Gender and Ethnicity Specific Generic Elastic Models from a Single 2D Image for Novel 2D Pose Face Synthesis and Recognition

Jingu Heo, *Member, IEEE,* and Marios Savvides, *Member, IEEE*

Abstract—In this paper, we propose a novel method for generating a realistic 3D human face from a single 2D face image for the purpose of synthesizing new 2D face images at arbitrary poses using gender and ethnicity specific models. We employ the Generic Elastic Model (GEM) approach, which elastically deforms a generic 3D depth-map based on the sparse observations of an input face image in order to estimate the depth of the face image. Particularly, we show that gender and ethnicity specific GEMs can approximate the 3D shape of the input face image more accurately, achieving a better generalization of 3D face modeling and reconstruction compared to the original GEM approach. We qualitatively validate our method using publicly available databases, by showing each reconstructed 3D shape generated from a single image and new synthesized poses of the same person at arbitrary angles. For quantitative comparisons, we compare our synthesized results against 3D scanned data and also perform face recognition using synthesized images generated from a single enrollment frontal image. We obtain promising results for handling pose and expression changes based on the proposed method.

Index Terms—Generic Elastic Models, Gender and Ethnicity Specific Models, Face Synthesis, Face Recognition

---

1 INTRODUCTION

Modeling three-dimensional (3D) faces for the purpose of synthesizing novel face images from a single two-dimensional (2D) image is one of the most difficult and challenging tasks in computer vision, partly because of large variations in human faces. Researchers have been developing technologies for 3D face modeling without relying on 3D sensors, due to the demands in many real-world operating scenarios which require efficient, un-cooperative and cost-effective solutions. Research topics in this field include shape-from-shading (SFS), shape-from-stereo, structure-from-motion (SFM) and shape-from-texture. However, the quality of 3D faces obtained from these methods is often not satisfactory, and more importantly many of these approaches require multiple images. Thus reconstructing a 3D face from a single 2D face image is extremely challenging.

To achieve realistic 3D face modeling, it becomes necessary to use prior knowledge of a statistical 3D face model [3] [4] [10] [11]. However, these methods are known to be computationally expensive, and may require manually annotated control points or camera calibration. More efficient 3D face modeling approaches are desired in numerous applications, which demand real-time computation and less user cooperation. Applications include automated face recognition in surveillance video, access control, entertainment and online 3D gaming.

Recently, Generic Elastic Models (GEMs) [12] were proposed as a new efficient and reliable 3D modeling method from a single 2D image. The author of [12] claimed that the depth information (z) of a face is not extremely discriminative when factoring out the 2D spatial location of facial features. In fact, he proposed that the depth information of a face can be approximated either from another person's depth, or from generic depth information by elastically deforming a "generic" 3D depth-map. This way, fairly accurate 3D models can be generated by using only a single frontal image at a relatively less computational expense compared to previous approaches.

In this work, we improve the quality of 3D generic models by incorporating gender and ethnicity specific information of the input face. This information, known as the top two features of the widely investigated "soft-biometrics" [29], is utilized to create gender and ethnicity specific GEMs. The assumption we make is that the depth information of faces is significantly less discriminative among the same gender and ethnicity group. We show that we can achieve better 3D reconstruction results if we utilize gender and ethnicity information in the basic GEM approach.

Therefore, our Gender and Ethnicity specific GEM (GE-GEM) 3D modeling approach can be summarized as follows. Given an input face, we first extract the observed facial features, i.e., 79 automatically detected facial landmarks using a modified Active Shape Model [6] [7] on the face image, along with gender and ethnicity information. Then, depending on the gender and

---

J. Heo, the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA e-mail: jheo@cmu.edu.
M. Savvides, the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA. e-mail: Marios.Savvides@ri.cmu.edu

ethnicity of the input face, we utilize the corresponding generic depth-model (depth-map), instead of using a single generic depth-model for all faces. Finally, the designated gender and ethnicity specific 3D depth-model is deformed based on the observed 2D facial features of the input image. The correspondences between the input face and the depth-map are re-established using subdivision [9], providing more accurate visual image reconstruction and pose synthesis results. We qualitatively validate our method using 2D faces obtained from publicly available images, by showing the corresponding reconstructed 3D shape rendered from a single image and several novel 2D pose views generated using this 3D reconstructed model. In addition, we demonstrate that we can effectively handle pose and expression changes for the purpose of face recognition on the Multi-PIE database [32], although we only use synthesized images generated from a single image of each person.

This paper is organized as follows. In Section II, we briefly review related work in this area. In Section III, we analyze depth variation in human faces. Gender and Ethnicity specific GEMs (GE-GEMs) are explained in Section IV, while, in Section V, we present face synthesis and recognition results by using GE-GEMs. Finally, in Section VI, we summarize the results and effectiveness of our proposed work.

2 BACKGROUND

In this section, we first briefly review and compare well-known techniques for 3D face modeling by using a single image - namely Active Appearance Models (AAMs) and 3D Morphable Models (3DMMs). Then we summarize other techniques which can achieve 3D face modeling. Finally, we introduce the concept of the GEM approach and discuss our motivation for the use of gender and ethnicity specific information in the basic GEM approach for achieving more robust 3D reconstruction results.

As one of the leading methods in face modeling, AAMs [8] [28] [15] and 3DMMs [5] have become increasingly popular in computer graphics for modeling human faces. Although face modeling can be more efficiently achieved by AAMs compared to 3DMMs, large rotations (particularly out-of-plane) cannot be generated by the traditional 2D-only warping [8] used in AAMs due to occlusions of facial regions. To handle such large pose changes, view based models [17] [31] [18] or 3D Active Appearance Models (3DAAMs) [16] have been introduced. However, 3DAAMs still have difficulty in synthesizing images under novel illumination conditions due to the sparse 3D shape representation. 3DMMs can overcome these problems because the appearance model of a 3DMM is defined per each 3D vertex (point), allowing us to understand image formulation of faces under various lighting and pose variations.

AAMs and 3DMMs use similar shape and appearance representations. The representation space of AAMs is 2D, whereas the representation space of 3DMMs is 3D. Additionally, AAMs and 3DMMs use similar functional fitting

procedures, which can be described by minimizing the following cost function:

$$E = \sum ||\mathbf{I}_{input} - \mathbf{I}_{model}||^2 \qquad (1)$$

where $\mathbf{I}_{input}$ is the input image and $\mathbf{I}_{model}$ is the reconstructed image, obtained by the model instance. $|| \cdot ||^2$ indicates the L2-norm. In case of AAMs, $\mathbf{I}_{model}$ considers 2D pose (scale, rotation and translation), 2D shape, and 2D appearance parameters. AAMs try to find a shape and texture which minimizes the above energy cost function by iteratively changing these parameters in 2D. On the other hand, in case of 3DMMs, the $\mathbf{I}_{model}$ function includes 3D-to-2D perspective projection, rendering, 3D rotation, 3D shape, and 3D appearance parameters. Therefore, 3DMMs try to find a 3D shape with a 3D illumination-normalized texture so that it generates the input image as closely as possible after a 3D rotation and the 3D-2D projection of the 3D shape. It is well known that the fitting procedure of 3DMMs is computationally expensive due to the problem of estimating dense 3D shapes iteratively in this non-linear optimization step.

More importantly, the texture representation of 3DMMs utilizes a linear subspace generated by Principal Component Analysis (PCA). Therefore, it is hard to ensure that this subspace will model textures of all possible faces and thus the forensic reconstruction capability of this approach is inherently limited. On the contrary, we will show that our proposed approach preserves original 2D texture - any moles, marks or scars, are maintained in our reconstruction.

