Efficient 3D Face Recognition in Uncontrolled Environment

Yuqi Ding[1(✉)], Nianyi Li[1], S. Susan Young[2], and Jinwei Ye[1]

[1] Louisiana State University, Baton Rouge, LA 70803, USA
{yding18,nli5,jinweiye}@lsu.edu
[2] U.S. Army Research Laboratory, Adelphi, MD 20783, USA

Abstract. Face recognition in an uncontrolled environment is challenging as body movement and pose variation can result in missing facial features. In this paper, we tackle this problem by fusing multiple RGB-D images with varying poses. In particular, we develop an efficient pose fusion algorithm that frontalizes the faces and combines the multiple inputs. We then introduce a new 3D registration method based on the unified coordinate system (UCS) to compensate for pose and scale variations and normalize the probe and gallery face. To perform 3D face recognition, we train a Support Vector Machine (SVM) with both 2D color and 3D geometric features. Experimental results on a RGB-D dataset show that our method can achieve a high recognition rate and is robust in the presence of pose and expression variations.

Keywords: 3D face recognition · Uncontrolled environment · RGB-D images · Pose fusion · 3D Face registration

1 Introduction

Face recognition (FR) is of great importance as it has numerous applications in access control, surveillance system, and law enforcement. Although the past decade has witnessed tremendous advances in 2D face recognition [5,26], robust recognition in an uncontrolled environment is still challenging as the facial appearance in 2D images is sensitive to illumination, viewpoint, pose, and expression variations. In addition, body movement and/or head motion can cause large occlusions and result in missing facial features. It is urgent to overcome these challenges and provide successful solutions to benefit both civilian and military applications. A viable solution is to apply 3D face models for recognition [27]. As directly associated with the face geometry, a 3D face model is inherently invariant to scene properties (*e.g.*, illumination and viewpoint). However, a complete and accurate 3D face acquisition is often time-consuming and requires expensive devices [22]. Some approaches [2,32] recover 3D face models from one or multiple 2D images by fitting statistical models (*e.g.*, 3DMM [4]). Although no specialized hardware is needed, these methods are usually computationally intensive and the recovered face models are less accurate.

G. Bebis et al. (Eds.): ISVC 2019, LNCS 11844, pp. 430–443, 2019.
https://doi.org/10.1007/978-3-030-33720-9_33

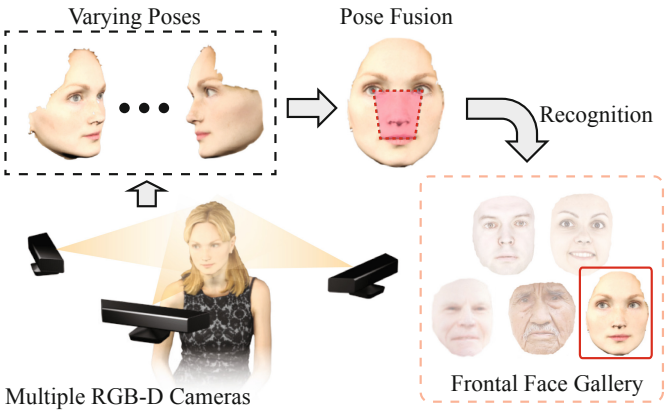

Fig. 1. Schematic illustration of our FR scheme in an uncontrolled environment.

In this paper, we use RGB-D images for 3D face recognition since RGB-D cameras are cheap and easily accessible. However, self-occlusions caused by pose variations result in missing facial features. Most RGB-D image-based methods utilize symmetric filling to complete the 3D face model [16,20,25]. Nevertheless, facial asymmetry renders the completion inaccurate. When multiple 3D scans are available, the Iterative Closest Point (ICP) algorithm [3] can be adopted to combine partial 3D face models. But in case of large pose changes where the partial models have small overlap, the ICP algorithm might fail to align them. The seminal work of KinectFusion [33] generated high-quality 3D reconstruction using a moving RGB-D camera. However, this method is not suitable for face recognition with stationary sensors.

