Neurocomputing 410 (2020) 12–27

Contents lists available at ScienceDirect

Neurocomputing

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

3D face reconstruction from mugshots: Application to arbitrary view face recognition

Jie Liang [a,1], Huan Tu [a,1], Feng Liu [a], Qijun Zhao [a,c,*], Anil K. Jain [b]

[a] College of Computer Science, Sichuan University, China
[b] Department of Computer Science and Engineering, Michigan State University, United States
[c] School of Information Science and Technology, Tibet University, China

ARTICLE INFO

Article history:
Received 30 November 2019
Revised 6 May 2020
Accepted 24 May 2020
Available online 1 June 2020

MSC:
00-01
99-00

Keywords:
Mugshot
3D face reconstruction
Arbitrary view face recognition

ABSTRACT

Mugshots while routinely acquired by law enforcement agencies are under utilized by automated face recognition systems. In this paper, we propose a regression based approach to reconstruct textured full 3D face models from multi-view mugshot images. Using landmarks from the input frontal and profile mugshots of a subject, our method reconstructs his\her 3D face shape via either linear or nonlinear regressors. The texture of the mugshot images is mapped to the reconstructed 3D face shape via an efficient seamless texture recovery scheme. Compared with existing 3D face reconstruction methods, the proposed method more effectively utilizes the three-view mugshot face images collected during booking. The reconstructed 3D faces are used to generate realistic multi-view face images to enlarge the gallery and facilitate arbitrary view face recognition. Evaluation experiments have been done on BFM and Bosphorus databases in terms of reconstruction accuracy, and on Multi-PIE and Color FERET databases in terms of recognition accuracy. The results show that the proposed method can reduce the 3D face reconstruction error of the best competitive method from 2.31 mm to 1.88 mm, and improve the recognition accuracy of state-of-the-art deep learning based face matchers by as much as ~4% on Multi-PIE and ~2% on Color FERET despite the high baseline set by them.

1. Introduction

Mugshot face images are widely used for identity recognition in forensic applications. They usually consist of 2D frontal and profile face images of each person (see Fig. 1), which are routinely collected by law enforcement agencies. The frontal and profile face images provide complementary information of a face, and are thus believed to be useful for pose-robust face recognition if they are effectively utilized [1,2]. However, existing automated face recognition methods mostly assume that only 2D frontal face images are enrolled in gallery, and recognize off-angle probe images by extracting pose-robust features [3] or normalizing the probe faces to frontal pose [4]. Typically, these methods use three-dimensional (3D) face models to assist pose normalization or pose-adaptive feature extraction [5–7,4,8].

The usefulness of 3D face models in recognizing arbitrary view face images has also been shown by other researchers [9–12]. They assume the availability of 3D face data in gallery, and use these 3D face data to assist face recognition. The 3D face data could be acquired by using 3D scanners [13,14], or reconstructed from 2D face images [15–19]. In contrast to the high cost of 3D scanners, 2D face image acquisition devices (such as surveillance cameras and web cameras) are much more cost-effective and widespread in both forensic and civilian applications. Therefore, it is highly desirable to develop efficient methods to reconstruct 3D face models from 2D face images.

Despite the large amount of research on 3D face reconstruction, very few studies have been reported about reconstructing 3D face models from mugshot face images. Ip and Yin [21] and Ansari and Abdel-Mottaleb [22] proposed methods to reconstruct 3D face models from orthogonal-view face images. However, they required that the input face images should be calibrated, and were thus not suitable for mugshot face images that are not calibrated. Also, some methods can reconstruct 3D face models from multi-view face images. Choi et al. [23] utilized sparse bundle adjustment to reconstruct 3D landmarks from multi-view images, which were further used to deform a generic 3D face model to the final shape. To better investigate the multi-view constraint on face images, Lin et al. [24] used five images (including profile views), and inferred first the accurate poses of cameras in all views, and then a dense

* Corresponding author.
E-mail address: qjzhao@scu.edu.cn (Q. Zhao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neucom.2020.05.076
0925-2312/

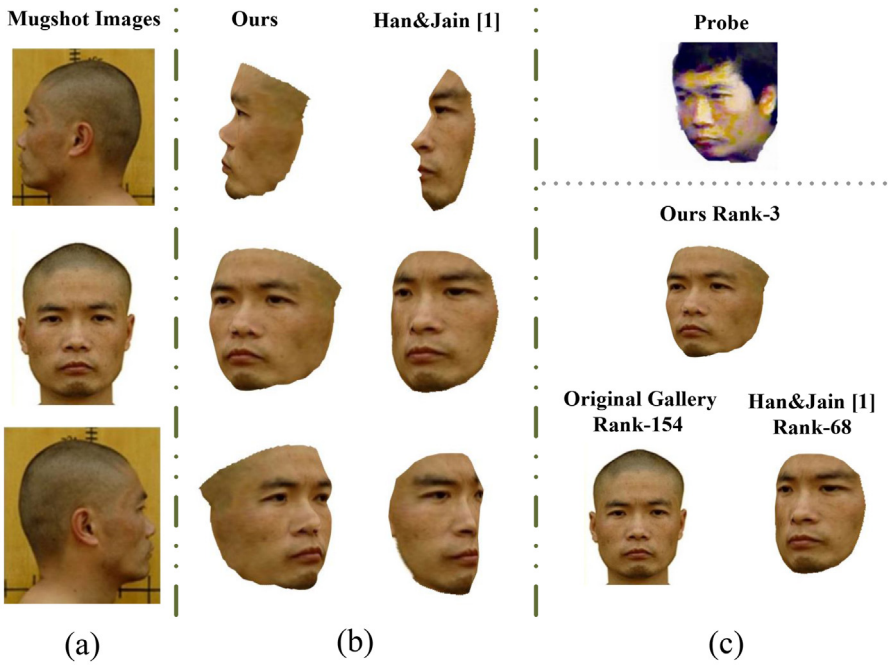

Fig. 1. (a) Mugshot images of a suspect (obtained from the Internet). (b) The reconstructed 3D faces in three different views by the proposed method and Han & Jain's method [1]. (c) Recognition results of SphereFace face matcher [20] in a gallery of tens of thousands of subjects for a probe image of the suspect. The first five rank results on the original gallery and the gallery enlarged by Han and Jain's method both find wrong subjects; the highest ranks of the true subject are 154 and 68, respectively in these two cases. Our method hits the true subject at rank 3.

3D face model. However, the calibration process in their method may fail and the final reconstructed shape may be severely affected.

Recently, there are some attempts to use deep learning to solve multi-view image (not exactly mugshot) based face reconstruction. Dou et al. [25] regress 3DMM parameters from both a deep convolutional neural network and a recurrent neural network that aggregate the identity specific contextual information in multi-view images. Wu et al. [26] regress 3DMM parameters from three-view inputs with an end-to-end Convolutional Neural Network leveraging a novel self-supervised view alignment loss. Both approaches only focus on shape reconstruction and are hardly applied to face recognition. Some researchers [1,2] have endeavored to utilize the mugshot images to improve the automated face recognition accuracy for arbitrary view images by reconstructing 3D face models. Despite the promising results they obtain, they mainly focus on shape reconstruction, but do not fully explore the texture on the three images. Further, the rotation angles of probe images they used are mostly within 70 degrees. Moreover, the baseline face matchers they use are traditional ones, and it is not clear how beneficial the mugshot images are with respect to the state-of-the-art deep learning (DL)-based face matchers, which have achieved significant progress in automated face recognition.

The goal of this paper is threefold: (i) improve the 3D face shape reconstruction accuracy via effectively exploiting the frontal and profile images in mugshot databases, (ii) generate dense full 3D face models with texture stitching from frontal and profile images, and (iii) investigate the effectiveness of mugshot images in improving the arbitrary view face recognition accuracy of state-of-the-art DL-based face matchers. To this end, we propose both linear and nonlinear regression approaches for reconstructing full 3D face shapes based on the facial landmarks on mugshot images, and an effective texture recovery method that can cope with potential illumination variations among the mugshot face images (see Fig. 2). Once the full textured 3D face models are obtained, we enlarge the gallery with multi-view face images generated from the models, and recognize probe face images based on the enlarged gallery. Extensive evaluation experiments demonstrate the superi-

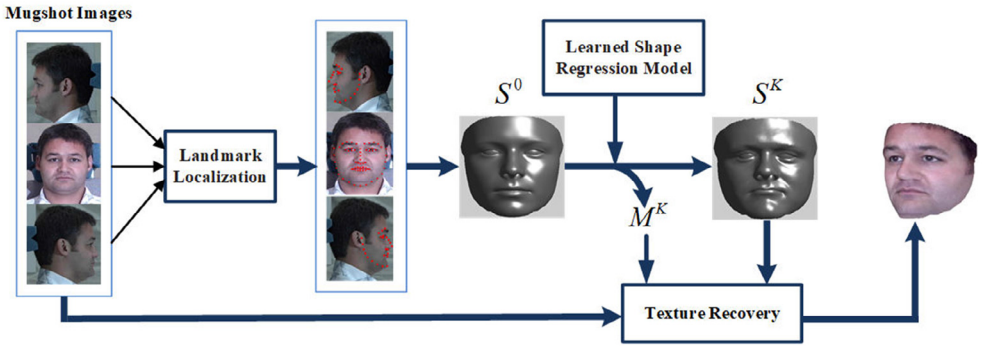

Fig. 2. Flowchart of the proposed mugshot-based reconstruction method. S^0 and S^K are, respectively, the initial and final estimated 3D models, and M^K are the 3D-to-2D projection matrix in the last iteration.

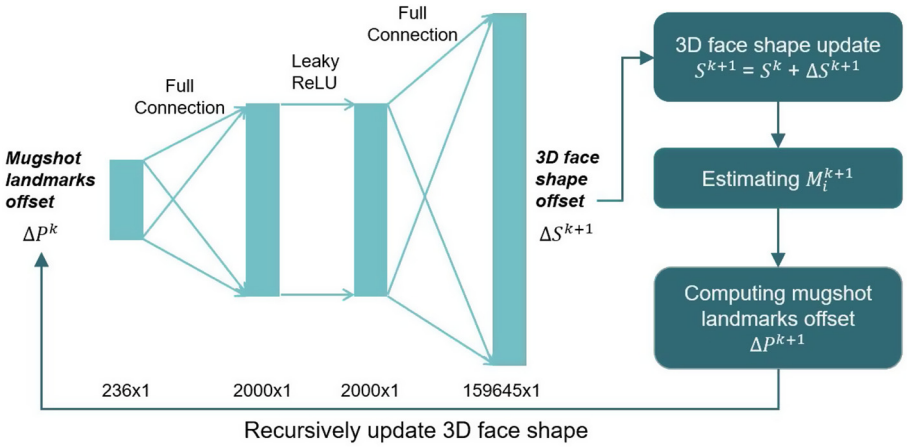

Fig. 3. Pipeline of the nonlinear implementation of the proposed shape reconstruction method.