Instead of utilizing statistical shape information for face representation, SFS techniques emphasize depth estimation purely from images, as detailed in [1] [2] [3]. However, reconstruction from SFS itself is still a very challenging problem and may require some additional constraints, such as symmetry [24] and statistical shape information [23], for enhanced 3D modeling.

On the other hand, SFM methods [13] require multiple images for depth recovery - they may not be applicable to 3D reconstruction from a single frontal image. Due to the difficulty associated with correspondence problems and non-rigid changes in face images, many improved techniques have been developed [16] [25] [14].

From all the reviewed developments in 3D face modeling and reconstruction, it is known that statistical shape information with dense points is often required to achieve accurate 3D face models with reasonable reconstruction quality. However, due to the costs associated with capturing and processing a large corpus of 3D shapes which can represent any face with expression changes, building a generic statistical 3D face model is still an ongoing research problem. To alleviate this problem, GEMs were proposed as an alternative method for 3D face modeling. In the basic GEM approach, each face's depth information is estimated by using generic face depth information based on input 2D facial feature observations, without jointly modeling 3D shape variation of faces.

In this work, we extend the basic GEM approach in order to improve the reconstruction quality of 3D models by incorporating the gender and ethnicity specific information of the input 2D face image. We show that this information can enhance our 3D model reconstructions greatly. This has been inspired by the automatic gender and ethnicity classification work [26] which improves face recognition tasks by taking into account this information. Similarly, it has been reported by cognitive scientists that human face recognition is greatly affected by gender and ethnicity [19] [20] [27]. To automate this identification process in computer vision and machine learning research fields, scientists have been developing a gender and ethnicity classification system by using a variety of intelligent techniques in machine learning [22] [21]. However, we believe that the task of perfectly classifying gender and ethnicity is not an easy problem, not only for humans but also for machines. Therefore, we implement gender and ethnicity classifiers in a semi-automated manner. We design an interactive interface which utilizes the classification results of Support Vector Machines (SVMs) [30] [22] and an operator can confirm or modify the results of SVMs. The SVMs achieve an accuracy of 90%, and the remaining 10% of the classifier output is modified by the operator.

It is important to note that our main contribution is not to design a reliable gender and ethnicity classifier system; rather, it is our ultimate goal to show that we can improve 3D modeling if we utilize gender and ethnicity specific information. In our work, we consider two genders (Male and Female) and two ethnicity groups (Asian and Caucasian). More diverse ethnicity groups will be considered in our future work.

3 DEPTH ANALYSIS USING $\mathbf{xyz}$, $\mathbf{xy[z]}$ AND $\mathbf{xy[z}_{GE}]$ REPRESENTATIONS

In this section, we analyze depth variation in 3D human faces with three different PCA representations: PCA on $\mathbf{xyz}$, $\mathbf{xy[z]}$ and $\mathbf{xy[z}_{GE}]$ representations. The $\mathbf{xyz}$ representation uses the standard 3D ($\mathbf{xyz}$) Cartesian coordinate system, whereas $\mathbf{xy[z]}$ can be considered a $\mathbf{z}$-only representation since we sample each $\mathbf{z}$ coordinates with the same $\bar{\mathbf{x}}\bar{\mathbf{y}}$ (where $^-$ indicates the mean operator) spatial locations. In this representation we have essentially factored out the 2D spatial variation of facial features, i.e., we observe the depth variation of each face as its 2D facial features are deformed to a canonical mean ($\bar{\mathbf{x}}\bar{\mathbf{y}}$) 2D face. Finally, for the case of the $\mathbf{xy[z}_{GE}]$ representation, we consider gender and ethnicity specific information in the $\mathbf{z}$-only representation by deforming all facial features to a set of canonical mean 2D faces and measure only the depth information variation depending on gender and ethnicity.

In order to model only depth changes in $\mathbf{xy[z]}$ and $\mathbf{xy[z}_{GE}]$, it is necessary to sample all depth information at the same relative spatial locations of $(\mathbf{x}, \mathbf{y})$. We regard this sampling step as a warping process ($\mathbf{W}$) from an
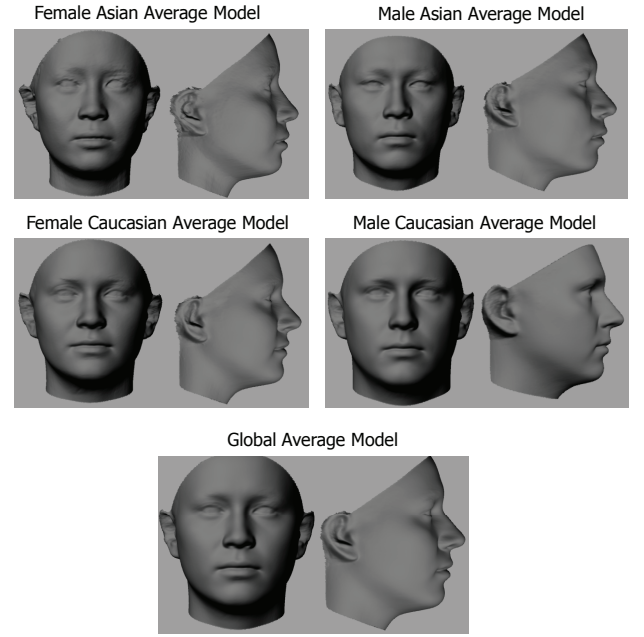

Fig. 1. Five generic models, such as Female Asian (FA), Male Asian (MA), Female Caucasian (FC), Male Caucasian (MC), and global average models, are used for our 3D reconstruction.
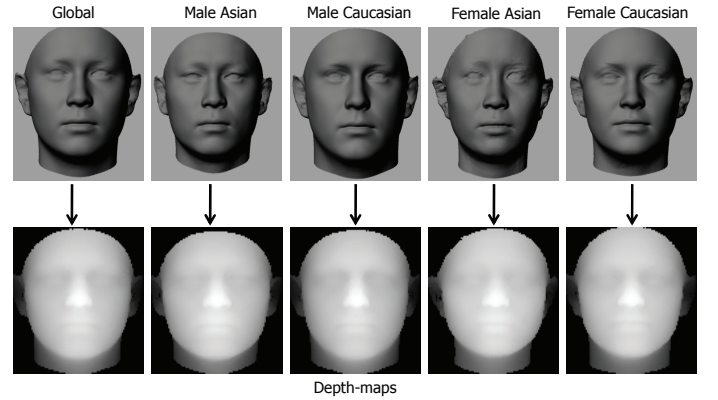

Fig. 2. Depth-maps of the 3D generic models.

input ($\mathbf{x}, \mathbf{y}$, $\mathbf{z}$) to a canonical 2D shape ($\bar{\mathbf{x}}, \bar{\mathbf{y}}$). We write this as:

$$\mathbf{Z}_{sf} = \mathbf{W}(\mathbf{x}, \mathbf{y}, \mathbf{z}; \bar{\mathbf{x}}, \bar{\mathbf{y}}) \qquad (2)$$

where $\mathbf{Z}_{sf}$ (a matrix version of $\mathbf{z}_{sf}$) is the shape-free (in terms of $\mathbf{x}$ and $\mathbf{y}$) depth-map (we convert depth values in $\mathbf{z}$ into intensity values) which allows us to model only depth changes in $\mathbf{xy[z]}$ and $\mathbf{xy[z}_{GE}]$.