To tackle the occlusion problem caused by pose variations, we propose a system consisting of multiple low-cost RGB-D cameras (*e.g.*, Microsoft Kinect) for 3D face recognition (see Fig. 1). The RGB-D cameras surround the subject to capture varying poses. Conceptually, our system can be deployed in various indoor environments (*e.g.*, a building interior and cave) and is able to handle uncontrolled conditions. To fuse face models with varying poses, we develop an efficient fusion algorithm that first frontalizes the faces and then merges the partial face models in a uniform grid. We then introduce a new 3D registration method based on the unified coordinate system (UCS) to compensate for pose and scale variations. We use the UCS to normalize the gallery (frontal face model) and probe (combined model). To perform 3D face recognition, we train a Support Vector Machine (SVM) using both 2D color and 3D depth features. Specifically, we extract 2D features from a normalized canonical color image using a convolutional neural network (CNN) based on a deep feature extractor. We then use the expression-invariant geodesic distances between facial landmarks that are computed on 3D facial meshes as 3D geometric features. We finally concatenate these 2D and 3D features to train the SVM for FR. The processing pipeline of our approach is shown in Fig. 2. Through experiments

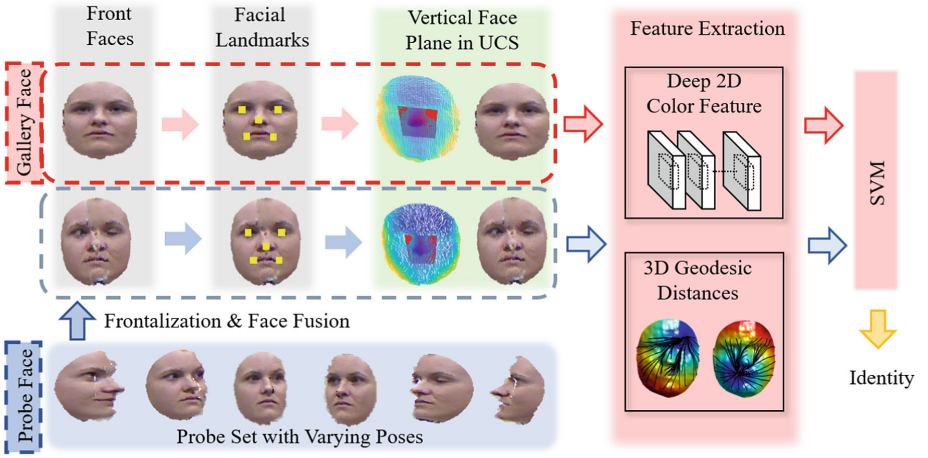

Fig. 2. Processing pipeline of our 3D FR algorithm.

on an RGB-D dataset [16] and comparisons with the state-of-the-art methods, we show that our approach has a high FR accuracy but relatively low computational complexity. More importantly, our approach is robust to pose and expression variations.

2 Related Work

A FR system usually consists of three basic modules [24]: (1) a face detection module for detecting the facial region [31]; (2) a feature detection and alignment module for data normalization [11,35]; and (3) a recognition module applied on the normalized faces [34]. As we present a new pose-invariant FR algorithm in this paper, we briefly review related studies on face registration and 3D FR.

Face Registration. According to [6,13], pose variation is a major factor that leads to reduced FR accuracy. Hence pose-invariant face recognition is of great importance. Face registration aims to align the faces of different poses to a canonical pose. In this way, pose variation is eliminated. Existing face registration methods can be classified into three major categories: (1) one to one registration; (2) all to mean face registration, and 3)registration by coordinate normalization. Given 3D face point clouds, the first-class registers the probe point cloud to each reference face in the gallery by an iterative procedure [23]. The Iterative Closest Point (ICP) algorithm is usually adopted for optimization. However, without good initialization of the parameters, the ICP may fail to converge. Instead of mapping the probe to each face in the gallery, the second class aligns all the face models to a mean face model learned from a training set [9,12]. Each face only needs one-time registration. Therefore, the computational cost is significantly decreased. However, these methods may suffer from large