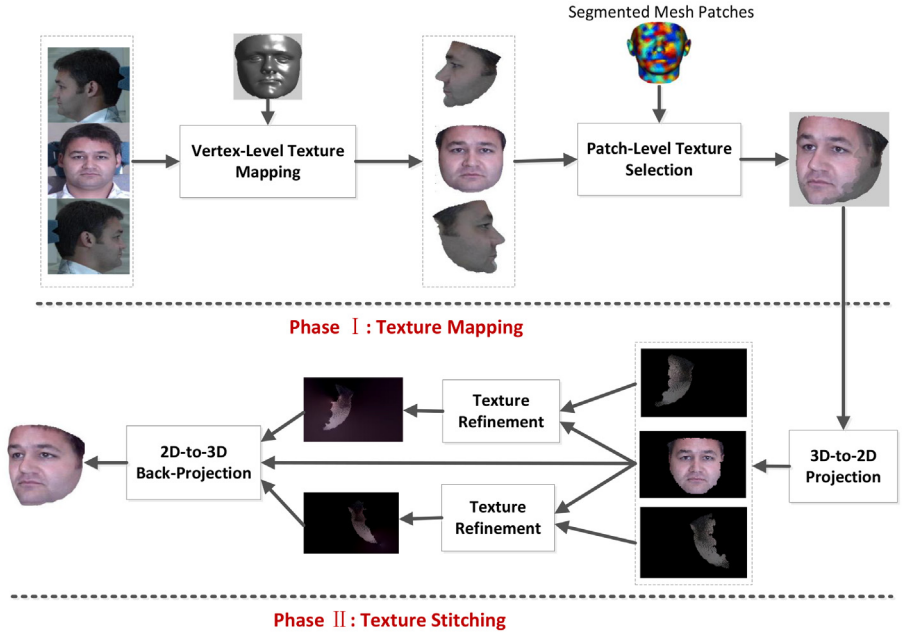

Fig. 4. Pipeline of the texture recovery module of the proposed method.

ority of our method over existing mugshot-based methods in reconstructing 3D faces, as well as the effectiveness of mugshot-generated 3D face models in improving arbitrary view face recognition accuracy of state-of-the-art DL-based face matchers. See Fig. 1 for an example.[2]

A preliminary version of this work was published in the 2018 24th International Conference on Pattern Recognition (ICPR) [27]. We extend this work from three aspects. (i) We implement the mugshot-based shape reconstruction method with both linear and non-linear regressors. (ii) We extend in detail the application of the proposed method to face recognition. (iii) We carry out more comprehensive evaluation with comparisons to state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 briefly reviews related work. Section 3 introduces the proposed mugshot-based 3D face reconstruction method in detail. Section 4 introduces the 3D-assisted face recognition method employed in this paper and Section 5 reports the evaluation results. Finally, a conclusion is drawn in Section 6.

2. Related work

Only a few studies on mugshot-based face reconstruction methods for automated face recognition have been reported in the public domain. The first work on exploring mugshot images for face recognition is due to Wallhoff et al. [28]. They proposed to synthesize profile face images from frontal face images through a neural network that was trained with pairs of frontal and profile face images in mugshot databases. They focused specifically on recognizing profile faces, rather than arbitrary view faces. The mugshot face images were used to train the profile face synthesis neural network, but used only frontal faces as gallery when recognizing

[2] The name of the suspect in Fig. 1 is Kehua Zhou. The mugshot images were captured in 2005 when he was arrested in Yunnan province of China. The probe image was from a surveillance video in 2011 when he committed a crime in Chongqing of China. At that time, the automated face matcher of the police failed to correctly identify the suspect.

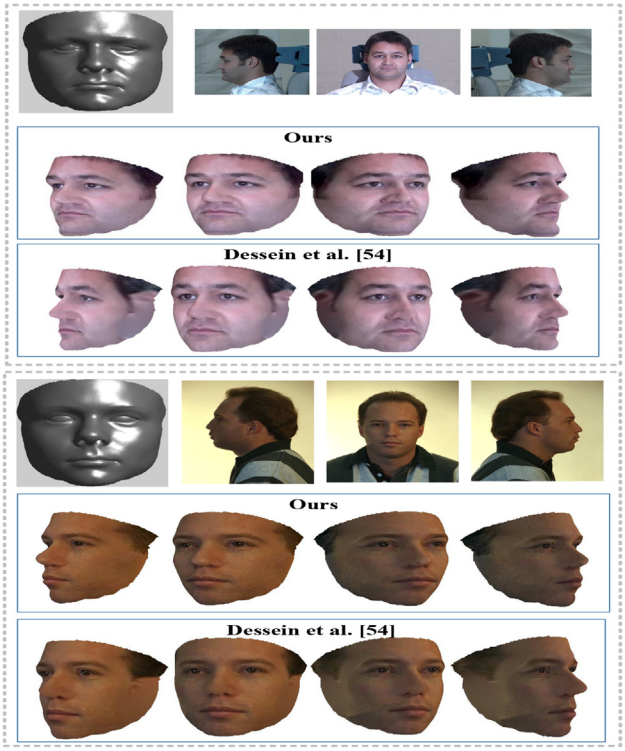

Fig. 5. Texture recovery results of our method and the method of Dessein et al. [54] for one subject in Multi-PIE and one subject in Color FERET. All 3D face models are shown in four different views.

profile faces. Next, we will introduce the related work from the aspects of mugshot-based reconstruction and pose-invariant face recognition.

2.1. Mugshot-based face reconstruction

Single-view face reconstruction has attracted lots of attentions and made a great progress [29–34,70]. For instance, Tran and Liu [29–31] proposed an innovative network to learn a nonlinear 3DMM model from a large set of in-the-wild face images. Liu [34] proposed a joint face alignment and 3D face reconstruction method by exploring the relationship between 2D landmarks and 3D shapes. Most of above methods are focused on the shape reconstruction, and there are a few works on mugshot-based face reconstruction.

Zhang et al. [2] presented a novel approach to recognize faces in arbitrary pose using frontal and side view mugshot face images as gallery images. They generated virtual view face images to enlarge the gallery based on the 3D face models reconstructed from the gallery mugshot images. To reconstruct the personalized 3D face model of a subject, they employed a hierarchical multilevel variation minimization approach for 3D shape modelling and pixel-wise texture analysis considering diffuse and specular reflections from human face. Their proposed method was evaluated on the CMU PIE database consisting of 68 subjects with PCA (Principal Component Analysis) based and LBP (Local Binary Patterns) based methods as baselines. The rotation angles of the probe images in their experiments are within 70 degrees.

Lee et al. [35] obtained the 3D shape by deforming a generic model in accordance with the extracted facial features from both frontal and profile face images. They first extracted salient features of the frontal and profile face images, by using ACM (Active Contour Model) and deformable ICP (Iterative Closest Point) methods, and then generated a 3D face model by deforming a generic model so that the 3D face model conforms to the extracted facial features. This method is limited in recovering fine facial details and could be easily dominated by the generic model. Notable deformations can be observed especially when the reconstructed model is rotated under large view point changes. As a consequence, the reconstructed 3D face model could not be utilized to assist in face recognition.

Han and Jain [1] proposed a 3DMM (3D Morphable Model) based 3D face reconstruction method. They first employed a simplified 3DMM to reconstruct the 3D face shape from the frontal face image, then refined the shape according to the landmarks on the profile face image, and finally directly mapped the texture from only the frontal face image to the reconstructed 3D shape. They evaluated the contribution of reconstructed 3D faces to face recognition in two ways, i.e., enlarging the gallery and normalizing the probe face images to frontal view. Their evaluation results showed that the face recognition accuracy of contemporary commercial face matchers was improved significantly by both these approaches. For the sake of efficiency, they used a sparse 3DMM that had a relatively small number of vertices in the 3D face model. As a result, the reconstructed 3D face shapes could have serious distortion especially when observed in profile views (see Fig. 1). Hence, the rotation angles of most of the probe images in their experiments were within 60 degrees. In addition, because the underlying 3DMM is a PCA based global statistical model, this

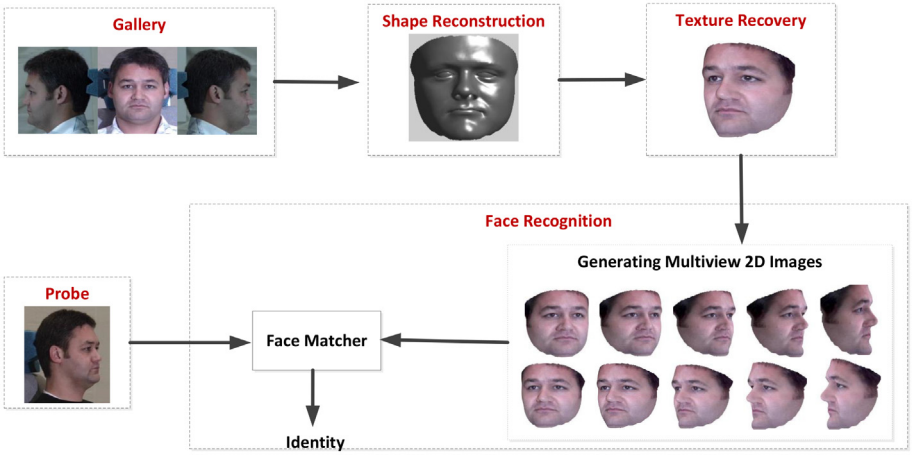

Fig. 6. Flowchart of the proposed mugshot-based arbitrary view face recognition method.

