesize images under various pose and illumination conditions. In this manner, with a single 3D shape prior, we make pose and illumination augmentation with only one gallery sample, given a small bootstrap set of images. Synthesized images of one subject under 7 target poses and 20 illuminations are shown in Fig. 5. The detailed procedures are summarized in Algorithm 1.

4. Face recognition via Pose-Aware Metric Learning

Given a single frontal gallery image, the extended generic elastic model synthesizes facial images under varying 3D shape (depth) and illumination variations at the target pose. To reduce the shape and illumination variability, Pose-Aware Metrics are indi-

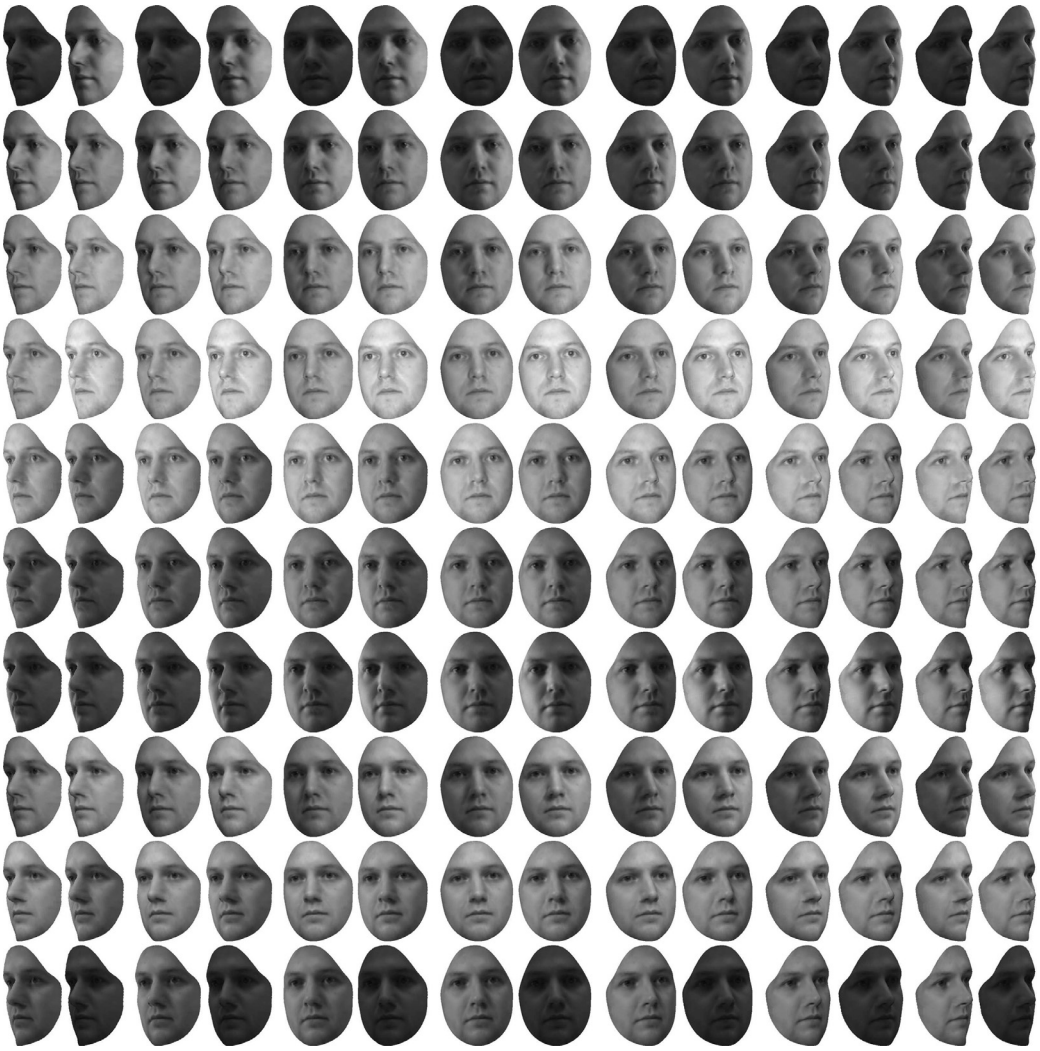

Fig. 5. E-GEM based synthesized images of a single individual to simulate the illumination and pose conditions in the Multiple-PIE database. The synthesized images have been divided into seven pose subsets with 20 light source directions. Every pair of columns shows the images from a particular pose (−45 through 45° from left to right). Note that all the images were generated from the **a single frontal image** using our Extended GEM with the depth parameter $\alpha = 1$.

vidually learnt by linear regression analysis [16] of every quantized pose. Pose-invariant recognition is performed by the metric space at the corresponding pose. This section introduces the recognition pipeline and the linear regression analysis [16] technique used in PAML.