registration errors. The third class aligns faces of different poses by normalizing the coordinates of detected facial landmarks [29]. A set of facial landmarks is first detected and then transformed into a common coordinate system. The resulting transformation is applied to the entire face point cloud for registration. This class of methods is more efficient than the other two in terms of computational cost. However, the accuracy still largely depends on the quality of detected landmarks, which might be missing in the presence of large pose variations. In this paper, we propose to first fuse faces with different poses and then use a unified coordinate system (UCS) to align the 3D face models. Our fused face is more robust to pose variations because it incorporates partial features from each pose to a complete set. Our UCS-based 3D face registration considers the depth information and is, therefore, more accurate than conventional 2D face registration.

3D Face Recognition. We refer readers to [21] for a comprehensive overview of 3D FR. Here, we focus on how 3D FR algorithms handle challenges in pose and expression variations as well as corrupted data. We classify existing 3D FR algorithms into three categories: (1) local descriptor-based; (2) global/model descriptor-based [1]; and (3) learning-based techniques [14,36]. The first category utilizes the characteristics of a small local neighborhood such as curvatures, shape index, and normals for matching. To list a few, Mian *et al.* [18] fused 3D keypoints with 2D Scale Invariant Feature Transform (SIFT) to identify 3D faces. Gupta *et al.* [10] matched the 3D Euclidean and geodesic distances between pairs of facial landmarks for 3D FR. Yet, these approaches are sensitive to facial expressions. The second category usually derives a 3D morphable face model and fits it to all probe faces. The best-matched 3D face is then used for recognition. For example, Gilani *et al.* [8] conducted FR by matching keypoint-based features on a statistical morphable model. However, these methods do not explicitly capture the actual 3D information, with a low-quality depth, they are less accurate, and do not work well on the RGB-D data. Although the third category has achieved great success in 2D FR, its application to 3D FR is still limited due to the absence of massive 3D face datasets. For instance, Kim *et al.* [14] fine-tuned the VGG-Face network on depth images to generate 3D feature descriptors and then used Principal Component Analysis (PCA) for feature matching. Gilani *et al.* [36] developed an end-to-end 3D FR framework by training an augmented 3D face dataset. Although these methods are successful, their datasets and network models are not publicly available. In this paper, we leverage a pre-trained deep neural network for 2D color feature extraction and then integrate expression-invariant 3D features for 3D FR.

3 Pose Fusion Using RGB-D Images

In this section, we describe our pose fusing algorithm using RGB-D images. The goal of the algorithm is to obtain a face model with a complete set of facial landmarks (*i.e.*, four eye corners, nose tip, and two mouth corners, as shown in Fig. 3(a)) by fusing RGB-D images with varying poses.

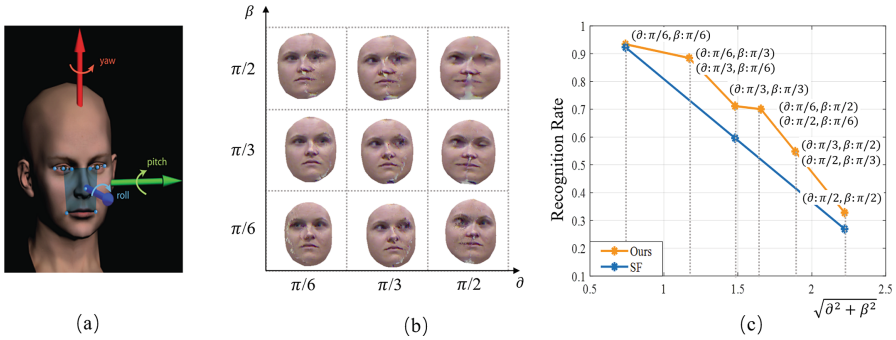

Fig. 3. (a) 3D rotations (yaw, pitch, and roll) with respect to our coordinate axes. (b) Pose fusion results that combine input poses with different rotation angles (α and β refer left and right yaw angles). (c) Recognition rates for using various pose fusion results in (b) as the probe. We compare with the completion results by symmetric filling (SF).