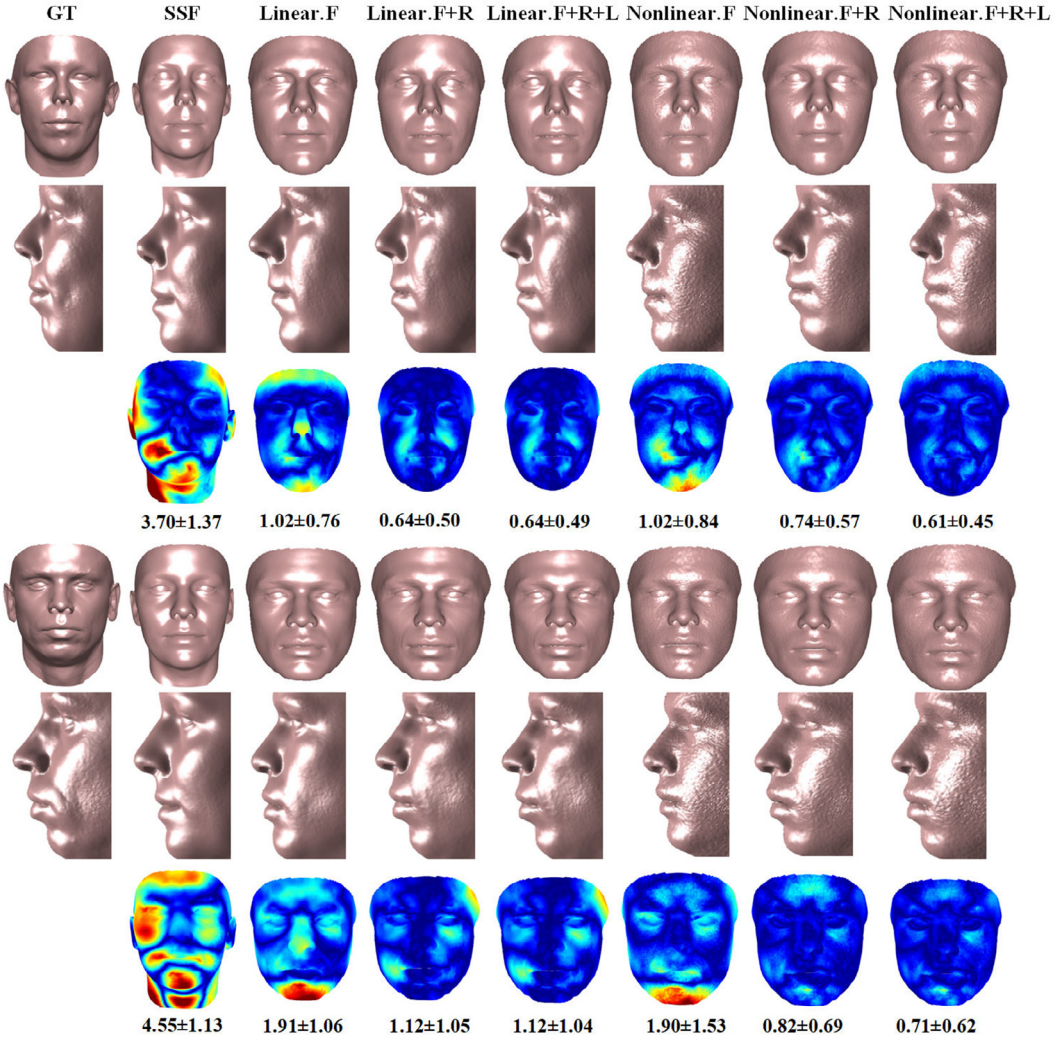

Fig. 7. Reconstruction result of different methods for subjects No.1 and No.8 in BFM. From left to right: the ground truth 3D face models, the reconstructed 3D face models by the SSF method [64] and our proposed method using only frontal view (denoted as 'F'), frontal and right profile views ('F + R'), and frontal and both right and left profile views ('F + R + L'). All 3D face models are shown in two different views. Error maps, mean and standard deviation of errors (in terms of PDE) are also shown. The colormap goes from dark blue to dark red (corresponding to an error between 0 and 5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Reconstruction errors (Mean Absolute Error (MAE)) of different 3D face reconstruction methods on the BFM test data. 'F' represents using only frontal view as input, 'F + R' using frontal and right profile views, and 'F + R+ L ' using frontal and both left and right profile views.

Method	Input Image	MAE (mm)
SSF [64]	F	6.18
MFF [63]	F	6.24
VRN [19]	F	4.96
3DSR [60]	F	2.31
3DMM-CNN [62]	F	2.46
Qu et al. [61]	F	7.34
	F + R + L	5.78
Proposed (Linear)	F	**2.13**
	F + R	1.92
	F + L	**1.91**
	F + R + L	1.88
Proposed (Nonlinear)	F	2.27
	F + R	**1.89**
	F + L	1.92
	F + R + L	**1.87**

The best results for different inputs are highlighted in bold.

method is limited in recovering fine details and could be easily dominated by the mean 3D face model.

Zeng et al. [36] proposed an exemplar based method for mugshot-based face reconstruction. The method first reconstructed a coarse 3D face model from each of the mugshot face image by using a shape from shading (SFS)-based approach, and then fused the three coarse 3D face models to form a fine 3D face model via an energy minimization process based on a diverse set of reference 3D face models. A limitation of this method is its high computational cost due to the involved online optimization. The authors [37] applied the method to face recognition by using the reconstructed 3D face shape to establish the correspondence between the semantic patches on the arbitrary view probe image and those on the gallery mugshot face images. They directly compared the LBP features of the corresponding patches, and fused the matching results of different patches to obtain the final decision on the probe identity. This method effectively utilized the texture information in both frontal and profile views. Evaluation results on the Bosphorus and Color FERET databases demonstrated its

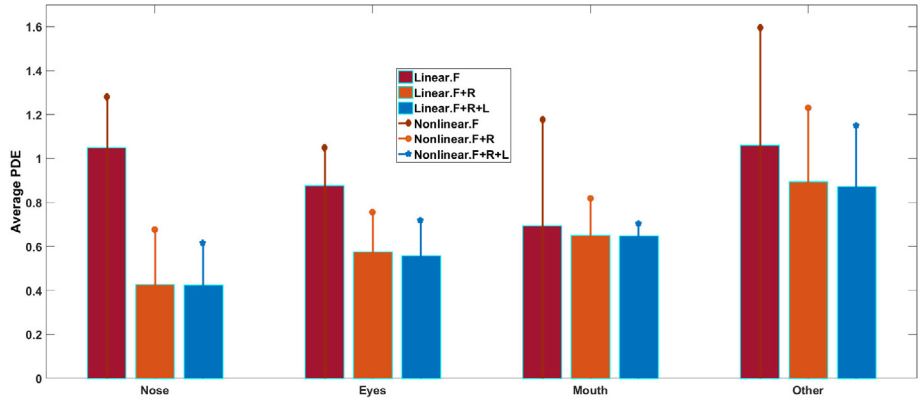

Fig. 8. The reconstruction errors on the BFM test data of our method over different face regions when only frontal ('F') images, frontal and right profile ('F + R') images, and frontal and both right profile and left profile ('F + R + L') images are used.

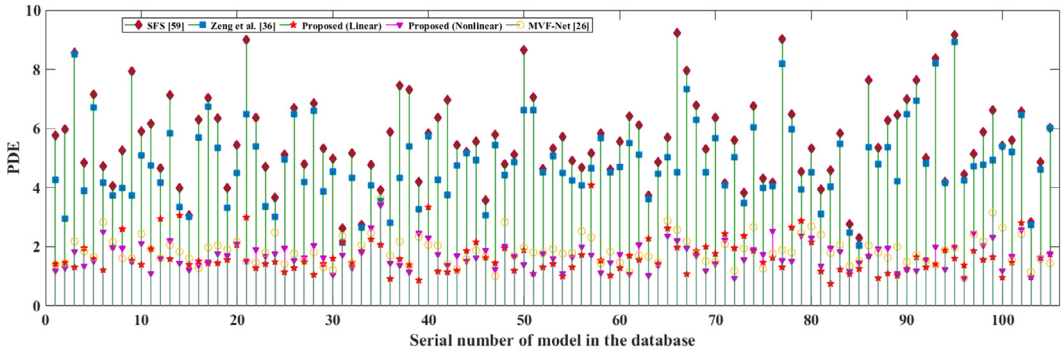

Fig. 9. Mean depth errors obtained by the SFS [59], examplar-based method [36], MVF-Net [26] and our proposed method for the 105 subjects in the Bosphorus database. The overall average errors of our proposed method are 1.66 for linear implementation and 1.69 for nonlinear implementation, while that of MVF-Net method is 1.86, the examplar-based method is 4.79, and SFS method is 5.60.

advantages over the baseline LBP-based matcher. However, the baseline face matcher used in their work is not state of the art.

Dou et al. [25] proposed a Deep Recurrent 3D Face Reconstruction (DRFAR) method to regress the 3DMM shape parameters from a set of facial images by a deep convolutional neural network (DCNN) and a recurrent neural network (RNN). The DCNN disentangles the facial identity and the facial expression components for each single image, while the RNN fuses identity-related features from the DCNN and aggregates the identity specific contextual information from the whole set of images to predict the facial identity 3DMM parameters. Wu et al. [26] proposed MVF-Net to extract identity features and pose parameters of three-view face images separately with a weighting-sharing CNN, and concatenate those identity features to regress the 3DMM shape parameters. The photometric reprojection error and optical-flow-based alignment error are used to supervise the training process. Same as DRFAR, MVF-Net focuses only on shape reconstruction, and is limited by the capacity of 3DMM. Consequently, both methods are hardly applicable to face recognition.

2.2. Pose-invariant face recognition

Pose variation is a major problem in face recognition, and existing approaches to this problem can be mainly divided into two categories. One is to extract pose-invariant features directly from the original images; the other is to normalize face images to frontal pose, and then feed the pose normalized images to feature extractors. Tran et al. [38,39] proposed DR-GAN which disentangles the identity features in arbitrary view images via an elaborated GAN framework, and uses these features for face recognition. Huang et al. [40] proposed a Two-Pathway Generative Adversarial Network (TP-GAN) to synthesize frontal facial images by simultaneously sensing global structures and local details. Zhao et al. [41,42] proposed a Dual-Agent Generative Adversarial Network (DA-GAN) model, which can improve the realism of synthetic profile face images while preserving the identity information. Zhao et al. [43] also proposed 3D-Aided Deep Pose-Invariant Face Recognition Model (3D-PIM), which automatically recovers realistic frontal faces from arbitrary poses through a 3D face model. Zhao et al. [44,69] proposed a Pose Invariant Model (PIM), which consists of Face Frontalization sub-network and Discriminative Learning sub-network. The two sub-networks are combined for end-to-end training to get better frontal images and identity feature representation. Different from the above methods, we propose in this paper to generate dense full 3D face models from mugshot images and enlarge the gallery to improve arbitrary view face recognition accuracy.

To summarize, this paper makes the following contributions:

- We propose a novel mugshot-based 3D face shape reconstruction method implemented by both linear and nonlinear regression, which effectively integrates and utilizes the information provided by the frontal and profile images in mugshot database.
- We generate dense full 3D face models with texture stitching from frontal and profile images, which eliminates texture inconsistency caused by varying illuminations in mugshot images.
- We improve the arbitrary view face recognition accuracy with 3D-enhanced method, by enlarging the gallery with multi-view face images generated from obtained full 3D face models.

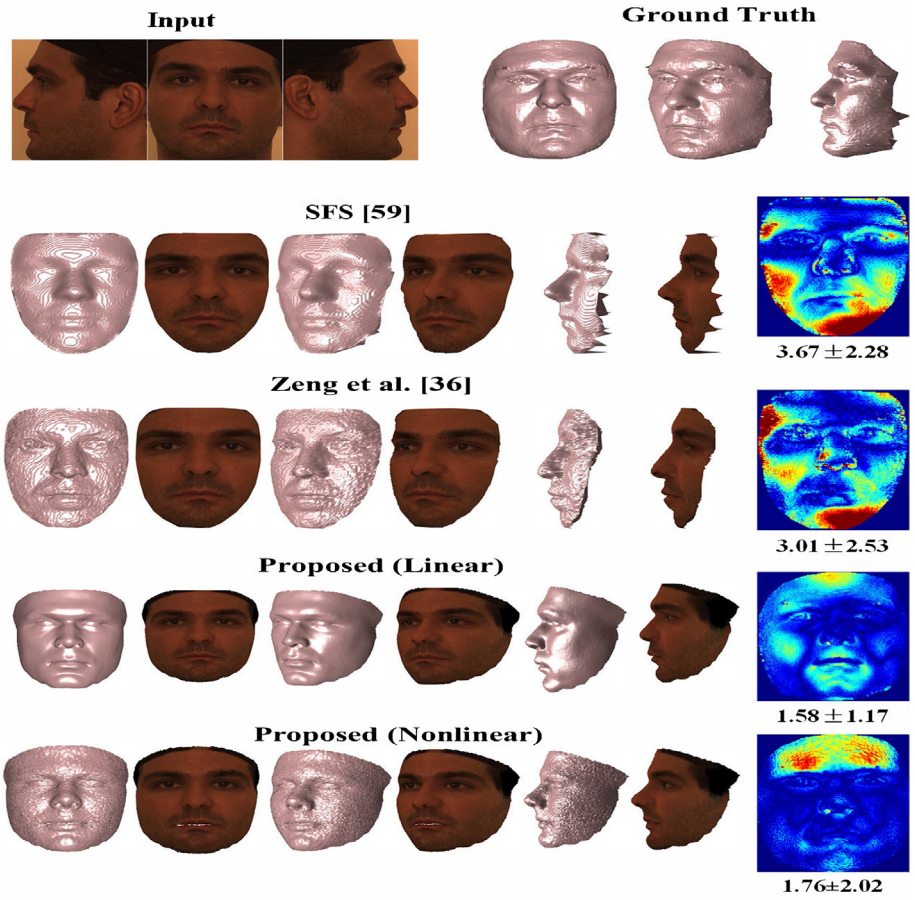

Fig. 10. Example reconstruction results on the Bosphorus database by SFS [59] (second row), examplar-based method [36] (third row), our proposed method (fourth and fifth rows). The first row shows the input mugshot images and the ground truth 3D model from different viewpoints. Error maps of the three methods are shown in the right-most heat map in the second to fourth rows.