Then the depth information of each image can be represented by a linear combination of the depth basis vectors ($\mathbf{V}_{z_{sf}}$) through PCA:

$$\mathbf{Z}_{sf} = \overline{\mathbf{Z}}_{sf} + \sum_{i=1}^{m} p_i \mathbf{V}_{z_{sf}}^{i}, \qquad (3)$$

where p_i is the projection coefficient. In case of $\mathbf{xy[z]}$ and $\mathbf{xy[z}_{GE}]$, we first obtain the depth information in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

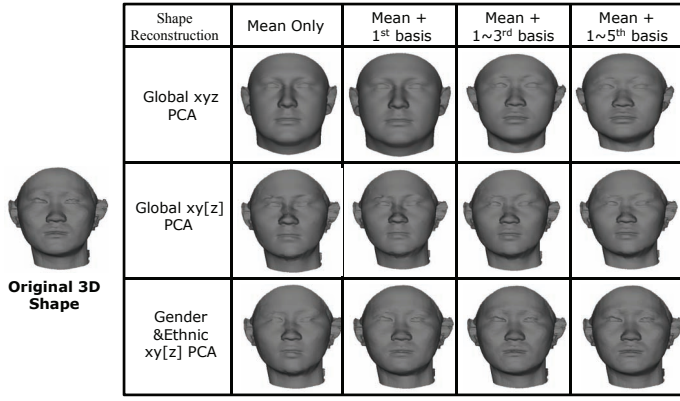IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

4

Fig. 3. 3D reconstruction results of a female Asian, obtained by three different PCAs with varying number of basis vectors.

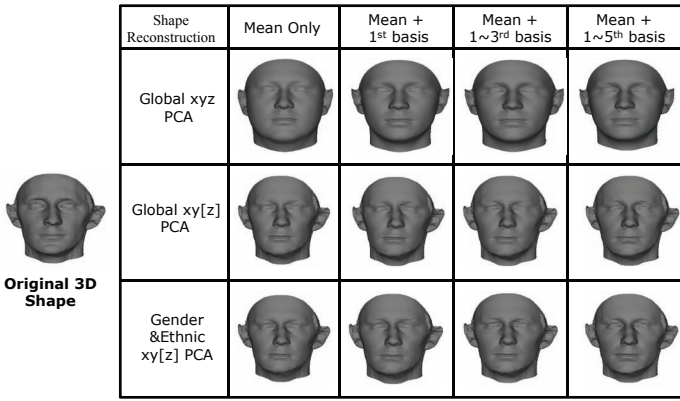Fig. 4. 3D reconstruction results of a male Caucasian.

Fig. 5. MSE comparison results of Asians with three PCA representations. (a) Female Asians and (b) male Asians.

the canonical 2D shape domain (shape-free), then apply the inverse warping $\mathbf{W}^{-1}$ for 3D reconstruction. We also exchange depth in the shape-free domain with other depth information and warp back into the $\mathbf{xyz}$ space for 3D reconstruction. Therefore, we represent 3D face shapes as $(\mathbf{x}, \mathbf{y}, \mathbf{z} = \mathbf{W}^{-1}(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \mathbf{Z}_{sf}; \mathbf{x}, \mathbf{y}))$.

Fig. 1 shows examples of five different generic average models ($\bar{\mathbf{Z}}_{sf}$), such as female Asian, male Asian, female Caucasian, male Caucasian and global average models. The corresponding depth-maps of these average models, where the intensity of the images becomes the length in depth, are shown in Fig. 2. These global $\mathbf{z}$-only depth information ($\mathbf{xy}[\mathbf{z}]$) and gender and ethnicity $\mathbf{z}$-only depth information ($\mathbf{xy}[\mathbf{z}_{GE}]$) are synthesized by using the USF database [5].

In order to measure the effectiveness of GE-GEMs both qualitatively and quantitatively, we compare the results of three different PCA reconstruction methods by using reconstructed 3D shapes and Mean Squared Error (MSE) values. Figures 3 and 4 display qualitative results. In these figures, we enumerate the number of basis vectors (up to five) used for each 3D reconstruction to examine the quality of 3D models. As can be seen in the "Mean Only" column in each figure, the depth exchanged results with generic depth information still
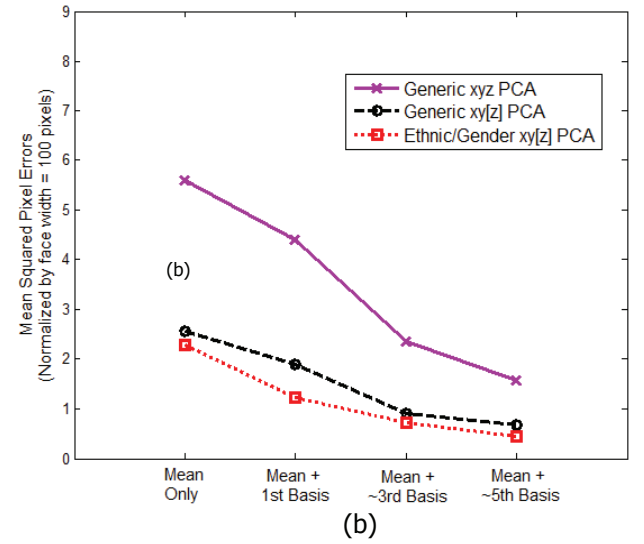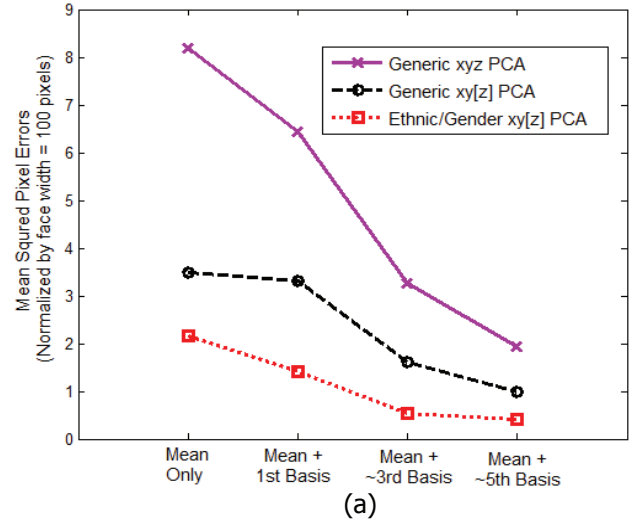
tend to preserve original 3D shapes in $\mathbf{xy}[\mathbf{z}]$ and $\mathbf{xy}[\mathbf{z}_{GE}]$ representations compared to PCA on raw 3D data. In order to achieve near perfect reconstruction in $\mathbf{xy}[\mathbf{z}_{GE}]$, we need only five leading basis vectors. Although the $\mathbf{xy}[\mathbf{z}]$ PCA representation also renders these 3D models with reasonable quality, better 3D reconstruction can be achieved if we utilize gender and ethnicity information with a relatively less number of basis vectors.