Given RGB-D images as input, we first pre-process the data to recover a point cloud using depth values and crop out the local face region by fitting a sphere centered at the nose tip. The radius of the sphere is determined according to the face scale. The face point cloud is further translated with the nose tip as the origin of coordinates. We then set out to combine the partial face point clouds with varying poses to restore the complete face model. A naïve approach is to directly apply the ICP algorithm to merge the partial point clouds. However, ICP fails easily in case of extreme poses with small overlaps. Instead, we apply 3D rotation to frontalize the faces and merge them into a uniform grid.

In an uncontrolled condition, the 3D face point cloud can exhibit three types of rotations: in-plane, pose, and tilt rotations, which are commonly referred to as roll, yaw, and pitch (see Fig. 3(a)). We use 3D rotation matrices to model these variations and revert the rotations to frontalize the face. To estimate the rotation parameters, we compare the nose region in our 3D face with a nose template from a standard frontal mean face. We compare intensities in the range images instead of the 3D models to reduce the computational cost.

Assuming the range images of our sampled nose region and the mean face nose template to be $N_R(i,j)$ and $N_T(i,j)$, $i,j \in \mathbb{N}$. We use a weighted normalized cross correlation (WNCC) Γ to assess the similarity between N_R and N_T:

$$\Gamma(N_R, N_T) = w \cdot \frac{\sum\limits_{i,j\in\mathbb{N}} \left(N_R(i,j) - \bar{N}_R\right) \cdot \left(N_T(i,j) - \bar{N}_T\right)}{\sqrt{\sum\limits_{i,j} \left(N_R - \bar{N}_R\right)^2} \cdot \sqrt{\sum\limits_{i,j} \left(N_T - \bar{N}_T\right)^2}} \tag{1}$$

where $\bar{N}_R$ and $\bar{N}_T$ are the mean values of N_R and N_T. The weight w is introduced to compensate for the missing data in the low-quality RGB-D images and is computed as the percentage of valid range data in the whole grid.

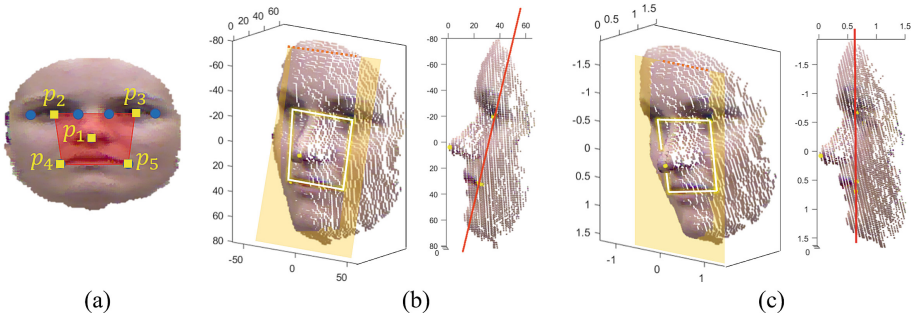

Fig. 4. UCS transformation and the face plane. (a) We illustrate the facial landmarks for constructing the UCS vertical face plane; (b) The initial face plane in the world coordinate system is not vertical due to pose variations; (c) The face plane becomes vertical after UCS transformation.

We then conduct a coarse-to-fine search to find the optimal rotation angle. Firstly, we use a relatively large search step, initialized as $\pi/60$, in the coarse search phase to maximize the Γ and shrink the search space to $\theta_c \pm \pi/30$, where θ_c corresponds to the coarse optimal rotation angle. Then, we implement a more precise search on the refined search space to further optimize Γ. We conduct this search algorithm by reducing the search step iteratively until Γ converges. To speed up our searching process, we employ a bilateral face ratio η that can effectively reduce the initial searching space: $\eta = num(I_l)/(num(I_l)+num(I_r))$, where $num(I_l)$ and $num(I_r)$ are facial pixel counts in the left and right sub-images. If $\eta > 0.5$, we search θ in $[0, \pi/2]$; otherwise, $\theta \in [-\pi/2, 0]$.