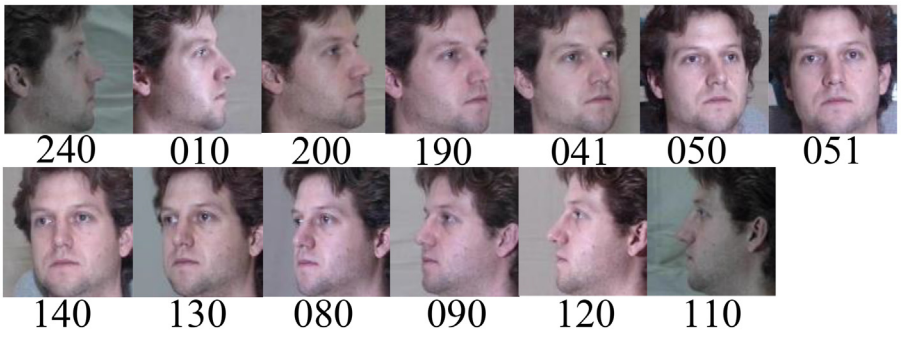

Fig. 11. Thirteen poses in Multi-PIE [47]. Labels under images indicate different yaw rotations: '240': 90°, '010': 75°, '200': 60°, '190': 45°, '041': 30°, '050': 15°, '051': 0°, '140': −15°, '130': −30°, '080': −45°, '090': −60°, '120': −75°, '110': −90°.

- We achieve state-of-the-art mugshot-based 3D face reconstruction performance on BFM [45] and Bosphorus [46] databases. We demonstrate the effectiveness of our proposed 3D-enhanced face recognition method in improving state-of-the-art deep learning based face matchers on Multi-PIE [47] and Color FERET [48] databases.

3. 3D face reconstruction

In this section, we introduce our proposed mugshot-based 3D face reconstruction method. As shown in Fig. 2, input to our method contains three mugshot face images of a person, including one frontal and two profile views. To recover a point-cloud-based full 3D face model with texture, the method first detects 2D facial landmarks on the mugshot images, then updates the reconstructed 3D face shape with either linear or nonlinear regressors such that the 3D face shape is consistent with the 2D landmarks, and finally computes the texture for each vertex in the 3D face shape by exploiting both frontal and profile face images with a patch-based texture stitching method. In this section, we will go through the details with emphasis on (i) 2D facial landmark extraction, (ii) 3D face shape reconstruction, and (iii) texture recovery.

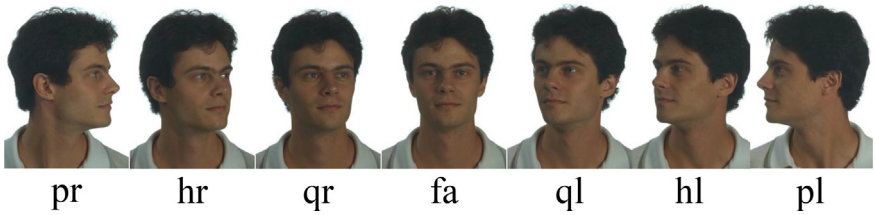

Fig. 12. Pose variations in the Color FERET database [48]. Labels under images indicate different yaw rotations: 'pr': 90°, 'hr': 67.5°, 'qr': 22.5°, 'fa': 0°, 'ql': −22.5°, 'hl': −67.5°, 'pl': −90°.

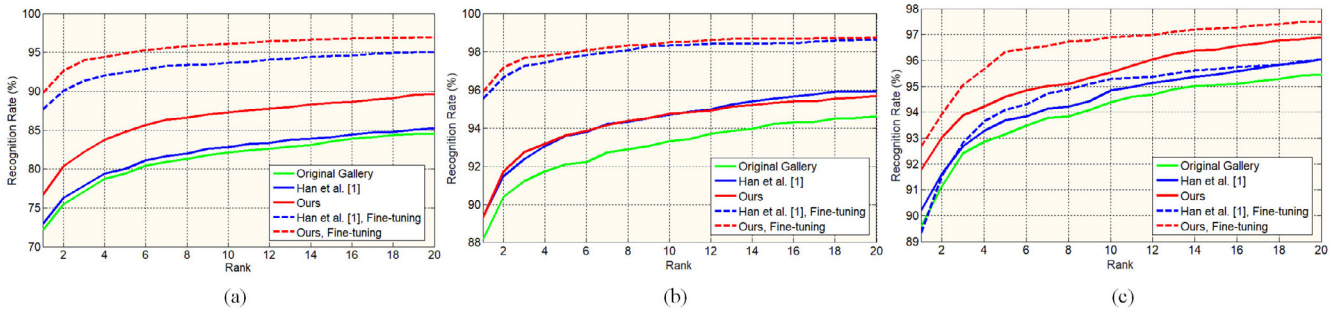

Fig. 13. CMC curves of the (a) LightCNN [65], (b) CenterLoss [66] and (c) SphereFace [20] matchers on Multi-PIE before and after enlarging the gallery, and with and without fine-tuning, using Han and Jain's and our methods.

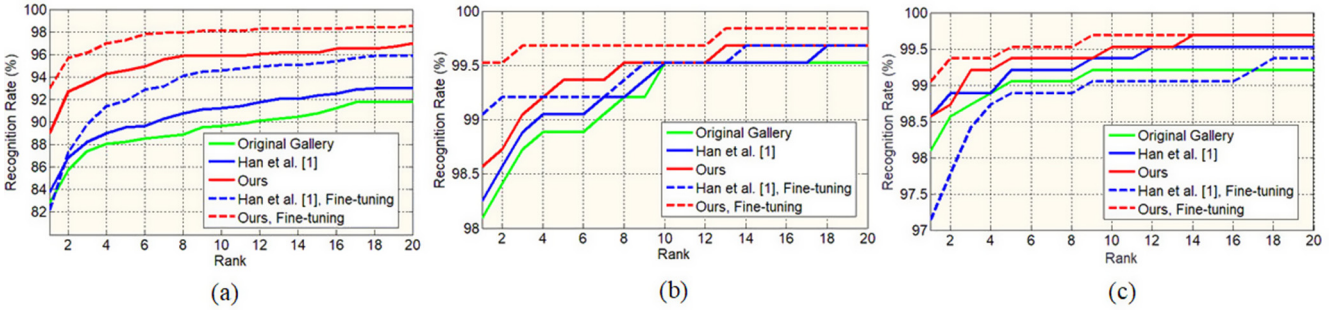

Fig. 14. CMC curves of the (a) LightCNN [65], (b) CenterLoss [66] and (c) SphereFace [20] matchers on Color FERET before and after enlarging the gallery, and with and without fine-tuning, using Han and Jain's and our methods.

3.1. 2D facial landmarks extraction

Our mugshot-based 3D face shape reconstruction method begins with extracting 2D facial landmarks on the mugshot face images. For the frontal face image, we extract a set of 68 facial landmarks using the DLIB implementation [49] of the algorithm in Ref. [50]. For profile face images, we extract 25 landmarks (see Fig. 2). However, since the DLIB model cannot be directly applied due to large pose variations, we re-train it with the face images in the 300 W-LP database [51] whose yaw angles are between 70° and 90°. Let P_F, P_R, and P_L denote the landmarks on frontal, right profile, and left profile face images, respectively. To fully utilize the correlation between the landmarks on frontal and profile faces, we concatenate them to form a unified 2D facial landmark vector $P=(P_F P_R P_L)^T=(u_1^F, v_1^F, \ldots, u_{68}^F, v_{68}^F, u_1^R, v_1^R, \ldots, u_{25}^R, v_{25}^R, u_1^L, v_1^L, \ldots, u_{25}^L, v_{25}^L)^T \in \mathbb{R}^{236\times 1}$ ('T' denotes transpose, and (u, v) 2D landmark coordinates) as input to the subsequent 3D face shape reconstruction step.

3.2. 3D face shape reconstruction

In this paper, we regress 3D face shape by exploiting relationship between 3D shape and its mugshot landmarks. Both linear regression and nonlinear regression are implemented to learn this 'relationship'. Next, we introduce the implementation in detail.

3.2.1. Reconstruction via Linear regression

In the process of 3D face shape reconstruction, we assume that dense correspondences have been established for 3D face shapes, and the indices of the vertices corresponding to facial landmarks are known. We represent the full 3D face shape of a subject as a vector $S=(x_1, y_1, z_1, \ldots, x_n, y_n, z_n)^T \in \mathbb{R}^{(3n)\times 1}$, where (x, y, z) are 3D coordinates of the shape vertex, and n is the total number of vertices. To reconstruct the 3D face shape, we start from an initial 3D face shape S^0 (e.g., the mean 3D face shape of training samples), and iteratively seek shape offsets ΔS to update the 3D face shape towards its true value. Motivated by the recent single-image-

based 3D face shape reconstruction methods [52,34], we estimate ΔS via regression over the deviations of 2D landmarks from their true positions.

Let P^* denote the set of detected 2D landmarks on mugshot images that are taken as the ground truth, and S^k be the reconstructed 3D face shape after k iterations. According to Ref. [52], S^k can be projected to 2D image plane to obtain its corresponding 2D landmarks P^k through weak perspective projection M. Note that the 3D-to-2D projection matrix is computed via least squares fitting for each of the frontal, right profile and left profile views such that the projections of the landmark vertices in S^k are as close as possible to the ground truth landmarks on mugshot images, i.e.,

$$M_j^k = \left(\left(D_j^k\right)^T D_j^k\right)^{-1}\left(D_j^k\right)^T P_j^*, \tag{1}$$

where $j \in \{f, l, r\}$ denotes the frontal, left or right view in mugshot images, and D^k denotes the corresponding landmark vertices in S^k.