4.1. Pose estimation and alignment

Given a testing (non-frontal) face, five facial fiducial landmarks are automatically located by using face alignment algorithms[1] such as SDM [42,43], including the centers of two eyes, the tip of the nose, two corners of the mouth. A linear regression framework, enlightened by the face 3D alignment process in [44], is employed to conduct pose estimation based on the five landmarks. Although this method provides a weak pose estimation, experimental results shows that it is sufficient to assure reasonable recognition accuracy.

Each 3D model in the database is rendered at the estimated pose and 2D images are synthesized after 2D projection. The virtually rendered images and probe image are aligned by an affine transformation, in order to compensate for scaling and in-plane rotation, using two eyes and the midpoint of two mouth corners. Specifically, all faces are aligned to 65×75 pixels with eyes position of (15, 20) and (50, 20) and the midpoint of mouth corners position of (32.5, 60). Example aligned images are show in Fig. 6(a).

4.2. Recognition via Pose-Specific Metric

The 3D continuous pose space can be divided into a limited number of quantized poses according to the requirement of the application. For example, our experiments render the synthetic faces along the yaw with a range of $\pm 50°$ [2] at steps of 5° (see Fig. 6(a)). At each quantized pose, we render the 3D models with virtually varying shape and illumination of each gallery subject as the training set. The basic assumption of PAML is that, at each pose angle, the synthesized images of a subject with varying shapes and illuminations span a low dimensional subspace. Based on this assumption, PAML learns a transformation matrix $W^{(p)}$ by which the

[1] In the experiment, we have manually labeled the feature points on the failure cases of face alignment.

[2] In the experiment, we apply the affine transformation (determined by centers of two eyes and mouth) to align the face images. The alignment is applicable to the face under relatively small pose angles. If the pose angle is large, the similarity warp designed in [45] can be used.

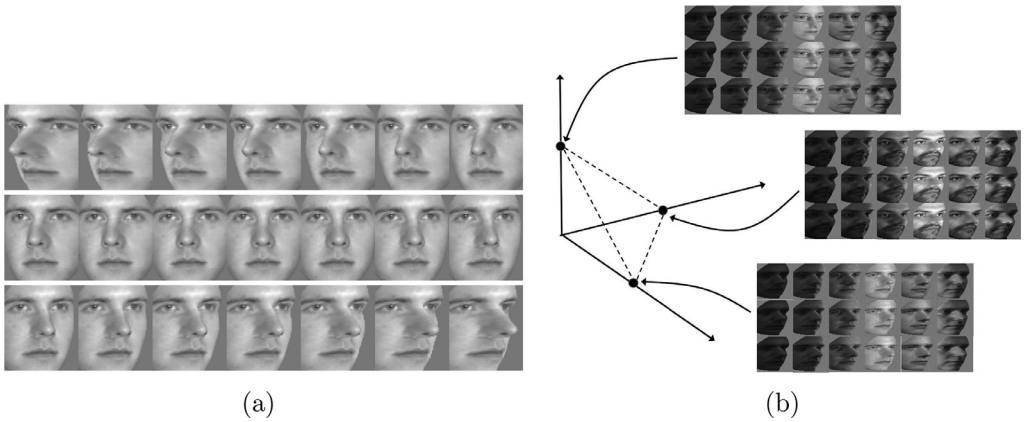

Fig. 6. (a)Aligned faces of quantized poses (yaw -50° to $+50^\circ$ in step of 5°) (b)The geometric interpretation of a Pose-Aware Metric space at the yaw 45° pose. Three classes of rendered images under the same pose, but with varying 3D depths ($\alpha = \{0.9, 1.0, 1.1\}$) and lightings, are mapped to $[1; 0; 0]^T$, $[0; 1; 0]^T$, $[0; 0; 1]^T$ respectively.