The quantitative results with the reconstructed MSE values for Asians and Caucasians using three PCA subspace representations by varying number of basis vectors used for reconstruction are shown in Figures 5 and 6, respectively. In the above experiments, we normalized all 3D shapes with a face width of 100 pixels to evaluate MSE values in a consistent manner. Although we can achieve better 3D reconstruction if we utilize more basis vectors with known depth information, it is typically acknowledged that obtaining exact depth information from a single image is still a challenging and error-prone task. From the rendered 3D shapes shown
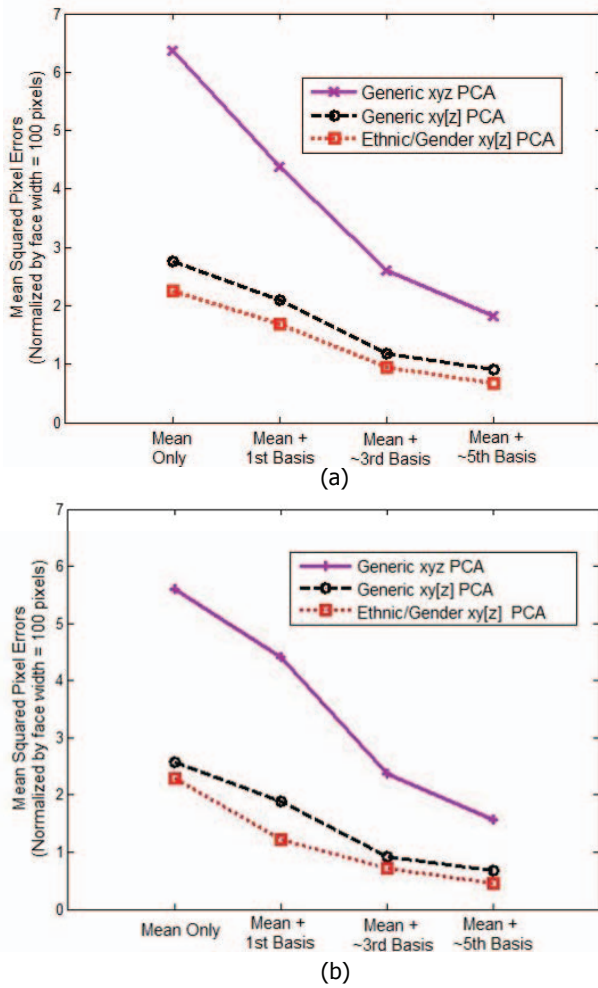
(a)

(b)

Fig. 6. MSE comparison results of Caucasians with three PCA representations. (a) Female Caucasians and (b) male Caucasians.
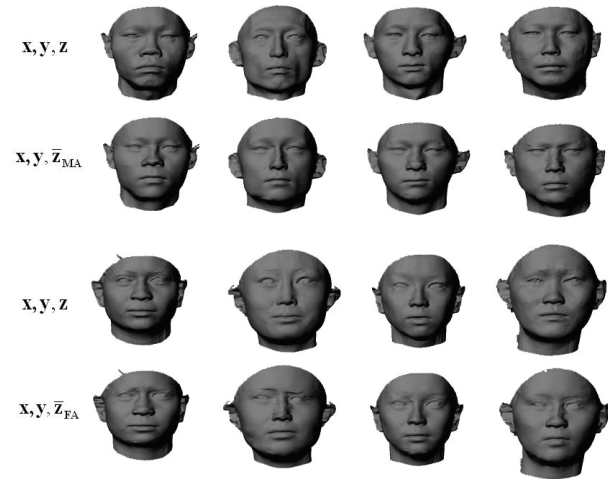

Fig. 7. Depth exchange using Asian average models.
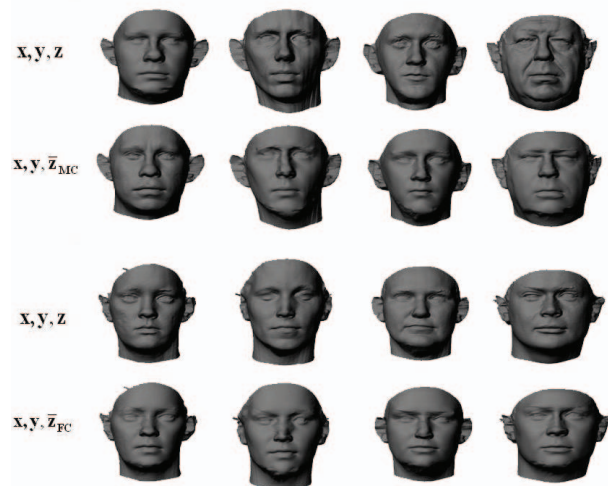
Fig. 8. Depth exchange using the Male Caucasian (MC) and Female Caucasian (FC) average models.

in these figures, we see that it is reasonable to make an assumption that the average gender and ethnicity specific depth information (i.e., the mean only basis) is sufficient to generate accurate 3D modeling without utilizing additional basis vectors.

More results of depth exchange with Asian and Caucasian generic depth-models are shown in Figures 7 and 8. In Fig. 7, the first row images are the original 3D (rendered) images and the second row images are rendered by using the $\mathbf{x}$ and $\mathbf{y}$ spatial locations of the corresponding column images of the first row and the depth information ($\bar{\mathbf{z}}_{MA}$) from the Male Asian average model. Similarly the depth information ($\bar{\mathbf{z}}_{FA}$) of the Female Asian (FA) average model is used for modeling the female Asians. Depth exchange results by using the Male Caucasian (MC) and Female Caucasian (FC) average models are shown in Fig.8.

From these results, we can see that gender and ethnicity based GEMs generate 3D models with reasonable quality comparable to the original 3D data, allowing us to further confirm that the depth information ($\mathbf{z}$) of faces

is not very discriminative among persons within a specific gender and ethnicity group. What makes persons discriminative is the spatial locations of $\mathbf{z}$, which are the $\mathbf{x}$ and $\mathbf{y}$ coordinates of facial fiducial features (such as the eyes, nose, and mouth). Due to different $\mathbf{x}$ and $\mathbf{y}$ positions, which sample the depth information $\mathbf{z}$, and the texture information, people look completely different from each other.

4 GENDER-ETHNICITY SPECIFIC GENERIC ELASTIC MODELS (GE-GEMS)

In this section, we introduce the basic GEM framework [12] for modeling a realistic 3D face from a single image. Formally, this problem can be stated as follows. "Given a face image ($\mathbf{I}$), automatically extract input face landmarks ($\mathbf{S}_{2\mathbf{x}n}$) and assign generic depth information." To achieve this task, the $\mathbf{x}$ and $\mathbf{y}$ information needs to be estimated from the input image and the depth
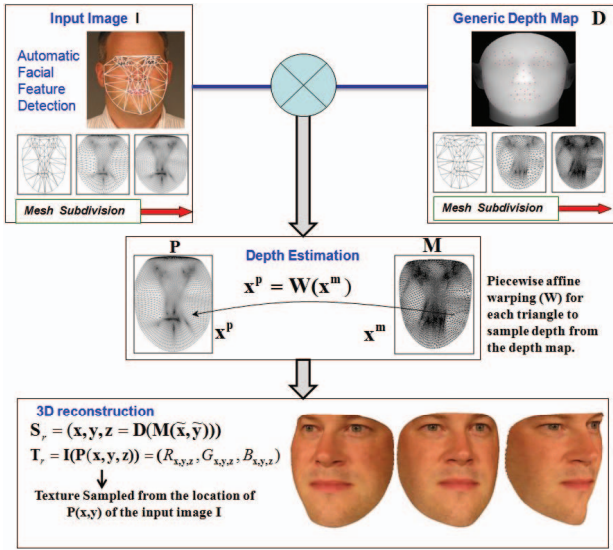
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

6

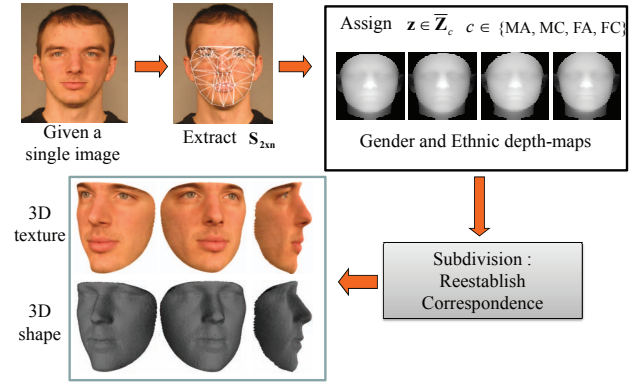Fig. 9. The basic GEM approach for 3D modeling from a single image.