The shape offset to S^k can be then calculated by

$$\Delta S^{k+1} = R^{k+1}\left(\Delta P^k\right), \tag{2}$$

where $\Delta P^k = \left(P^* - P^k\right)$ are 2D landmark deviations, and R^{k+1} is the regressor at $(k+1)^{\text{th}}$ iteration. The updated 3D face shape after $(k+1)$ iterations is finally obtained as

$$S^{k+1} = S^k + \Delta S^{k+1}. \tag{3}$$

The regressors $\left\{R^k\right\}$ involved in the above shape reconstruction process are learnt based on N training samples of mugshot images together with their ground truth 3D face shapes $\{S_i^* | i = 1, 2, \ldots, N\}$ and 2D landmarks $\{P_i^* | i = 1, 2, \ldots, N\}$. Specifically, R^k at the k^{th} iteration is obtained by solving the following optimization problem over the N training samples,

$$\arg\min_{R^k} \sum_{i=1}^{N} \| \left(S_i^* - S_i^{k-1}\right) - R^k\left(P_i^* - P_i^{k-1}\right) \|_2^2. \tag{4}$$

Under the assumption that $\left\{R^k\right\}$ are linear regressors, the above optimization problem can be solved by using least squares with a closed-form solution as

$$R^k = \Delta\mathbb{S}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T\left(\Delta\mathbb{P}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T\right)^{-1}, \tag{5}$$

where $\mathbb{S} \in \mathbb{R}^{(3n)\times N}$ and $\mathbb{P} \in \mathbb{R}^{236\times N}$ denote the composition of all training samples' 3D face shapes and 2D landmarks with each column corresponding to one sample, and $\Delta\mathbb{S}^{k-1} = \left(\mathbb{S}^* - \mathbb{S}^{k-1}\right)$ and $\Delta\mathbb{P}^{k-1} = \left(\mathbb{P}^* - \mathbb{P}^{k-1}\right)$ are, respectively, the 3D shape offsets and the 2D landmark deviations.

In order to avoid over-fitting, we further incorporate a regularization term into the objective function in Eq. (4), resulting in

$$\arg\min_{R^k} \sum_{i=1}^{N} \| \left(S_i^* - S_i^{k-1}\right) - R^k\left(P_i^* - P_i^{k-1}\right) \|_2^2 + \lambda \|R^k\|_2^2, \tag{6}$$

with its closed-form solution as

$$R^k = \Delta\mathbb{S}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T\left(\Delta\mathbb{P}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T + \lambda E\right)^{-1}, \tag{7}$$

where E is identity matrix and λ is the regularization parameter.

According to Eqs. (5) and (7), the inverse of $\Delta\mathbb{P}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T$ is required to compute the regressors. In order to accurately evaluate these two equations, the rank of $\Delta\mathbb{P}^{k-1}\left(\Delta\mathbb{P}^{k-1}\right)^T$ should not be larger than the number of training samples. Fortunately, as we use low-dimensional 2D features, it is easy to satisfy this requirement by using a small number of training samples. Algorithm 1 summarizes the process of learning the cascaded linear regressors.

Algorithm 1. Training process of learning the cascaded linear regressors

Input: Training data $\left\{\left(I_i, S_i^*, P_i^*\right)\right\} | i = 1, 2, \ldots, N\}$, initial shape S^0

Output: Cascaded regressors $\left\{R^k\right\}_{k=1}^K$

1: **for** $k = 1, \ldots, K$
2: Estimate weak perspective projection matrix M^{k-1} for each subject via Eq. (1), where D^{k-1} can be obtained from shape S^{k-1};
3: Compute 2D projection landmark $P^{k-1} = D^{k-1}M^{k-1}$;
4: Compute 2D landmark deviations and 3D face offsets for all samples: $\Delta P^{k-1} = P^* - P^{k-1}$, $\Delta S^{k-1} = S^* - S^{k-1}$;
5: Estimate cascaded regressor R^k via Eq. (7);
6: Update 3D face $S^k = S^{k-1} + R^k\left(\Delta P^{k-1}\right)$
7: **end for**

3.2.2. Reconstruction via nonlinear regression

The 'relationship' between 3D face shapes and mugshot landmarks can also be learned by nonlinear regressors. Here, we employ multiple layer perceptions (MLP) with LeakyReLU activation functions as the nonlinear regressors. As shown in Fig. 3, unlike the linear implementation, our nonlinear implementation updates the reconstructed 3D face shape in a recursive rather than cascaded way. Given the mugshot images of a subject, its 3D face shape is initialized as the average 3D face shape, and the 3D-to-2D projection matrices are estimated according to Eq. (1). The mugshot landmarks offset is taken as the input to the MLP, which consists of two full connection layers and a LeakyReLU activation layer. The output of the MLP is the offset required to update the currently reconstructed 3D face shape. After updating the 3D face shape according to Eq. (3), the 3D-to-2D projection matrices M_i are re-estimated, and the new mugshot landmarks offset is recalculated accordingly. This 3D face shape update procedure is recursively called with the same MLP until convergence.

To train the MLP, in each recursion call, the Euclidean loss between the estimated 3D face shape offset and the ground truth offset (i.e., the difference between the ground truth 3D face shape and the currently reconstructed 3D face shape at the begin of the recursion call) is employed. Note that the same MLP is shared between different recursion calls. Hence, once the training of one recursion call is completed, the new mugshot landmarks offset as well as the new ground truth 3D face shape offset are computed and used as the training data for the next recursion call. This way, the MLP as a nonlinear regressor for 3D face shape reconstruction is gradually fine-tuned to generate more accurate 3D face shapes.

3.3. Texture recovery

The purpose of texture recovery module is to assign texture value for each of the vertices in the reconstructed full 3D face shape. As shown in Fig. 4, our proposed texture recovery method consists of two phases: texture mapping and texture stitching. The texture mapping phase coarsely determines the texture values,

while the texture stitching phase refines the texture values to deal with the seams or texture inconsistency caused by varying illuminations present in mugshot images.

Texture mapping is done at two levels, i.e., vertex and patch levels. We first map, at vertex level, the texture values from each of the mugshot images to the reconstructed 3D face shape at its corresponding view angle, resulting in three partially textured 3D faces of frontal, right profile and left profile views. The correspondences between the shape vertices and the image pixels are determined by the projection matrices that are obtained during 3D face shape reconstruction (refer to Section 3.2). Inspired by Ref. [53], the visibility of each vertex is estimated according to the intersection angle between the surface normal at the vertex and the view direction from the face to the observer. Smaller intersection angles indicate higher visibility, and vertices with angles beyond 90° are invisible and thus do not contain any texture because no corresponding pixels exist for them in the mugshot images.

To remove the seams in the textured full 3D face is essentially a problem of stitching target and source textures (i.e., frontal and profile face images). While many methods [54–56] have been proposed for seamless texture stitching, only a few of them have been designed for 3D data. In addition, most of them operate in gradient domain and are thus time-consuming. Recently, a fast texture stitching method was proposed in Ref. [57]. It treats texture stitching as a re-sampling process of the source texture constrained by the target texture. It directly refines the source texture in the intensity domain rather than in the gradient domain. Specifically, it derives three filter kernels that can approximate the Poisson blending process in gradient domain, and applies pyramid convolutions to the source texture image by filtering the image with the three kernels at multiple scales. Motivated by this method, we first project the coarsely textured 3D face onto 2D images in three views, then refine the profile view texture according to the frontal view texture by using pyramid convolutions, and finally back-project the three 2D images onto the 3D face. Algorithms 2 and 3 summarize the processes of patch-level texture selection and texture refinement, respectively.

Algorithm 2. Process of patch-level texture selection and projection

Input: Textures of different views: $\{T_v|v\in\{f,l,r\}\}$, projection matrix of different views: $\{M_v|v\in\{f,l,r\}\}$, 3D face shape: S
Output: Projection images of selected texture fragments of different views: $\{a_v|v\in\{f,l,r\}\}$
1: Compute 100 overlapping patches (P) via method introduced in [58];
2: **for** p in patches P
3: Compute visibility of each vertice in p:
$vis=\frac{1}{2}\left(1+sgn\left(\vec{n}\cdot\left(\frac{M_1}{\|M_1\|}\times\frac{M_2}{\|M_2\|}\right)\right)\right)$
(sgn is sign function, $\vec{n}$ is normal, M_1 and M_2 are first three elements of the first and second rows, respectively);
4: Compute visibility of patch p on each view:
$\left\{vis_j^v=\sum_{j\in p}(vis_j|v\in\{f,l,r\}\right\}$;
5: Texture of p is set to be the value of the most visible one among three views;
6: **end for**
7: The selected patches on each view forms a texture fragment;
8: Project the selected texture fragments of each view and obtain images $\{a_v|v\in\{f,l,r\}\}$.

Algorithm 3. Process of texture refinement

Input: Projection images of selected texture fragments from different views: $\{a_v|v\in\{f,l,r\}\}$
Output: Refined texture fragments on left and right views: $\{\hat{a}_v^0|v\in\{l,r\}\}$
1: Determine the number of levels L and convolution filter h_1,h_2 and g according to [57];
2: **for** a in $\{a_v|v\in\{l,r\}\}$
3: Calculate the texture difference between a and a_f at the seam, and modify the texture value at the seam of a to the texture difference;
4: {Forward transform (analysis)}:
5: $a^0=a$;
6: **for** each level $j=0,\ldots,L-1$
7: $a_0^j=a^j$;
8: $a^{j+1}=\downarrow(h_1\circledast a^j)$ ($\downarrow$ denotes the downsampling operator, $\circledast$ denotes convolution operator);
9: **end for**
10: {Backward transform (synthesis)}:
11: $\hat{a}=g\circledast a^j$;
12: **for** each level $j=L-1,\ldots,0$
13: $\hat{a}^j=h_2\circledast(\uparrow\hat{a}^{j+1})+g\circledast a_0^j$ ($\uparrow$ denotes the upsampling operator);
14: **end for**
15: The $\hat{a}^0$ is the texture refinement result of a.
16: **end for**

Fig. 5 shows the recovered texture by our method and the state-of-the-art method proposed by Dessein et al. [54] for subjects in Multi-PIE and Color FERET databases. As can be seen, our obtained textured 3D faces no longer have apparent seams, and more importantly, retain facial details (especially in nose and cheek regions) that could be over-smoothed by the counterpart method. Moreover, the time complexity of our method is $O(n_p)$, where n_p is the number of image pixels, whereas that of the counterpart method is $O\left(n_p^3\right)$.

According to Ref. [54], patch-level texture mapping can better ensure local texture consistency than vertex-level texture mapping. Therefore, we use the method in Ref. [58] to segment the 3D face to 100 overlapping patches with a coefficient value $\sigma=0.6$ and a fast-marching-based farthest-point strategy. For each patch, its texture values are set to be the values of the most visible one among the corresponding patches on the aforementioned three partially textured 3D faces. Fig. 4 shows an example of coarsely textured full 3D face after texture mapping. Some obvious seams can be observed due to the inconsistent illuminations among different views of the input mugshot images.

4. Application to face recognition

Pose, illumination and expression (PIE) are well-known challenges in face recognition. With respect to the pose challenge, previous studies [2] show that face recognition accuracy would be obviously degraded as the facial pose becomes larger and enrolling subjects with face images of multiple views could enhance the robustness of face recognition to pose variations. Therefore, in this paper, we enlarge the mugshot gallery with multiple view face images that are generated by projecting the reconstructed textured full 3D faces onto 2D plane at different view angles. Given an arbi-

trary view probe 2D face image, its similarity with an enrolled subject is determined as the maximum of its match scores with the multi-view images of the subject, and its identity is decided as the subject that has the highest similarity with it. The flowchart of the proposed mugshot-based arbitrary view face recognition method is shown in Fig. 6.

5. Experiments

In this section, we evaluate the effectiveness and efficiency of the proposed method. First, we compare the proposed method with state-of-the-art methods in terms of 3D face reconstruction accuracy on BFM and Bosphorus databases. Then, we evaluate the contribution of the reconstructed 3D faces to face recognition with three deep learning based face matchers as baseline on Multi-PIE and Color FERET databases. Finally, we report the computational efficiency of the proposed method.