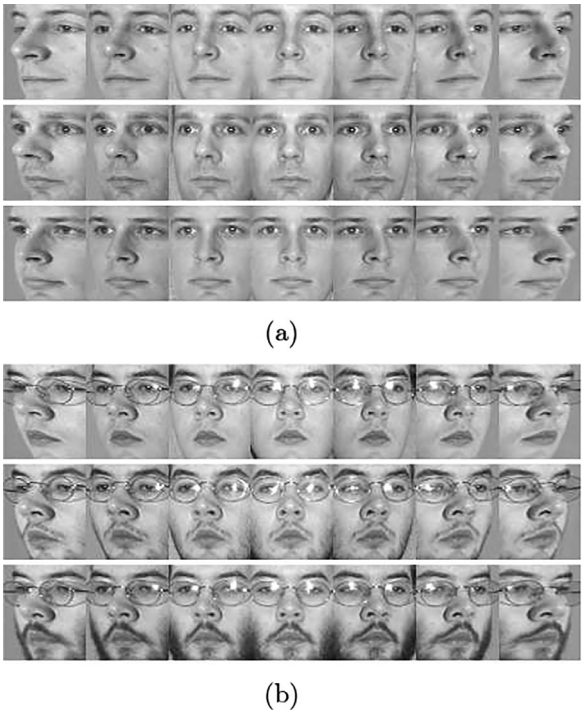

Fig. 7. Example aligned facial images used in our experiments. (a) The images of three subjects in the MultiPIE single-session testing. (b) The images of one subjects in MultiPIE multiple-sessions testing, which display significant appearance change by time interval.

illumination subspace of each subject (at each pose) is mapped to the metric space where the images of each subject locate at the corresponding class indicator vector. Fig. 6(b) illustrates a geometric interpretation of this mapping at the yaw pose of 45°. The purpose of this mapping is to eliminate the illumination uncertainty under a pose by mapping the illumination subspace to a single point. Moreover, this mapping could enhance the discrimination among similar faces by mapping them to equidistant targets (Fig. 7).

Linear regression analysis (LRA) is a parameter-free method that has achieved state-of-the-art performance on the SSFR for frontal face images [16]. In the PAML framework, we extend this technique to solve the metric learning problem at each quantized pose separately. Specifically, For the recognition of K subjects, countless equidistant embeddings are feasible by shifting and rotating a $K-1$ regular simplex. For the efficiency purpose, the class indicator vector $y_i \in \mathbb{R}^K$ is applied to represent the ith subject, where $y_i = [0, \cdots, 1, \cdots, 0]^T$ has a single 1, i.e. its ith component. This setting of equidistant embedding avoids the time-consuming nearest-neighbor search for recognition, since the nearest prototype can be efficiently found by the maximum element of the vector.

Table 3
Comparison of training images.

Methods	Training images
Li [18]	7cm100 Identities × 7 Poses × 20 Illuminations, Totally **14000** Images
RL [48]+LDA	
CPF [19]	
PAML	12 Identities × 1 Frontal Pose × 20 Illuminations, Totally **240** Images

At each quantized pose, let $X_k \in \mathbb{R}^{l \times N}$, $k = 1, \ldots, K$ denote the stacked feature vectors of the N synthesized images (Under variable depth and illumination) for the subject k. Using class indicator targets y_i as multivariate outputs of the gallery samples $X = [X_1, \cdots, X_K] \in \mathbb{R}^{l \times (NK)}$, we can formulate the linear regression model in matrix notation

$$T = WX + E \tag{6}$$

where T is a $K \times NK$ target matrix whose columns are the one-hot class indicator vectors, W is a $K \times l$ mapping matrix and $E = [e_1, e_2, \cdots, e_{NK}]$ is a $K \times NK$ matrix of errors. In order to minimize the mean square error, i.e. $\mathrm{Tr}\{E^T E\}$, the optimal transformation matrix can be readily computed as follows.

$$W = TX^\dagger \tag{7}$$

where $X\dagger$ denotes the generalized inverse of X. For each quantized pose indexed by superscript p, the optimal transformation matrix $W^{(p)} \in \mathbb{R}^{K \times l}$ is derived in the identical manner.

When a novel test image near the quantized pose p is presented to the LRA based classifier, the feature vector of the image, denoted by x, is first extracted and then normalized to zero mean and unit length. The response vector $r \in \mathbb{R}^K$ is derived by a linear transformation: $r = W^{(p)}x$. Finally, the recognition result is determined by the largest component of the response vector:

$$\omega = \arg \max_{i=1,\ldots,K} r^{(i)} \tag{8}$$

where $r^{(i)}$ denotes the ith element of the response vector r.

For automatic recognition, we estimate the pose of the probe image using head pose estimator, find the nearest quantized pose

Table 4
Average Recognition Rate (Percent) on Different Poses under **Setting-III**. The Best Performance Are in **Bold**. Pose Strategy is as **PS** for simplification.