Fig. 10. A visual illustration of gender and ethnicity GEMs. Four gender and ethnicity models, Female Asian (FA), Male Asian (MA), Female Caucasian (FC), and Male Caucasian (MC) models are used for 3D reconstruction in the GEM framework.
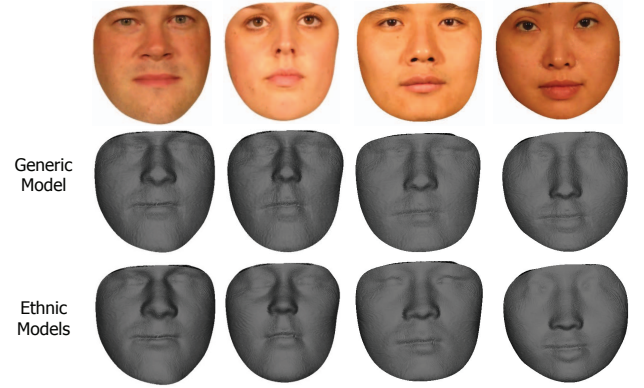

Fig. 11. 3D shape comparison results between a single global GEM and GE-GEMs. These images are obtained from the MBGC database [33].

information can be recovered by elastically deforming the GEM depth-map based on the input 2D facial feature observations.

The overall procedure of the basic GEM approach is shown in Fig. 9. Based on the detected shape, each face (**I**) is partitioned into a mesh of triangular polygons (**P**). Similarly, the generic depth-model (**D**) is partitioned into a mesh (**M**) from predefined landmark points. After registering points between the input image and the generic depth-map, we increase the point density simultaneously for both using Loop subdivision [9]. The subdivision method used here can be considered an intermediate step for establishing dense correspondence between the input mesh and the depth-model. A piecewise affine transform **W** [15] is used for warping the depth-model (**D**), sampled at the spatial locations of **M**, onto the input triangle mesh (**P**) in order to estimate the depth information. Each point in the input image has an exact corresponding point in the depth-model and the intensity of the depth-model can be used for the estimation of depth in the input image. Finally, the intensity of the input image $\mathbf{I}(\mathbf{P}(\mathbf{x}, \mathbf{y}))$, sampled at the spatial locations of $\mathbf{P}(\mathbf{x}, \mathbf{y})$, is mapped on to the 3D shape. Therefore, we represent the reconstructed 3D face by:

$$\mathbf{S}_r = (\mathbf{x}, \mathbf{y}, \mathbf{z} = \mathbf{D}(\mathbf{M}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})))$$
$$\mathbf{T}_r = \mathbf{I}(\mathbf{P}(\mathbf{x}, \mathbf{y}, \mathbf{z})) = (R_{\mathbf{x},\mathbf{y},\mathbf{z}}, G_{\mathbf{x},\mathbf{y},\mathbf{z}}, B_{\mathbf{x},\mathbf{y},\mathbf{z}}) \quad (4)$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ in **M** are the registered points **x** and **y** in image **P**. This representation with shape ($\mathbf{S}_r$) and texture ($\mathbf{T}_r$) can be achieved using 3DMMs. By avoiding the computationally-expensive 3D reconstruction process in 3DMMs, the GEM approach can achieve 3D modeling in an extremely short time (1-2 seconds). Furthermore, unlike 3DMMs that use a PCA subspace to reconstruct

2D texture, we utilize the actual texture of the input 2D face. Therefore, we can preserve the input 2D texture and maintain any forensic features such as scars, marks or moles while a generic subspace method (PCA) may have difficulty in reconstructing such features.

In order to improve the quality of 3D models in the GEM framework, we utilize gender and ethnicity information. We assign a set of generic depth-models ($\mathbf{z} \in \mathbf{Z}_c$), where $c \in (MA, MC, FA, FC)$, according to the input gender and ethnicity group of the face. More accurate 3D modeling can be obtained if we consider the corresponding gender and ethnicity specific GEM framework since these designated depth-models have shown to provide more accurate depth information for reconstruction. A visual illustration of our GE-GEMs is illustrated in Fig. 10.

To show the effectiveness of the GE-GEMs, we first compare the 3D modeling results obtained from a global generic GEM and Gender-Ethnicity specific models (GE-GEMs). The results are shown in Fig. 11. As shown by these images, more visually accurate 3D shapes can be obtained by the use of GE-GEMs. Although a variety

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

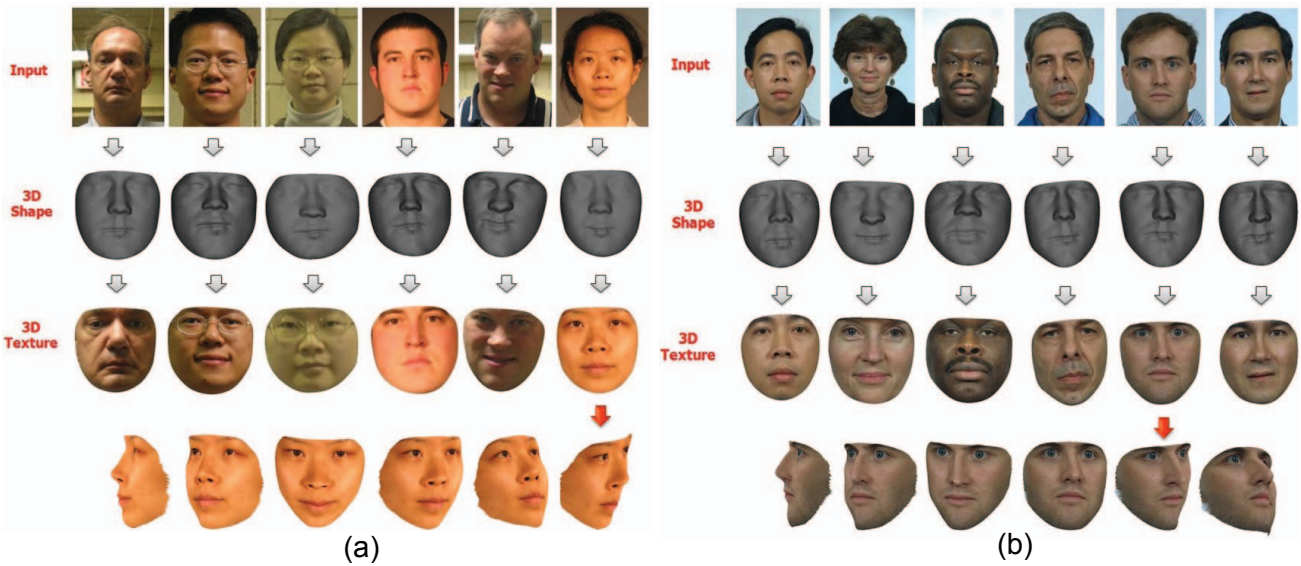IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

7

Fig. 12. Examples of 3D shape modeling from a single image by using GE-GEMs obtained from (a) the MBGC database [33], (b) the FERET database [34], respectively.

of gender and ethnicity information might assist in improving the quality in our 3D reconstruction process, we have shown that the reconstructed 3D shapes, even with only four different gender and ethnicity models in this paper, are enough to be used for 2D face synthesis. Many examples of 3D modeling by using GE-GEMs are shown in Fig. 12. In each figure, the second row images contain the estimated 3D shapes and the bottom row images illustrate the results of novel pose synthesis. These 3D models are generated from their corresponding input images in the first row. Due to the difficulty of obtaining the ground truth for the 3D shapes of these images, we demonstrate the estimated 3D shapes in a qualitative manner.