5.1. 3D face reconstruction accuracy

Databases and Metrics. We use the 3DMM model in the Basel Face Model (BFM) database [45] to randomly generate 3D faces of 1000 subjects, and synthesize multi-view 2D face images (of 1024×768 pixels) from them. The facial landmarks on these images are directly obtained during the synthesis process according to their corresponding vertices in the 3D face shapes (note that all the 3DMM-generated 3D faces have vertex-to-vertex dense correspondences). The landmarks together with the 3D face shapes are used as ground truth data to train our proposed method to learn the linear and nonlinear regressors for 3D face shape reconstruction. In the experiments of linear regression, we empirically set $\lambda = 3000$ in Eq. (6), and observe that the training process converges typically in five iterations, i.e., $K = 5$. The obtained linear regressors are also used in the face recognition experiments in the next subsection.

The Bosphorus [46] database and the BFM database are used to assess the 3D face shape reconstruction accuracy of our proposed method and several state-of-the-art methods,[3] including the improved shape-from-shading (SFS) method [59], an examplar-based method [36], a regressor based method (3DSR) [60], a volumetric CNN Regressor based method (VRN) [19], a multi-view 3DMM based method [61], a 3DMM CNN based method (3DMM-CNN) [62], a Multi-view 3DMM regression based method (MVF-Net) [26] and two improved single-view 3DMM methods, namely MFF [63] and SSF [64]. The BFM database provides 10 out-of-sample 3D faces for evaluation. The Bosphorus dataset has ground truth 3D faces of 105 subjects. In both datasets, we choose only the frontal and profile 2D face images as input, and compare the reconstructed models with the ground truth shapes in terms of two metrics: Mean Absolute Error (MAE) and Per-vertex Depth Error (PDE). MAE is defined as

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left(\|S_i^* - \widehat{S}_i\|/n\right), \tag{8}$$

where S_i^* is the ground truth 3D shape of the i^{th} model out of a total of m test samples, $\widehat{S}_i$ is the corresponding reconstructed shape, and n is the number of points in 3D face shape. PDE is computed by

$$E_z(x_i, y_i) = \mid z_i^*(x_i, y_i) - \hat{z}_i(x_i, y_i) \mid, \tag{9}$$

where z_i^* and $\hat{z}_i$ are the ground truth and reconstructed depth values of the i^{th} vertex.

Results on BFM. Fig. 7 shows the reconstruction results of our method compared with SSF on the BFM dataset. As the training data used in MVF-Net[26] includes 300W-LP dataset that is based on BFM, it's not fair to compare with MVF-Net on this dataset. The reconstruction error in terms of PDE is also plotted. Both linear and nonlinear implementations of our proposed method can produce more visually pleasing results. Compared with SSF, our method leads to much smaller reconstruction errors.

Table 1 summarizes the reconstruction accuracy of different methods. Note that for the nonlinear implementation of our proposed method, we recursively call the MLP for two times. According to our experimental results, when applying the MLP only once without recursive calls, the MAEs are 2.28, 1.94, 1.95 and 1.93 for using frontal, frontal and right profile, frontal and left profile, and frontal and both left and right profile face images, respectively, which are worse than the MAEs when recursively calling the MLP. When only frontal 2D images are used as input, our method is obviously better than the existing state-of-the-art 3DMM-based methods. When multi-view mugshot images are used, our method surpasses the counterpart multi-view 3DMM-based method with a large margin. These results demonstrate the superiority of our proposed method in fully utilizing the information in multi-view mugshot face images.

We can also see from Table 1 that using only right-profile images or left-profile images and using both right-profile and left-profile images do not make significant differences in enhancing the reconstruction accuracy. This is reasonable considering that the two profile views are nearly mirror views of each other. We further investigate the contribution of additional profile images on the reconstruction of different face regions. The results are shown in Fig. 8. Note that the nose region benefits the most. This is probably because the profile views provide richer geometric details for the nose region than for other regions.

Results on Bosphorus. Fig. 9 shows the average depth error of our proposed method as well as the SFS [59], examplar-based method [36] and MVF-Net [26] over the 105 subjects in the Bosphorus database. It can be seen that our method achieves the lowest error on almost all the subjects. Furthermore, our method successfully reduces the average reconstruction error from 5.60 to 1.66 (Linear) and 1.71 (Nonlinear). Some example results of reconstruction are shown in Fig. 10. Compared with SFS and examplar-based method, our proposed method can produce full 3D face models with overall smoothness and detailed geometry.

5.2. Face recognition accuracy

5.2.1. Databases and protocols

The Multi-PIE face database [47] and the Color FERET face database [48] are used for face recognition accuracy evaluation. To follow the real-world applications in law enforcement agencies, the galleries in our experiments are further expanded by 30,000 mugshot photographs (three images per subject) from a private database (which can not be published due to copyright and privacy issues).

The Multi-PIE database contains 755,370 images from 337 subjects under various poses, illuminations and expressions. These face images were captured in four sessions during different periods. In our experiments, we only consider the effect of arbitrary poses on face recognition. Therefore, we choose the 11,921 images from 337 subjects with 13 poses, normal illumination and neutral expression to evaluate the arbitrary view face recognition accuracy, and take the frontal view (pose 051) and side views (poses 110 and 240) of each subject from a session as the gallery mugshot images (see Fig. 11). Note that the gallery images are chosen from

[3] Most of these methods are designed for single-image-based reconstruction. For a fair comparison, we reconstruct 3D face shapes for each of the three views in mugshot images using these methods, and we find the frontal view can get the best accuracy, so we report the accuracy of frontal view only.

Table 2
Rank-1 identification rates (%) of different methods at different poses of probe images on Multi-PIE. The best results of each face matcher at each pose are highlighted in bold.

Method	Matcher	Pose of probe images												Average
		−90°	−75°	−60°	−45°	−30°	−15°	0°	15°	30°	45°	60°	75°	90°	
Original Gallery	LightCNN	48.4	31.4	56.8	**93.9**	**99.0**	**99.5**	**100**	**100**	**99.5**	90.6	45.0	20.0	30.1	70.35
Han and Jain [1]		**75.7**	84.3	86.8	92.0	92.5	94.4	90.4	94.4	92.4	90.6	86.8	75.7	**65.5**	86.30
Ours		69.7	**87.1**	**93.0**	93.4	93.0	95.8	94.3	97.2	94.4	**94.4**	**90.1**	**78.6**	62.4	**87.94**
Original Gallery	SphereFace	**72.7**	71.9	89.2	99.0	**100**	**100**	**100**	**100**	**100**	98.1	81.6	68.5	65.5	88.22
Han and Jain [1]		65.6	76.6	89.2	**99.1**	99.5	**100**	**100**	**100**	**100**	96.7	81.6	70.4	55.9	87.30
Ours		69.7	**86.7**	**92.0**	98.1	99.5	99.5	**100**	99.5	**100**	**98.6**	**90.6**	**80.0**	**67.7**	**90.93**
Original Gallery	CenterLoss	74.7	82.8	85.9	92.0	92.4	93.9	90.4	94.3	92.4	90.1	84.5	74.2	65.5	85.68
Han and Jain [1]		83.8	**95.2**	96.2	97.2	**99.0**	99.0	**98.1**	99.1	99.1	98.5	92.9	90.4	78.5	94.41
Ours		**84.8**	93.3	**97.2**	**98.6**	98.6	**99.1**	97.1	**99.1**	**99.5**	**98.5**	**94.8**	**90.9**	**81.7**	**94.88**
LDF-Net [67]	LDA	63.9	87.3	93.0	98.1	98.6	97.2	-	**100**	99.1	98.6	94.4	85.0	66.7	90.1

Table 3
Rank-1 identification rates (%) of different methods at different poses of probe images on Color FERET. The best results of each face matcher at each pose are highlighted in bold.

Method	Matcher	Pose of Probe Images					Average
		22.5°	67.5°	−22.5°	−67.5°	90°	
Original Gallery	LightCNN	**100**	65.1	**100**	73.4	26.6	73.05
Han and Jain [1]		92.6	75.9	95.3	70.9	33.3	73.62
Ours		99.3	**91.1**	97.9	**89.9**	**33.3**	**82.33**
Original Gallery	SphereFace	**100**	98.1	**100**	96.2	**80.0**	94.86
Han and Jain [1]		**100**	94.9	**100**	95.5	**80.0**	94.10
Ours		**100**	**99.3**	**100**	**98.7**	**80.0**	**95.62**
Original Gallery	CenterLoss	**100**	96.8	**100**	97.4	80.0	94.86
Han and Jain [1]		**100**	97.2	**100**	**100**	83.3	96.11
Ours		**100**	**97.5**	**100**	**100**	**86.7**	**96.83**

one session while the probes are collected in other sessions. The Color FERET database contains a total of 11,338 facial images from 994 subjects at various angles, over the course of 15 sessions between 1993 and 1996. We conduct experiments on a subset of the Color FERET database because not each subject in the Color FERET database has mughsot images. Similarly, we choose the 2974 images from 343 subjects with various angles to evaluate the face recognition accuracy. Among them, a frontal face image (pose fa) and two profile face images (poses pl and pr) of each subject are used as the gallery mugshot images (see Fig. 12). From the mugshot images, 3D faces are reconstructed with linear regressors and used to generate multi-view 2D face images with yaw rotation angles ranging from −90° to +90° at an interval of 15° to enlarge the gallery.The face recognition is conducted in identification mode, and rank-1 identification rate and cumulative match characteristic (CMC) curves are reported. In order to investigate the scalability of our proposed method, we consider two galleries, one of which is an expansion of the other with additional 10,000 subjects.

5.2.2. Improved face recognition by enlarging gallery

To evaluate the effectiveness of the proposed method in enhancing face recognition, we enlarge the mugshot gallery with multiple view face images that are generated by reconstructed 3D model. we employ the following state-of-the-art deep learning based face matchers, including the Lightened Convolutional Neural Networks (LightCNN) [65], the CenterLoss model (CenterLoss) [66], and the SphereFace method (SphereFace) [20], which are available in the public domain. We conduct two series of experiments to (i) compare the performance of these matchers before and after enlarging the gallery, and (ii) investigate the effectiveness of fine-tuning the matchers with the generated multi-view images. The training data for fine-tuning are chosen in a similar way to the Setting-I protocol in [67]. Specifically, for the Multi-PIE database we choose the images of the first 229 subjects as training data, and the images of the remaining 108 subjects as test data (~23 images per subject for test), and for the Color FERET database we choose the images of the first 200 subjects as training data, and the images of the remaining 143 subjects as test data (~5 images per subject for test). To demonstrate the superiority of our reconstructed 3D face models, we compare our method with one existing mugshot based face recognition method in [1] and one state of the art cross-pose face recognition method in [67].

Figs. 13 and 14, respectively, show the CMC curves of different methods based on Multi-PIE database and Color FERET database. 'Original Gallery' in the figures indicates the results of the original models on un-enlarged gallery. From these results, the following two observations can be made. (i) The accuracy of the DL-based matchers is improved after enlarging the gallery, and our method provides more improvement than the counterpart method in Ref. [1]. We believe that this is because our method effectively exploits both the shape and texture information in mugshot images. (ii) Fine-tuning the original DL-based face matchers with the generated multi-view face images further improves the recognition accuracy with a large margin in most cases,[4] and our method is consistently better in terms of rank-1 identification rate.