Methods	−45°	−30°	−15°	+15°	+30°	+45°	Avg.
Li [18]	63.5	69.3	79.7	75.6	71.6	54.6	69.3
RL [39]+LDA	67.1	74.6	86.1	83.3	75.3	61.8	74.7
CPF [19]	73.0	81.7	89.4	89.5	80.4	70.3	80.7
PAML (PS ♯1)	76.5	88.3	**98.5**	**99.2**	**95.4**	84.3	90.4
PAML (PS ♯2)	76.3	89.5	97.0	98.3	94.1	85.1	90.1
PAML (PS ♯3)	**79.0**	**90.3**	97.0	98.3	94.7	**87.4**	**91.1**

p, and recognize the image using the corresponding Pose-Aware Metric (called **pose quantization strategy**). The recognition can be readily conducted by the linear regression: $r = W^{(p)}x$. Also, another strategy called **pose quantization plus search range** can be defined as using the corresponding Pose-Aware Metrics of the two nearest quantized poses to determine the recognition result. For our LRA method, the fusion is naturally conducted by simply adding up the output scores of each gallery identity. Given a weak pose estimator, this strategy is expected to compensate for the incorrect estimation and obtain more accurate recognition rate.

5. Experiments and results

In this section, we evaluate the effectiveness of the proposed PAML method on the Multi-PIE face database [20]. The Multi-PIE face database contains 754,204 images of 337 identities, where each identity has images captured under 15 poses and 20 illuminations in four sessions during different periods.

5.1. Recognition across poses by MD-GEM

We evaluate pose-invariant face recognition using two commonly used settings as follows.

- *Single-session Setting* adopts images of different poses and neutral illumination marked as ID 07. Only the images in session one are used, which only has 249 identities. The images of the first 100 identities are for model training (PAML does not use these external training data), and the images of the remaining 149 identities for test. In the test stage, one frontal image of each identity in the test set is selected in the gallery. The remaining images from $-45^\circ \sim +45^\circ$ except 0° are selected as probes.
- *Multi-sessions Setting* adopts images of different poses and neutral illumination marked as ID 07. It evaluates the robustness to pose variations. For Setting-I, the images of the first 200 identities in all the four sessions are chosen for training (PAML does not use these external training data), and the images of the remaining 137 identities for test. During test, one frontal image (i.e. 0°) of each identity in the test set is selected to the gallery, so there are 137 gallery images in total. The remaining images from $-45^\circ \sim +45^\circ$ except 0° are selected as probes.

These two settings have been widely used to evaluate pose-invariant face recognition, and our experiments compare our PAML method to a few existing methods, including VAAM [35], MDF [18] StackFlow [36], CCA [31], PLS [32], GMA [33], LGBP [28], LE+LDA [29], CRBM+LDA [30], RR [18], DAE [37], SPAE [38], FIP+LDA [39], RL+LDA [39], and MVP [46]. For PAML, the MD-GEM renders $N = 6$ images with the depth parameter $\alpha = \{1.1, 1.0, 0.9, 0.8, 0.7, 0.6\}$ for each gallery subject at each of the 6 test poses. The first experiment involves only the facial images of session 1 in which the gallery and probe images are collected at the same time under identical lighting condition. The purpose is to evaluate the accuracy of the 3D PAML model for off-pose 2D matching, assuming the pose-angle of the face image is given in the recognition stage. All images that we selected are converted to gray scale. To characterize the detailed texture of the synthesized images, the LBP feature ($LBP^{u2}_{8,1}$ histograms of the cells of 3×3 pixels [47]) is extract to represent the images for PAML.

Table 1 enumerates the comparative accuracy of 10 methods on the experiment of setting-I. Statistical subspace learning methods, such as PLS, CCA, GMA, perform the worst since they only linearly learn the association among the images of different poses. Considering that this test has not introduced any real-world factors (lighting and time changes), the accuracy lower than 80% indicates the subspace analysis of 2D images may be not a suitable technique to address pose-invariant recognition problem. By exploiting 3D information learned from the 100 subjects of the same database, VAAM and MDF improve the accuracy to about 90%. This demonstrates the usefulness of 3D model but it may not be satisfactory, especially considering the Single-session setting of this experiment. Basic deep learning method such as DAE reports a reasonable accuracy of 83%, while recent advanced deep learning methods, such as SPAE, RL+LDA, boost the accuracy to 91% and 98%, respectively. These results seem promising, but one should be aware of the same pose angles of the 200-subject training set and the test set. This concern becomes evident when a simple ridge regression method based on the same training data can achieve 97% accuracy.