5 FACE SYNTHESIS AND RECOGNITION USING GE-GEMS

Our key assumption in this paper, that the depth information (z) of a face is not significantly discriminative among the same gender and ethnicity group, serves to further emphasize the insignificance of exact depth estimation of each face for 2D face synthesis and face recognition. We have shown that the depth information of a face can be easily synthesized by using a set of generic depth information. This way, we can represent 3D face shapes very efficiently without the need to acquire accurate depth information (z) of each individual by using 3D scanners. More importantly, we also analyzed and showed that facial depth information is not significantly discriminative in 3D faces; this assumption is the premise that makes our proposed GE-GEM approach feasible. Indeed, generic depth information is shown to be sufficient to model depth information for synthesizing novel 2D faces images under different views.

In this section, we first present our comparison results of synthesized 2D images obtained by using GEMs and GE-GEMs against those of the ground-truth scanned 3D data, followed by comparison results of face recognition using GEMs and GE-GEMs.

5.1 GEM vs. GE-GEM for Face Synthesis

This section contains two different comparison results of GEMs and GE-GEMs against actual 3D scanned data for the purpose of generating novel 2D views. Based on the same frontal image used in the 3D scanned data, we modeled a 3D face and obtained 2D images by using GEM and GE-GEM approaches, respectively. Then, each synthesized image is aligned to the view of the ground-truth 2D image (2D projection of the 3D scanned data). Finally, we compute distances between them by using the normalized sum squared distance measure. These comparison results are shown in Fig. 13 and Fig. 14, respectively. As demonstrated by these figures, GE-GEM seems to generate novel views more robustly with relatively less errors compared to GEMs.

5.2 Face Recognition using Synthesized Images

In order to quantitatively show the effectiveness of our 3D modeling method for face recognition, we compare 2D and 3D approaches for handling pose and expression variations. We utilize a subset of the MPIE database [32] with three different illumination sets ('07','08','09'), each containing images of 100 persons taken under 5 views, with/without expression changes. Examples of the MPIE database are shown in Fig. 15. It is important to address that there is some degree of illumination variation for the same illumination set images, although these images are taken at the same session. We first present 2D methods for handling pose changes followed by comparison

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
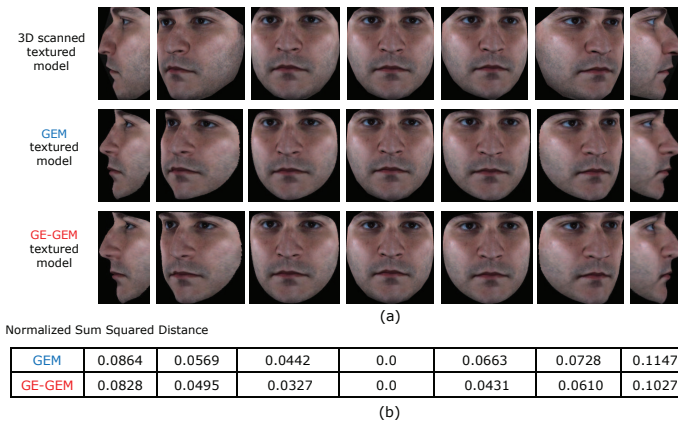IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

8

Fig. 13. Comparison results of GEMs and GE-GEMs against 3D scanned data. In (a), the first row images are synthesized 2D images under different poses using a 3D scanned data and the second row images are the synthesized images using GEMs while the third row images are generated by using GE-GEMs. The bottom table indicates the distances between the ground-truth 2D images and 2D images obtained by using GEMs and GE-GEMs, respectively.
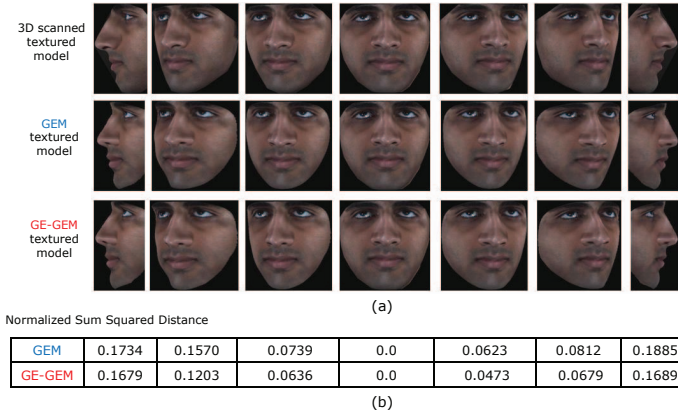

Fig. 14. Another comparison of GEMs and GE-GEMs against 3D scanned data.

results of 2D and 3D methods. It is worth mentioning that there is no training stage in these experiments. We perform face recognition using the rank-1 recognition rate with the smallest cosine distance (1 - cosine distance) measure for the performance evaluation.

TABLE 1
Rank-one face recognition performance across views using the eye-based normalization.

| | Angle (degree) | Query | | | | |
		-30	-15	0	15	30
Gallery	-30	100%	51%	26%	20%	10%
	-15	46%	100%	59%	43%	18%
	0	32%	70%	100%	67%	10%
	15	23%	55%	72%	100%	62%
	30	20%	21%	26%	50%	100%

Fig. 15. Examples of the images used for our experiments (angles in degree).

TABLE 2
Rank-one face recognition performance across views using the 2D warping method.

| | Angle (degree) | Query | | | | |
		-30	-15	0	15	30
Gallery	-30	100%	97%	50%	17%	10%
	-15	96%	100%	90%	43%	18%
	0	61%	94%	100%	90%	66%
	15	45%	82%	92%	100%	98%
	30	10%	39%	68%	99%	100%

5.2.1 2D Methods

In the 2D methods, we use two approaches for computing a distance between two faces: eye-based normalization and global 2D warping methods. In case of the global 2D warping, we first align two face images and then we exchange the two shapes and obtain new intensity (texture) images based on these exchanged shapes. By doing this, we can align the two faces with increased pixel-level correspondence and can achieve pose correction as well. We define this warping process as "shape-preserving warping", instead of using a mean shape for warping, known as shape-free warping, which is widely used in AAMs [8]. However, in our experiments, 2D warping can effectively handle 3D face pose changes if the pose difference is less than ± 15 degrees approximately. However, if the pose difference is more than ± 15 degrees (out of plane rotations), the resulting warped images typically produce strong shearing artifacts. These distortions are caused by the texture sampled in relatively small regions. Comparison results of face recognition using these eye-based normalization and 2D warping methods are presented in Table 1 and Table 2, respectively. As shown in Table 1, the eye-based normalization scheme cannot handle pose changes effectively. On the other hand, the 2D warping method handles pose changes effectively within around ± 15 degrees. In these tables, we have shown the results of the same pose recognition performance in order to emphasize pose problems in face recognition.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

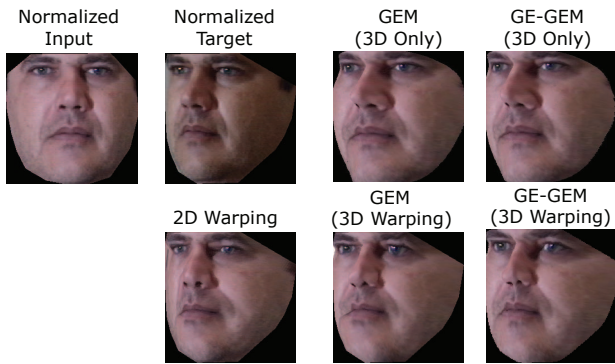IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

9

Fig. 16. Examples of pose synthesized images obtained from 3D Only and 3D Warping schemes by using GEMs and GE-GEMs. The input and target images are taken at the same session with some degree of illumination changes.