Tables 2 and 3 give the rank-1 identification rates of different methods under various pose variations of probe images on Multi-PIE database and Color FERET database, respectively. While the face matchers deteriorate significantly with the pose variations in probe images on the original gallery, they become more robust to pose variations after using the reconstructed 3D faces to enlarge the gallery and to fine-tune the matchers. Moreover, our method achieves the best rank-1 identification rate, on average, among

[4] The results of SphereFace become worse after being fine-tuned with the images generated by the method in Ref. [1]. One possible reason is because SphereFace is more sensitive to the low quality images in the relatively small training dataset, while the images generated by the method in [1] at large poses look unrealistic with obvious artifacts.

Table 4
Rank-1 identification rates (%) of different methods for small and large galleries, and the improvement made under different enlarged galleries. The entries shown as "–" in the table indicate no improvement.

Matcher	Gallery (No. of Subjects)	Rank-1 Accuracy of Different Methods			Improvement (%) y Han and Jain [1]	Improvement (%) by Ours
		Han and Jain [1]	Ours	Original Gallery		
LightCNN	Multi-PIE (337)	89.55	92.15	80.05	9.5	12.1
	Multi-PIE (337) + Private (10,000)	87.65	89.8	72.16	15.49	17.64
	FERET (343)	85.85	94.28	87.6	–	6.68
	FERET (343) + Private (10,000)	82.19	93	82.83	–	10.17
CenterLoss	Multi-PIE (337)	96.74	97.44	93.52	3.22	3.92
	Multi-PIE (337) + Private (10,000)	95.54	95.99	88.19	7.35	7.8
	FERET (343)	99.52	100	99.36	0.16	0.64
	FERET (343) + Private (10,000)	99.05	99.52	98.09	0.96	1.43
SphereFace	Multi-PIE (337)	92.52	95.21	93.68	–	1.53
	Multi-PIE (337) + Private (10,000)	89.34	92.69	89.59	–	3.1
	FERET (343)	98.57	99.68	99.05	–	0.63
	FERET (343) + Private (10,000)	97.14	99.05	98.09	–	0.96

Table 5
Rank-1 identification rates (%) for probe images of Mongolian race and other races.

Matcher	Gallery	Percentage of Mongolians in Gallery (%)	Probe images of Mongolian race			Probe images of other races		
			Ours	Original Gallery	Improvement (%)	Ours	Original Gallery	Improvement (%)
LightCNN	Multi-PIE	26.36	90.83	72.84	17.99	92.57	82.31	10.26
	Multi-PIE + Private	99.19	86.61	60.38	26.23	90.94	75.85	15.09
	FERET	8.96	96.43	78.57	17.86	94.07	88.48	5.59
	FERET + Private	98.69	87.5	66.07	21.43	93.54	84.47	9.07
CenterLoss	Multi-PIE	26.36	96.54	85.47	11.07	97.72	96.04	1.68
	Multi-PIE + Private	99.19	91.7	70.24	21.46	97.23	93.81	3.42
	FERET	8.96	100	98.21	1.79	100	99.48	0.52
	FERET + Private	98.69	96.43	85.71	10.72	99.83	99.3	0.53
SphereFace	Multi-PIE	26.36	92.56	88.06	4.5	96.04	95.44	0.6
	Multi-PIE + Private	99.19	84.43	75.09	9.34	95.28	94.14	1.14
	FERET	8.96	98.21	98.21	0	99.83	99.13	0.7
	FERET + Private	98.69	92.86	91.07	1.79	99.65	98.78	0.87

the methods considered.

Table 4 compares the improvement made by our method and that made by the method in Ref. [1] when different numbers of subjects are in the gallery. As can be seen, as the gallery becomes larger, the face recognition accuracy improves more for both methods. This is because it is more likely to hit wrong subjects in a larger gallery if no additional information is explored. While both our method and the method in Ref. [1] enlarge the gallery with images of extra views, our method can improve the face recognition accuracy more significantly due to its ability to generate better quality synthetic multi-view images.

Considering that the subjects in the private database are all of Mongolian race, we expect that the recognition accuracy of the Mongolian probe images is improved more than that of others. To investigate such *race effect*, we divide the probe images into two subsets according to whether they are of Mongolian race or not, and calculate the rank-1 identification rates for them separately. The results are shown in Table 5. Not surprisingly, the improvement for Mongolian probe images is obviously higher, since a larger portion of the gallery is Mongolian. All these results demonstrate the superior scalability of our method.

5.2.3. Further discussion on face recognition

The above experiments show that our proposed method obtains better overall face recognition accuracy; however, it does not consistently work across all pose directions. Based on this observation, we argue that the synthetic images and the original gallery images can complement each other. Yet, due to the modality gap between

Table 6
Rank-1 identification rate (%) for different methods at different poses of probe images on Multi-PIE. The best results at each pose are highlighted in bold.

Method	Pose of Probe Images						Average
	±90°	±75°	±60°	±45°	±30°	±15°	
TP-GAN [40]	64.0	84.1	92.9	98.6	99.9	99.8	89.9
3D-PIM [43]	76.1	94.3	98.8	99.3	99.5	99.8	94.7
PIM [44]	75.0	91.2	97.7	98.3	99.4	99.8	93.6
Original Gallery	65.4	88.6	98.3	**99.8**	**100.0**	**100.0**	92.0
Ours without Fusion	66.2	86.0	97.0	99.5	99.9	99.9	91.4
Ours with Fusion	**81.5**	**94.1**	**99.1**	**99.8**	**100.0**	**100.0**	**95.8**

Table 7
Testing time and model size of different methods. The entries shown as "–" in the table indicate that the corresponding metrics are not provided or unavailable.

Method	Test Time (s)		Model Size (MB)	
	Shape	Texture	Shape	Texture
Zhang et al. [2]	985.80	225.60	–	–
MVF-Net [26]	0.25	–	33.4	–
Ours	0.04	1.20	702.5	3.7

them, simply using the synthetic images to enlarge the gallery might not be the best way to improve face recognition accuracy. Therefore, in this experiment, we modify the implementation of the face recognition method by matching the probe against the original and the synthetic images in the enlarged gallery, respectively. Once the comparison between the probe and the images of an object is done, we sum up the maximum score among the synthetic images and the maximum score among the original images as the final match score between the probe and the subject. The identity of the probe is finally determined as the subject that has the highest match score with the probe. We refer to this implementation of our method as ours with fusion. For a fair comparison with existing state-of-the-art pose-invariant face recognition methods, in this experiment we use the CosFace [68] as the face matcher and conduct comparison evaluation on Multi-PIE by using the images of 250 subjects with neutral expression in session one. The images with 13 poses within ±90° and under 20 illumination situations of the first 150 identities are used for fine-tuning. For testing, one frontal view with neutral expression and normal illumination (i.e., ID07) is used as the gallery image for each of the remaining 100 identities, and the rest images are used as probes.

Table 6 gives the evaluation results. As can be seen, although the previous implementation in Section 5.2.2 is not consistently better than the counterpart methods, the implementation of our method in this experiment can better utilize the complementary features in the original and synthetic gallery images, and thus obtain consistently better recognition accuracy under varying pose angles than the existing state-of-the-art methods.

5.3. Computational efficiency

In order to evaluate the computational efficiency of our method, here, we analyze the computational complexity of the linear implementation of our proposed method. According to the closed-form solution in Eqs. (5) and (7), the involved computation mainly includes three matrix multiplications and one matrix inverse. Let N be the number of training samples, n the number of vertices in the output 3D model, and m the number of used facial landmarks. The time complexity is then of the order $O(3nmN + m^2N + m^2N + m^3)$, in which the first three terms correspond to the three matrix multiplications and the last term to the matrix inverse. Note that our method uses a sparse set of facial landmarks. Thus the training complexity is primarily linear to the number of training samples and the point cloud density of the output 3D model. Further, we execute the MATLAB implementation of the linear version of our proposed 3D face reconstruction method on a PC with i7-4710 CPU and 16 GB memory. When using 1000 samples to train our method, it takes 133 s (s) to converge. The testing time and the model size of our method are summarized in Table 7 with comparison to some existing methods. Obviously, our method is more efficient in terms of running time. However, the space complexity of our method is relatively higher.

6. Conclusion

We have proposed a novel method for reconstructing textured full 3D faces from mugshot images (frontal and profile views). In our method, personalized face models are reconstructed via a linear or nonlinear regression pipeline for 3D shape reconstruction and an efficient texture recovery module. Extensive experimental results have demonstrated that our proposed method can generate more accurate 3D face shapes and more realistic facial texture for the full 3D faces. An application of the reconstructed textured full 3D faces to arbitrary view face recognition is presented. In the recognition experiments, multi-view 2D face images are generated from the textured full 3D faces and used to enlarge the gallery to improve the arbitrary view face recognition accuracy. Our recognition results demonstrate the effectiveness of the proposed method, and show that DL-based face matchers, though being more robust to pose variations than conventional face matchers, do benefit from the textured full 3D faces reconstructed from mugshot images. A more significant improvement over the recognition accuracy can be obtained after they are fine-tuned with the generated multi-view face images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jie Liang: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Investigation, Validation, Visualization. **Huan Tu:** Conceptualization, Methodology, Software, Writing - original draft, Investigation, Validation. **Feng Liu:** Conceptualization, Methodology. **Qijun Zhao:** Conceptualization, Supervision, Writing - review & editing, Funding acquisition. **Anil K. Jain:** Supervision, Writing - review & editing.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2017YFB0802300), the National Natural Science Foundation of China (61773270, 61971005), and the Miaozi Key Project in Science and Technology Innovation Program of Sichuan Province (No. 2017RZ0016).

References

[1] H. Han, A.K. Jain, 3D face texture modeling from uncalibrated frontal and profile images, in: Proceedings of the IEEE Conference on Biometrics: Theory, Applications and Systems, 2012, pp. 223–230.
[2] X. Zhang, Y. Gao, M.K. Leung, Recognizing rotated faces from frontal and side views: an approach toward effective use of mugshot databases, IEEE Transactions on Information Forensics and Security 3 (4) (2008) 684–697.
[3] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1063–1074.
[4] X. Zhu, Z. Lei, J. Yan, D. Yi, S.Z. Li, High-Fidelity pose and expression normalization for face recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 787–796.
[5] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu, M. Rohith, Fully automatic pose-invariant face recognition via 3d pose normalization, in: Proceedings of the IEEE Conference on Computer Vision, 2011, pp. 937–944.
[6] C. Ding, D. Tao, Pose-invariant face recognition with homography-based normalization, Pattern Recognition 66 (2017) 144–152.
[7] I. Masi, S. Rawls, G. Medioni, P. Natarajan, Pose-aware face recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4838–4846.
[8] D. Yi, Z. Lei, S.Z. Li, Towards pose robust face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3539–3545.

[9] Y. Hu, D. Jiang, S. Yan, L. Zhang, et al., Automatic 3d reconstruction for face recognition, in: Proceedings of the IEEE Conference on Automatic Face & Gesture Recognition, 2004, pp. 843–848.

[10] X. Liu, T. Chen, Pose-robust face recognition using geometry assisted probabilistic modeling, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 502–509..

[11] H.T. Ho, R. Chellappa, Pose-invariant face recognition using markov random fields, IEEE Transactions on Image Processing 22 (4) (2013) 1573–1584.