The proposed MD-GEM based PAML method achieves perfect (100%) accuracy on all tested pose sets, which clearly validates the superiority of PAML over other 3D models for off-pose 2D matching. More importantly, different from other methods based on homologous training data, PAML could generalize equivalently to any target pose, since it does not rely on external training data outside the gallery. The perfect accuracy comes mainly from the 3D structure of MD-GEM, rather than the implicit correlation between the training and the test set.

The second experiment involves the images of all the four sessions of the Multi-PIE database. The purpose is to evaluate the accuracy of the 3D PAML model for off-pose 2D matching, as well as the robustness against other real-world factors, such as appearance changes caused by mustache and glasses. Table 2 enumerates the comparative accuracy of 9 methods on the experiment of setting-II. LGBP is a highly discriminative descriptor for frontal face matching, but reports the lowest accuracy of 59%. By applying feature learning techniques to the 200-subject training set, LE+LDA and CRBM+LDA methods boost the accuracy to 93% and 88% respectively. However, these training models may not generalize well to other test poses or data sets. Similar to the first experiment, traditional 3D based methods, such as MDF, achieve very high accuracy (97%) but are surpassed by recently proposed deep learning based methods, such RL+LDA and MVP.

Though previously reported accuracies on this setting are already very high, our PAML method can further reduce the recognition errors by over a half (from 1.9% to 0.7%). At the most challenging pose (45°), it outperforms the other methods by a margin of 3% accuracy. Note that all the other methods have utilized a 200-subject training set to adapt the model to test poses, while PAML trains the class model only on the single gallery image per class.

It should be noted that the comparison among these techniques may be unfair, because their involved face alignment procedures are not identical.[3] Indeed, while some competitors such as [35] are fully automatic approaches, our results are semi-automatic because we have manually adjusted the feature points in case of the failure

[3] We would like to thank the reviewer for pointing out this unfair factor. The method in [35] clearly claimed that their pipeline was fully automatic. However, other methods have not clearly stated whether manual corrections are involved or not during the face detection and alignment.

Table 5
Average recognition rate (percent) on different illuminations conditions under **Setting-III**. The best performance are in **Bold**. Pose Strategy is as **PS** for simplification.

Methods	00	01	02	03	04	05	06	08	09	10
Li [18]	51.5	49.2	55.7	62.7	79.5	88.3	97.5	97.7	91.0	79.0
RL+LDA [39]	72.8	75.8	75.8	75.7	75.7	75.7	75.7	75.7	75.7	75.7
CPF [19]	59.7	70.6	76.3	79.1	85.1	89.4	91.3	92.3	90.6	86.5
PAML (PS ♯1)	85.7	**78.0**	82.3	**87.8**	92.5	96.0	98.7	99.0	97.4	95.1
PAML (PS ♯2)	85.8	75.1	81.3	87.1	93.0	96.7	99.1	99.0	**98.0**	95.0
PAML (PS ♯3)	**87.4**	76.4	**82.7**	87.7	**94.7**	**97.1**	**99.4**	**99.6**	**98.0**	**95.5**
	11	12	13	14	15	16	17	18	19	**Avg.**
Li [18]	64.8	54.3	47.7	67.3	67.7	75.5	69.5	67.3	50.8	69.3
RL+LDA [39]	75.7	75.7	75.7	73.4	73.4	73.4	73.4	72.9	72.9	74.7
CPF [19]	81.2	77.5	72.8	82.3	84.2	86.5	85.9	82.9	59.2	80.7
PAML (PS ♯1)	89.0	83.0	75.7	**93.3**	94.7	95.6	95.3	92.7	85.2	90.4
PAML (PS ♯2)	88.6	81.3	76.5	92.0	94.7	95.2	94.7	93.0	85.1	90.1
PAML (PS ♯3)	**89.8**	**84.1**	**77.7**	92.7	**95.3**	**96.9**	**95.9**	**94.9**	**85.8**	**91.1**

of detection. In this experiment, there are only about 1% of failure cases when using our extended SDM method [43] for face alignment, and the accuracy drop from 100% to 98.7% (single-session setting) and from 99.3% to 97.6% (multi-sessions setting) if a fully automated pipeline is applied. As expected, our accuracy is much higher than another fully automatic method [35]. Moreover, one can expect the accuracy loss would gradually decrease with the development of the face alignment algorithms in the future.