5.2.2 3D Methods

In case of the 3D pose synthesis case, two different 3D methods are considered in our face recognition experiments - namely, "3D Only" and "3D Warping" methods for handling a wide range of pose changes. The "3D Only" method utilizes a 3D rotation by using GEM or GE-GEM models and eye normalization, whereas the "3D warping" method uses a 3D rotation followed by the 2D warping method. In order to handle large pose changes with a single global model, these 3D approaches may provide a practical way of achieving face recognition systems without utilizing exact 3D depth information for each person. For clarity, examples of 2D and 3D normalization schemes are demonstrated in Fig.16.

We provide two experimental results depending on the test images with/without expression changes. Here, we only utilize frontal-neutral images in our gallery and obtain matching results after aligning our 2D synthesized images to the pose of the test images. 3D Only and 3D Warping methods perform very well across pose changes compared to the 2D based methods for both experiments. In case of neutral expression, the 3D Only method, which utilizes a 3D rotation and the eye-based normalization method, achieves better performance over the 3D Warping method. Additional 2D warping in the 3D Warping method seems to introduce some artifacts in the synthesized images. The proposed GE-GEM based 3D Only method seems slightly outperform the conventional GEM based 3D Only approach, especially in the 3D Only scheme. These comparison results are shown in Table 3.

However, for handling pose and expression changes together, the 3D Warping method produces robust face recognition performance results, which are shown in Table 4. In this case, we achieved similar results for both GEM and GE-GEM models, since additional 2D warping may compensate for some shape differences (expression) that occurred in both models.

TABLE 3
Comparison of rank-one face recognition performance across views using frontal neutral images in gallery (single image per person) and pose changes in query.

		Query			
	Methods \ **Pose Angle**	**-30**	**-15**	**15**	**30**
	2D eye-based	32%	70%	67%	10%
	2D Warping	61%	93%	90%	66%
Gallery	3D Only (GEM)	83%	93%	90%	85%
	3D Warping (GEM)	80%	95%	97%	80%
	3D Only (GE-GEM)	87%	96%	93%	90%
	3D Warping (GE-GEM)	80%	96%	97%	81%

TABLE 4
Comparison of rank-one face recognition performance across views using frontal neutral images in gallery (single image per person) and pose and expression changes in query.

		Query			
	Methods \ **Pose Angle**	**-30**	**-15**	**15**	**30**
	2D eye-based	35%	63%	58%	9%
	2D Warping	58%	84%	85%	53%
Gallery	3D Only (GEM)	47%	65%	63%	44%
	3D Warping (GEM)	74%	84%	86%	72%
	3D Only (GE-GEM)	50%	65%	64%	49%
	3D Warping (GE-GEM)	75%	85%	85%	73%

As shown in the table, we can achieve pose and expression tolerant face recognition from a single frontal image in our galley by utilizing our proposed GE-GEMs. Based on the pose and expression of the test image, we have the ability to synthesize the corresponding pose and expression images for face recognition.

So far, we have conducted our face recognition experiments based on the rank-one recognition rate, showing GE-GEMs slightly outperform GEMs. However, in the Receiver Operating Characteristic (ROC) Curve analysis, GE-GEMs perform much better than GEMs with an Equal Error Rate (EER) difference of 7 %, as shown in Fig. 17. The results are obtained by using the average performance at ±30 degrees with three different illumination sets. If the pose angle is smaller than ±30 degree, there is not much difference in performance between them in both rank-one recognition rates and ROC curve results.

5.3 Limitations

Currently, our modeling approach has difficulty in handling non-frontal images in which some portions of the face are occluded. The bilateral symmetry property of a face can resolve this problem; however, we do not utilize this property since many faces may not be symmetric. For this reason, we present our results by using a common scenario for face recognition. In other words, we construct a 3D face from each subject in our database by using only a single frontal image for the synthesis of novel views according to the pose of an input test image.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE
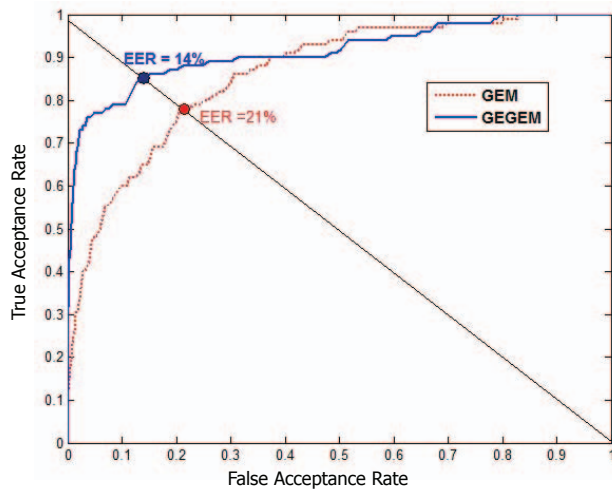
10

Fig. 17. ROC curves of average performance of GEMs and GE-GEMs at ± 30 degrees under three different illumination set in the MPIE database.

To handle pose changes in test images for face recognition, we limit our focus to within $\pm$ 30 degrees from a frontal view since our eye-detection and alignment methods need improvement for extremely off-angle images, although our 3D modeling approach can synthesize 2D images at any pose angle.

In addition, our proposed approach needs improvement to handle diverse illumination changes for face recognition. Our normalized cosine distance matching works only for images with modest global illumination changes. Our focus in the future will be to implement a complete face recognition system, handling more diverse poses, various illumination setting, and expression changes.

6 SUMMARY

In this paper, we have extended the original GEM framework and proposed GE-GEMs by utilizing gender and ethnicity information. We have shown that information from these additional "soft-biometrics" can assist in improving the overall quality of 3D models. We demonstrated that the depth information of human faces alone is not significantly discriminative when factoring out x and y positions, and it can be effectively estimated by using a few deformable Gender and Ethnicity specific models. By redefining correspondence between the input image and the designated gender and ethnicity depth-map with increased point density, we can obtain a realistic dense 3D face model from a single 2D image. We have shown our 3D reconstruction results using many examples of publicly available databases. Comparison results against 3D scanned data were provided to show the accuracy of the proposed work. We have also shown that we can achieve pose and expression invariant face recognition by using synthesized images generated from our proposed work, as tested on the MPIE dataset.