[12] X. Xu, H.A. Le, P. Dou, Y. Wu, I.A. Kakadiaris, Evaluation of a 3d-aided pose invariant 2d face recognition system, in: Proceedings of the IEEE International Joint Conference on Biometrics, 2017, pp. 446–455.

[13] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3d dynamic facial expression database, in: Proceedings of the IEEE Conference on Automatic Face & Gesture Recognition, 2008, pp. 1–6.

[14] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the IEEE Conference on Computer vision and pattern recognition, vol. 1, 2005, pp. 947–954..

[15] C. Qu, E. Monari, T. Schuchert, J. Beyerer, Adaptive contour fitting for pose-invariant 3D face shape reconstruction, in: Proceedings of the British Machine Vision Conference, 2015, pp. 87.1–87.12..

[16] J. Jo, H. Choi, I.-J. Kim, J. Kim, Single-view-based 3d facial reconstruction method robust against pose variations, Pattern Recognition 48 (1) (2015) 73–85.

[17] M. Piotraschke, V. Blanz, Automated 3d face reconstruction from multiple images using quality measures, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3418–3427.

[18] P. Dou, S.K. Shah, I.A. Kakadiaris, End-to-end 3d face reconstruction with deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 5, 2017..

[19] A.S. Jackson, A. Bulat, V. Argyriou, G. Tzimiropoulos, Large pose 3d face reconstruction from a single image via direct volumetric cnn regression, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1031–1039.

[20] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: deep hypersphere embedding for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[21] H.H. Ip, L. Yin, Constructing a 3D individualized head model from two orthogonal views, The Visual Computer 12 (5) (1996) 254–266.

[22] A.-N. Ansari, M. Abdel-Mottaleb, Automatic facial feature extraction and 3d face modeling using two orthogonal views with application to 3D face recognition, Pattern Recognition 38 (12) (2005) 2549–2563.

[23] J. Choi, G. Medioni, Y. Lin, L. Silva, O. Regina, M. Pamplona, T.C. Faltemier, 3d face reconstruction using a single or multiple views, in: Proceedings of the IEEE Conference on Pattern Recognition, 2010, pp. 3959–3962.

[24] Y. Lin, G. Medioni, J. Choi, Accurate 3d face reconstruction from weakly calibrated wide baseline images with profile contours, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1490–1497.

[25] P. Dou, I.A. Kakadiaris, Multi-view 3d face reconstruction with deep recurrent neural networks, in: Proceedings of the IEEE International Joint Conference on Biometrics, 2017, pp. 483–492.

[26] F. Wu, L. Bao, Y. Chen, Y. Ling, Y. Song, S. Li, K.N. Ngan, W. Liu, Mvf-net: multi-view 3d face morphable model regression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[27] J. Liang, F. Liu, H. Tu, Q. Zhao, A.K. Jain, On mugshot-based arbitrary view face recognition, in: Proceedings of the International Conference on Pattern Recognition, 2018, pp. 3126–3131.

[28] F. Wallhoff, S. Muller, G. Rigoll, Recognition of face profiles from the mugshot database using a hybrid connectionist/hmm approach, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 2001, pp. 1489–1492..

[29] L. Tran, X. Liu, Nonlinear 3d face morphable model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7346–7355.

[30] L. Tran, X. Liu, On learning 3d face morphable model from in-the-wild images, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019), In press.

[31] L. Tran, F. Liu, X. Liu, Towards high-fidelity nonlinear 3d face morphable model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1126–1135.

[32] X. Tu, J. Zhao, Z. Jiang, Y. Luo, M. Xie, Y. Zhao, L. He, Z. Ma, J. Feng, Joint 3d face reconstruction and dense face alignment from a single image with 2d-assisted self-supervised learning, ArXiv abs/1903.09359..

[33] F. Liu, R. Zhu, D. Zeng, Q. Zhao, X. Liu, Disentangling features in 3d face shapes for joint face reconstruction and recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5216–5225.

[34] F. Liu, Q. Zhao, X. Liu, D. Zeng, Joint face alignment and 3d face reconstruction with application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2020) 664–678.

[35] W.B. Lee, M.H. Lee, I.K. Park, Photorealistic 3d face modeling on a smartphone, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2011, pp. 163–168.

[36] D. Zeng, Q. Zhao, S. Long, J. Li, Examplar coherent 3d face reconstruction from forensic mugshot database, Image and Vision Computing 58 (2017) 193–203.

[37] D. Zeng, S. Long, J. Li, Q. Zhao, A novel approach to mugshot based arbitrary view face recognition, Journal of the Optical Society of Korea 20 (2) (2016) 239–244.

[38] L. Tran, X. Yin, X. Liu, Disentangled representation learning gan for pose-invariant face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1283–1292.

[39] L. Tran, X. Yin, X. Liu, Representation learning by rotating your faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2018) 3007–3021.

[40] R. Huang, S. Zhang, T. Li, R. He, Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2458–2467.

[41] J. Zhao, L. Xiong, J. Karlekar, J. Li, F. Zhao, Z. Wang, S. Pranata, S. Shen, S. Yan, J. Feng, Dual-agent gans for photorealistic and identity preserving profile face synthesis, in: Proceedings of the Annual Conference on Neural Information Processing Systems, 2017.

[42] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan, J. Feng, 3d-aided dual-agent gans for unconstrained face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2018) 2380–2394.

[43] J. Zhao, L. Xiong, Y. Cheng, Y. Cheng, J. Li, L. Zhou, Y. Xu, J. Karlekar, S. Pranata, S. Shen, J. Xing, S. Yan, J. Feng, 3d-aided deep pose-invariant face recognition, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2018.

[44] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, J. Karlekar, S. Pranata, S. Shen, J. Xing, S. Yan, J. Feng, Towards pose invariant face recognition in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2207–2216.

[45] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D face model for pose and illumination invariant face recognition, in: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2009, pp. 296–301.

[46] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: Proceedings of the European Workshop on Biometrics and Identity Management, 2008, pp. 47–56.

[47] R. Gross, I. Matthews, J. Cohn, T. Kanade, Multi-PIE, in: Proceedings of the IEEE Conference on Automatic Face & Gesture Recognition, 2008, pp. 1–8..

[48] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The feret evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

[49] URL:http://dlib.net/ (accessed: 2018-01-02)..

[50] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.

[51] X. Zhu, Z. Lei, X. Liu, H. Shi, S.Z. Li, Face alignment across large poses: a 3D solution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 146–155.

[52] F. Liu, D. Zeng, Q. Zhao, X. Liu, Joint face alignment and 3D face reconstruction, in: Proceedings of the European Conference on Computer Vision, 2016, pp. 545–560.

[53] P. Huber, P. Kopp, W. Christmas, M. Rätsch, J. Kittler, Real-time 3D face fitting and texture fusion on in-the-wild videos, IEEE Signal Processing Letters 24 (4) (2017) 437–441.

[54] A. Dessein, W.A.P. Smith, R.C. Wilson, E.R. Hancock, Seamless texture stitching on a 3D mesh by poisson blending in patches, in: Proceedings of the IEEE International Conference on Image Processing, 2015, pp. 2031–2035.

[55] A. Baumberg, Blending images for texturing 3D models, in: Proceedings of the British Machine Vision Association, 2002, pp. 404–413..

[56] J. Totz, A.J. Chung, G. Yang, Patient-specific texture blending on surfaces of arbitrary topology, in: Proceedings of the Workshop on Augmented environments for Medical Imaging and Computeraided Surgery, 2009, pp. 78–85..

[57] Z. Farbman, R. Fattal, D. Lischinski, Convolution pyramids, ACM Transactions on Graphics 30 (6) (2011) 1–8.

[58] A. Dessein, W.A.P. Smith, R.C. Wilson, E.R. Hancock, Symmetry-aware mesh segmentation into uniform overlapping patches, in: Proceedings of the Computer Graphics Forum, 2016.

[59] I. Kemelmacher-Shlizerman, R. Basri, 3D face reconstruction from a single image using a single reference face shape, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2) (2011) 394–405.

[60] F. Liu, D. Zeng, J. Li, Q. Zhao, On 3d face reconstruction via cascaded regression in shape space, Frontiers of Information Technology & Electronic Engineering 18 (12) (2017) 1978–1990.

[61] C. Qu, E. Monari, T. Schuchert, J. Beyerer, Fast robust and automatic 3D face model reconstruction from videos, in: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2014, pp. 113–118.

[62] A.T. Tran, T. Hassner, I. Masi, G. Medioni, Regressing robust and discriminative 3d morphable models with a very deep neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1493–1502.

[63] S. Romdhani, T. Vetter, Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 986–993..

[64] X. Zhu, D. Yi, Z. Lei, S.Z. Li, Robust 3D morphable model fitting by sparse sift flow, in: Proceedings of the IEEE Conference on Pattern Recognition, 2014, pp. 4044–4049.
[65] X. Wu, R. He, Z. Sun, T. Tan, A light cnn for deep face representation with noisy labels, IEEE Transactions on Information Forensics and Security 13 (2018) 2884–2896.
[66] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: Proceedings of the European Conference on Computer Vision, 2016, pp. 499–515.
[67] L. Hu, M. Kan, S. Shan, X. Song, X. Chen, LDF-Net: Learning a displacement field network for face recognition across pose, in: Proceedings of the IEEE Conference on Automatic Face & Gesture Recognition, 2017, pp. 9–16.
[68] H.J. Wang, Y. Wang, Z.-F. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, W. Liu, Cosface: Large margin cosine loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5265–5274.
[69] J. Zhao, J. Xing, L. Xiong, S. Yan, J. Feng, Recognizing profile faces by imagining frontal view, International Journal of Computer Vision 128 (2019) 460–478.
[70] X. Tu, J. Zhao, M. Xie, Z. Jiang, A. Balamurugan, Y. Luo, Y. Zhao, L. He, Z. Ma, J. Feng, 3D face reconstruction from a single image assisted by 2D face images in the wild, IEEE Transactions on Multimedia (2020).

Jie Liang received the M.Sc. degree in Computer Science from Sichuan University in 2019 and the B.Sc. degree in Computer Science from Shaanxi University of Technology in 2016. Her main research interests are computer vision, image processing and pattern recognition.

Huan Tu is currently a master student in the Biometrics Research Laboratory at Sichuan University. She obtained her B.Sc. degree in Electronic Commerce from Southwest University in 2017. Her main research interests are computer vision and pattern recognition.

Feng Liu is currently a post-doc researcher in the Computer Vision Lab at Michigan State University. He received the Ph.D. degree in Computer Science from Sichuan University in 2018. His main research interests focus on computer vision and pattern recognition, specifically for 3D modeling.

Qijun Zhao (Ph.D., 2010, The Hong Kong Polytechnic University; M.Sc., 2006, B.Sc., 2003, Shanghai Jiao Tong University; Post-Doc, 2010–2012, Michigan State University) is a professor in the College of Computer Science at Sichuan University. He is also a visiting professor in the School of Information Science and Technology at Tibet University. His research is in the fields of pattern recognition, image processing, and computer vision.

Anil K. Jain (Ph.D., 1973, Ohio State University; B. Tech., IIT Kanpur) is a University Distinguished Professor at Michigan State University where he conducts research in pattern recognition, machine learning, computer vision, and biometrics. He is a Fellow of ACM, IEEE, AAAS, and SPIE, and a Member of U.S. NAE.