5.2. Recognition across pose and illumination by E-GEM

The third experiment is conducted under the **Setting-III** that was introduced in [18,39] for the evaluation on the robustness against both poses and illuminations. This setting is more realistic than the first two settings. Specifically, Setting-III adopts images in session one for training and test, which has 249 identities. Images from -45° to $+45^\circ$ (seven poses) under 20 illuminations (marked as ID 00-19) are used. As listed in Table 3, previous studies used all the images of first 100 identities for model training (PAML does not use these external training data), and the images of the remaining 149 identities for test. In the test set, one frontal image under the natural lighting ID 7 of each identity is selected in the gallery. The remaining images from -45° to $+45^\circ$ except 0° are selected as probes. All images that we selected were converted to gray scale.

For the lighting synthesis of our E-GEM, we empirically select frontal images of 12 identities from the first 100 identities (id 001, 002, 007, 008, 011, 012, 016, 019, 025, 026, 042, 047), under illuminations marked as ID 00-19 except 07, as the bootstrap set in AQI. Such small size bootstrap set is sufficient to achieve reasonable re-lighting results. E-GEM (with the depth parameter $\alpha = 1$) renders $N = 19$ images for each gallery subject at each quantized pose. The quantized poses are sampled along the yaw with a range of $\pm 50^\circ$ at step of 5° .[4] To characterize the detailed texture of the synthesized images, the LBP feature ($LBP^{u2}_{8,1}$ histograms of the cells of 3×3 pixels [47]) is extract to represent the images for PAML.

Our experiments compare the PAML method with three well-known pose and illumination invariant methods. (1) Li et al. [18] represents a test face as a linear combination of training images, and utilizes the regularized linear regression coefficients as features for face recognition. (2) **RL+LDA** [39] first reconstructs the frontal-view face images using FIP features extracted from an image under any pose and illumination, and then applies LDA to further enhance class separation. (3) **CPF** [19] is a recent work which learns to rotate an arbitrary pose and illumination image to a target-pose face image by multi-task deep neural network.

For PAML, we have tested three pose-aware classification strategies: (1) Matching against true pose (assuming that true pose is pre-known); (2) Matching by **pose quantization** strategy; (3) Matching by **pose quantization plus search range** strategy. Table. 4 and Table. 5 report results of **Setting III**. In table 4, the recognition rates of a pose is averaged over all the possible illuminations (marked as id 00-19, 07 excluded). Similarly, in table 5, the recognition rate under one illumination condition is the averaged result of all possible poses ($-45^\circ \sim +45^\circ$, 0° excluded). The overall recognition rate of PS ♯2 is just 0.3% lower than that of PS ♯1, indicating that our pose estimator is reliable and performance is affected trivially when using the estimated pose instead of the true pose. PS ♯1 achieves best performance under -15°, $+15^\circ$, $+30^\circ$. As the pose angle becomes larger, PS ♯3 becomes the best, and boosts the performance of PS ♯1 under 45° by an average margin of 2.8%, showing that strategy of **pose quantization plus search range** works well under large angles.

The PAML method with PS ♯3 achieves the overall accuracy of 91.1% across variable pose and illumination, which is more than 10% better than the state-of-the-art multi-task deep learning methods [19]. The significant higher accuracy of PAML clearly shows the superiority of the pose-aware model for SSFR, although we use a very simple metric learning model. CPF [19] and RL+LDA [39] attempt to learn an unified deep neural network for pose-invariant feature extraction, but our results show they may not be the premier solution, even with a representative training image ensemble in this experiment. Although applied on the shallow LBP feature, PAML effectively explores the discriminative information contained among the gallery subjects, and thus shows the superior performance to the deep learning methods.

It is worth mentioning that our method has not used any non-frontal images from MPIE database to learn the cross-pose transformation. PAML just needs a few frontal images under different illumination conditions as bootstrap set for the quotient image (see Table. 3 for comparison). Although the pose-aware information is learnt from the synthesized images, PAML achieves better results using much less training samples (20 illumination $\times$ 12 identities $= 240$) than the other methods that require tens of thousands of representative samples to learn the models.