In summary, our proposed method only requires a single 2D photo to generate a 3D model and then synthesizes a novel pose at any angle for the purpose of unconstrained face matching. Furthermore, compared to 3DMMs, our proposed method can model a 3D face in very quick time (1-2 seconds) and forensically preserve 2D texture, i.e., scars, marks, moles and any other discriminating features, whereas other subspace based reconstruction methods cannot guarantee the reconstruction of such details.

REFERENCES

[1] J. D. Durou, M. Falcone and M. Sagona, "A survey of numerical methods for shape from shading," *Technical Report 2004-2-R*, IRIT, 2004.
[2] R. Zhang, P. S. Tsai, J. Edwin Cryer and M. Shah, "Shape from Shading: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690-706, 1999.
[3] J.J. Atick, P.A. Griffin and A.N. Redlich, "Statistical Approach to Shape from Shading: Reconstruction of 3D Face Surfaces from Single 2D Images," *Computation in Neurological Systems*, vol. 7, no. 1, 1996.
[4] V. Blanz, S. Romdhani and T. Vetter, "Face Identification across Different Poses and Illuminations with a 3D Morphable Model," *Proceedings of the Fifth international conference on Automatic Face and Gesture Recognition*, pp. 202-207, 2002.
[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, 2003.
[6] T.F. Cootes, C.J. Taylor, D. Cooper and J. Graham, "Active Shape Models: Their Training and Application," *Computer Vision and Image Understanding (CVIU)*, vol. 61, no. 1, pp. 38-59, 1995.
[7] K. Seshadri and M. Savvides, "Robust modified active shape model for automatic facial landmark annotation of frontal faces," *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 319-326, 2009.
[8] T. Cootes, G. Edwards and C. Taylor, "Active appearance models," *Proceedings of the European conference on Computer Vision (ECCV)*, vol. 2, pp. 484-498, 1998.
[9] C. T. Loop, "Smooth Subdivision Surfaces Based on Triangles," *M.S. Thesis*, Department of Mathematics, University of Utah, August 1987.
[10] S. F Wang and S. H. Lai, "Efficient 3D Face Reconstruction from a Single 2D Image by Combining Statistical and Geometrical Information," *Asian Conference on Computer Vision*, pp. 427-436, 2006.
[11] D. Fidaleo and G. Medioni, "Model-Assisted 3D Face Reconstruction from Video," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pp. 124-138, 2007.
[12] J. Heo, "Generic Elastic Models for 2D Pose Synthesis and Face Recognition," *Ph.D Thesis*, Department of Electrical and Computer Engineering, Carnegie Mellon University, Dec 2009.
[13] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision (IJCV)*, vol. 9, no. 2, pp. 137-154, 1992.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

11

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004.

[15] I. Matthews and S. Baker, "Active Appearance Models Revisited," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 135-164, 2004.

[16] J. Xiao, S. Baker, I. Matthews, R. Gross and T. Kanade, "Real-time combined 2d+3d active appearance model," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 535-542, 2004.

[17] T. Cootes, K. Walker and C. Taylor, "View-based active appearance models," *IEEE International Conference on Automatic Face Modeling and Gesture Recognition*, 2000.

[18] S. Romdhani, S. Gong and A. Psarrou, "A multi-view non-linear active shape model using kernel pca," *10th British Machine Vision Conference*, vol. 2, pp. 483-492, 1999.

[19] A. OToole, K. Deffenbacher, D. Valentin and H. Abdi, "Structural aspects of face recognition and the other race effect," *Memory & Cognition*, vol. 22, pp. 208-224, 1994.

[20] J. Brigham and P. Barkowitz, "Do they all look alike? the effect of race, sex, experience and attitudes on the ability to recognize faces," *Journal of Appl. Soc. Psychol*, vol. 8, pp. 306-318, 1978.

[21] S. Gutta, J. Huang, P. Phillips and H. Wechsler, "Mixture of experts for classification of gender, ethnic origin, and pose of human faces," *IEEE Trans. Neural Networks*, vol. 11, pp. 948-960, 2000.

[22] B. Moghaddam and M. Yang, "Learning gender with support faces, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 707-711, 2002.

[23] R. Dovgard and R. Basri, "Statistical Symmetric Shape from Shading for 3D Structure Recovery of Faces," *European Conference on Computer Vision (ECCV)*, vol. 3022, 2004.

[24] W. Zhao and R.Chellappa, "Illumination-insensitive face recognition using symmetric shape-from-shading," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 286-293, 2000.

[25] J. Xiao, J. Chai and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," *International Journal of Computer Vision (IJCV)*, vol. 67, pp. 233-246, 2006.

[26] X. Lu and A.K. Jain, "Ethnicity Identification from Face Images," *In Proceedings of SPIE*, vol. 5404, 2004.

[27] R. Malpass and J. Kravitz, "Recognition for faces of own and other race," Journal of Personality and Social Psychology, vol. 13, no. 4, pp. 330-334, 1969.

[28] X. Hou, S. Z. Li, H. Zhang and Q. Cheng, "Direct appearance models," *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 828-833, 2001.

[29] A. K. Jain, Hou, S. C. Dass, and K. Nandakuma, "Soft Biometric Traits for Personal Recognition Systems," *In Proc. IEEE Conf. on Biometric Authentication*, pp. 731-738, 2004.

[30] V. Vapnik, *Statistical Learning Theory*, JohnWiley and Sons, New York, 1998.

[31] A. Pentland, B. Moghaddam and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 84-91, 1994.

[32] R. Gross, I. Matthews, J. Cohn, T. Kanade and S. Baker, "Multi-PIE," *The International Conference on Automatic Face and Gesture Recognition (AFGR)*, 2008.

[33] "Multiple Biometric Grand Challenge (MBGC)", *http://face.nist.gov/mbgc/*, 2009.

[34] "The FERET Database", *http://www.itl.nist.gov/iad/humanid/feret/*, 2004.
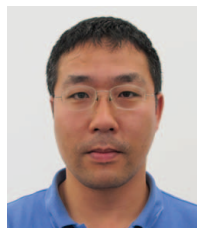
Jingu Heo received the BS degree in electronic and electrical engineering from Kyungpook National University, Daegu, Korea, in 1997, the MS degree in electrical and computer engineering from the University of Tennessee, Knoxville, in 2004, and the PhD degree in electrical and computer engineering from Carnegie Mellon University, in 2009. During 1997 and 2001, he worked for Samsung Electronics as a software engineer. He is currently working as a post doctoral researcher at Cylab, Carnegie Mellon University. He received the most cited paper award for Computer Vision and Image Understanding in 2007 and 2008, two years in a row and is also listed in 2009 Edition of Marquis' Who's Who in America. His research includes image processing, pattern recognition, computer vision and computer graphics, especially focusing on rapid 3D face modeling for face recognition.

Marios Savvides Marios Savvides is a Research Associate Professor at Carnegie Mellon University in the Electrical & Computer Engineering Department and is the Director of CMU CyLab Biometrics Center. He is one of the four chosen researchers to form the Office of the Director of National Intelligence (ODNI) Center of Academic Excellence in Identity Sciences (CASIS). His research is mainly focused on developing algorithms for robust face and iris recognition as well as in using pattern recognition, machine vision and computer image understanding for enhancing the performance of real time biometric systems. He is a program committee member on several biometric conferences such as IEEE BTAS, SPIE Biometric Identification, IEEE AutoID. He has organized and co-chaired Robust Biometrics Understanding the Science & Technology (ROBUST 2008) conference. He has authored and co-authored over 120 journals, conference publications, including several book chapters in the area of Biometrics and is an editor of the Springer's Encyclopedia of Biometrics.