6. Conclusion

In this paper, we address the pose variation problem for single-sample face recognition by the Pose-Aware Metric Learning (PAML) approach. Our primary idea is "*from one to many*": Synthesizing

[4] Our affine-transformation based alignment method warps the face improperly at the large pose angle, and thus we only render the synthetic images along the yaw within $\pm 50^\circ$. When testing the recognition problem with larger angles, we recommend to apply different face alignment method such as that in [45].

many images of the pose and illumination variability from the single gallery image, based on which metric learning approach can reduce the "synthesized" variations at each quantified pose. Given a single frontal image, two generic elastic model extensions are proposed to synthesize facial images under varying shape and illumination conditions at any pose. Pose-Aware Metrics are individually learnt by linear regression analysis at every quantized pose for recognition. Extensive experiments on the Multi-PIE database show that the PAML achieves 100% accuracy on the test setting across poses. Moreover, PAML does not rely on any external data for model training, while existing methods use a large generic image ensemble to learn the pose invariance. On the test setting across both poses and illuminations, PAML outperforms the recent deep learning methods by over 10% accuracy.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grants 61573068, 61471048, and 61375031, Beijing Nova Program under Grant No. Z161100004916088.

References

[1] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Face recognition from a single image per person:a survey, Pattern Recognit. 39 (9) (2006) 1746–1762.
[2] W. Deng, Y. Liu, J. Hu, J. Guo, The small sample size problem of ica: a comparative study and analysis, Pattern Recognit. 45 (12) (2012) 4438–4450.
[3] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognit. Neurosci. 3 (1) (1991) 71–86.
[4] W. Deng, J. Hu, J. Guo, Gabor-eigen-whiten-cosine: a robust scheme for face recognition, in: Analysis and Modelling of Faces and Gestures, Springer, 2005, pp. 336–349.
[5] J. Yang, D. Zhang, Two-dimensional pca: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.
[6] W. Deng, J. Hu, J. Lu, J. Guo, Transform-invariant pca: a unified approach to fully automatic face alignment, representation, and recognition, IEEE Trans. Pattern Anal. Mach.Intell. 99 (2013) 1. http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.194. PrePrints.
[7] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Trans. Image Process. 11 (4) (2002) 467–476.
[8] W. Deng, J. Hu, J. Guo, W. Cai, D. Feng, Emulating biological strategies for uncontrolled face recognition, Pattern Recognit. 43 (6) (2010) 2210–2223.
[9] T. Ahonen, A. Hadid, M. Pietikinen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[10] Y. Su, S. Shan, X. Chen, W. Gao, Adaptive generic learning for face recognition from a single sample per person, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 2699–2706.
[11] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[12] W. Deng, J. Hu, J. Guo, Extended src: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
[13] W. Deng, J. Hu, J. Guo, In defense of sparsity based face recognition, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 399–406.
[14] W. Deng, J. Hu, J. Guo, Face recognition via collaborative representation: its discriminant nature and superposed representation, IEEE Trans. Pattern Anal. Mach. Intell. (2017 (Accepted)).
[15] W. Deng, J. Hu, J. Guo, W. Cai, D. Feng, Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach, Pattern Recognit. 43 (5) (2010) 1748–1762.
[16] W. Deng, J. Hu, X. Zhou, J. Guo, Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning, Pattern Recognit. 47 (12) (2014) 3738–3749.
[17] T.-K. Kim, J. Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 318–327.
[18] A. Li, S. Shan, W. Gao, Coupled bias¿cvariance tradeoff for cross-pose face recognition, IEEE Trans. Image Process. 21 (7) (2014).
[19] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, J. Kim, Rotating your face using multi-task deep neural network, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 676–684, doi:10.1109/CVPR.2015.7298667.
[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vision Comput. 28 (5) (2010) 807–813.
[21] W. Deng, J. Hu, Z. Wu, J. Guo, Lighting-aware face frontalization for unconstrained face recognition, Pattern Recognit. 68 (2017) 260–271.
[22] P. Koppen, Z.-H. Feng, J. Kittler, M. Awais, W. Christmas, X.-J. Wu, H.-F. Yin, Gaussian mixture 3d morphable face model, Pattern Recognit. (2017).
[23] J. Heo, 3D Generic Elastic Models for 2D Pose Synthesis and Face Recognition, Citeseer, 2009 Ph.D. thesis.
[24] U. Prabhu, J. Heo, M. Savvides, Unconstrained pose-invariant face recognition using 3d generic elastic models, IEEE Trans. Pattern Anal. Mach. Intell. 33 (10) (2011) 1952–1961.
[25] Z. Wu, J. Li, J. Hu, W. Deng, Pose-invariant face recognition using 3d multi-depth generic elastic models, in: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, 1, IEEE, 2015, pp. 1–6.
[26] Z. Wu, W. Deng, Adaptive quotient image with 3d generic elastic models for pose and illumination invariant face recognition, in: Proceedings of 10th Chinese Conference on Biometric Recognition (CCBR 2015), Tianjin, China, Springer International Publishing, 2015.
[27] A. Shashua, T. Riklin-Raviv, The quotient image: class-based re-rendering and recognition with varying illuminations, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 129–139.
[28] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 1, IEEE, 2005, pp. 786–791.
[29] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2707–2714.
[30] G.B. Huang, H. Lee, E. Learned-Miller, Learned-miller. learning hierarchical representations for face verification with convolutional deep belief networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2518–2525.
[31] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4) (1936) 321–377.
[32] A. Sharma, D.W. Jacobs, Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 593–600.
[33] A. Sharma, A. Kumar, H.D. III, D.W. Jacobs, Generalized multiview analysis: a discriminative latent space, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
[34] V. Blanz, T. Vetter, Face recognition based on fitting a 3d morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.
[35] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu, M. Rohith, Fully automatic pose-invariant face recognition via 3d pose normalization, in: IEEE International Conference on Computer Vision (ICCV), 2011, pp. 937–944.
[36] A.B. Ashraf, S. Lucey, T. Chen, Learning patch correspondences for improved viewpoint invariant face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[37] Y. Bengio, Learning deep architectures for ai, Found.Trends Mach. Learn. 2 (1) (2009) 1–127.
[38] M. Kan, S. Shan, H. Chang, X. Chen, Stacked progressive auto-encoders (spae) for face recognition across poses, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1883–1890.
[39] Z. Zhu, X. Luo P. andWang, X. Tang, Deep learning identity preserving face space, in: IEEE International Conference on Computer Vision (ICCV), 2013, pp. 113–120.
[40] J. Heo, M. Savvides, Gender and ethnicity specific generic elastic models from a single 2d image for novel 2d pose face synthesis and recognition, IEEE Trans.Pattern Anal.Mach.Intell. 34 (12) (2012) 2341–2350.
[41] S. Milborrow, F. Nicolls, Active Shape Models with SIFT Descriptors and MARS, Computer Vision Theory and Applications (VISAPP), in: 2014 International Conference on, 2, IEEE, 2014, pp. 380–387.
[42] X. Xiong, F. De la Torre, Supervised descent method and its applications to face alignment, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 532–539.
[43] L. Liu, J. Hu, S. Zhang, W. Deng, Extended supervised descent method for robust face alignment, in: Asian Conference on Computer Vision, Springer International Publishing, 2014, pp. 71–84.
[44] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014, pp. 1701–1708.
[45] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, arXiv: 1411.7923(2014).
[46] Z. Zhu, P. Luo, X. Wang, X. Tang, Multi-view perceptron: a deep model for learning face identity and view representations, Advances in Neural Information Processing Systems (2014) 217–225.
[47] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[48] Z. Zhu, P. Luo, X. Wang, X. Tang, Deep learning identity-preserving face space, in: Computer Vision (ICCV), 2013 IEEE International Conference on, IEEE, 2013, pp. 113–120.

Weihong Deng received the B.E. degree in information engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2004 and 2009, respectively. From Oct. 2007 to Dec. 2008, he was a postgraduate exchange student in the School of Information Technologies, University of Sydney, Australia, under the support of the China Scholarship Council. He is currently an associate professor in School of Information and Telecommunications Engineering, BUPT. His research interests include statistical pattern recognition and computer vision, with a particular emphasis in face recognition.

Jiani Hu received the B.E. degree in telecommunication engineering from China University of Geosciences in 2003, and the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008. She is currently a lecturer in School of Information and Telecommunications Engineering, BUPT. Her research interests include information retrieval, statistical pattern recognition and computer vision.

Zhongjun Wu received the B.E. degree in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. He currently is a post-graduate student major in Information and Telecommunications Engineering. His research interests include pose-invariant face recognition and deep learning.

Jun Guo received B.E. and M.E. degrees from Beijing University of Posts and Telecommunications (BUPT), China in 1982 and 1985, respectively, Ph.D. degree from the Tohuku-Gakuin University, Japan in 1993. At present he is a professor and the vice-president of BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and network management. He has published over 200 papers, some of them are on world-wide famous journals or conferences including SCIENCE, IEEE Trans. on PAMI, IEICE, ICPR, ICCV, SIGIR, etc. His book “Network management” was awarded by the government of Beijing city as a finest textbook for higher education in 2004. His team got a number of prices in national and international academic competitions including: the first place in a national test of handwritten Chinese character recognition 1995, the first place in a national test of face detection 2004, the first place in a national test of text classification 2004, the first place of paper design competition held by IEEE Industry Application Society 2005, the second place in the competition of CSIDC held by IEEE Computer Society 2006.