Finally, we merge the frontalized point clouds and their corresponding range images by box/grid averaging to generate a resampled face model, denoting the face pre-model. We apply a box filter to the face pre-model to further smooth out the noise. Figure 3(b) illustrates our pose fusion results by combining input poses with various rotation angles. We can see that our method is able to handle large pose variations ($\pm\pi/2$). We also compare our fusion method with symmetric filling in terms of recognition rates (see Fig. 3(c)). We can see that although recognition rates go down as rotation angle increases, our method still works better than the baseline symmetric filling algorithm in all scenarios.

4 Face Normalization Using Unified Coordinate System

In order to normalize the face models for recognition, we introduce a unified coordinate system (UCS) to compensate pose and scale variations. Our UCS is invariant to sensor position with respect to the human subject when the face image is taken (*i.e.*, invariant relative to the head pose).

Our UCS transformation can be viewed as a 3D registration (or 3D stereotactic registration) method. Our method transforms the 3D face pre-model into the UCS to generate a 3D registered face model. The coordinate transformation

contains 3D rotation and scaling where the rotation generalizes yaw, pitch, and roll; and scaling normalizes the face size. After UCS transformation, all face models are in a common space and could be fairly compared for recognition.

Now we describe how to construct the UCS. When a human face in 3D is viewed from the side, the depth of the eyes and the mouth are different, which is unique for each person. Thus, when a human head is in a normal position, which is normally termed as a frontal pose, the centers of corners of each eye and the corners of mouth are on two different vertical planes. So the goal of our UCS transformation is to compensate this depth variation such that these four points are on the same vertical plane and we call it the face plane (see Fig. 4). We can see that in order to have the face plane being vertical, the human head is tilted from the normal frontal pose.

To sum up, our UCS is defined with the following steps: (1) we identify five facial landmarks (*i.e.*, nose tip p_1, center of the left eye corners p_2, center of the right eye corners p_3, left mouth corner p_4, and right mouth corner p_5) using the color image; (2) we form a plane use $p_2 \sim p_5$ (if they do not share the same plane, we use the plane in which the distances of the four points are minimal) and estimate the plane's orientation in the world coordinate system; (3) we perform a coordinate transformation such that p_1 is the origin and the face plane is vertical; (4) we scale the coordinate by dividing by a scale factor s, where s is the distance from p_1 to line p_2p_3 such that the sizes of the face models are normalized. The new coordinate system is our UCS where the pose and scale of the face models are normalized. We apply the UCS transformation on both the gallery face and the probe face.

5 RGB-D Face Recognition

In this section, we present a Support Vector Machine (SVM)-based RGB-D face recognition algorithm to identify the probe face in the gallery set. We propose a color and geometric (CG) feature extractor to retrieve the 2D features from the color images and 3D geometric features from the 3D face meshes to describe the face model. Our SVM is trained with CG features from the gallery. Specifically, we leverage a VGG-Face [19] pre-trained convolutional network (CNN) to extract 2D feature vectors from the input color images. To exploit the depth information, we compute the geodesic distances between the facial landmarks on the 3D face meshes and use them as 3D features to allow expression-invariant recognition. By jointly considering the 2D color and 3D depth information, our FR algorithm on the RGB-D images is able to achieve a high recognition rate and is robust under an uncontrolled environment with pose and expression variations. Our FR algorithm is illustrated in Fig. 5. In the following paragraphs, we describe each component in detail.

2D Color Feature. Recent deep networks [19,28] have had great success in 2D face recognition, and the datasets with millions of face images are used to train a robust face classifier. Deep convolutional networks (CNNs) use a cascade of multiple layers of processing units for feature extraction and transformation [30].

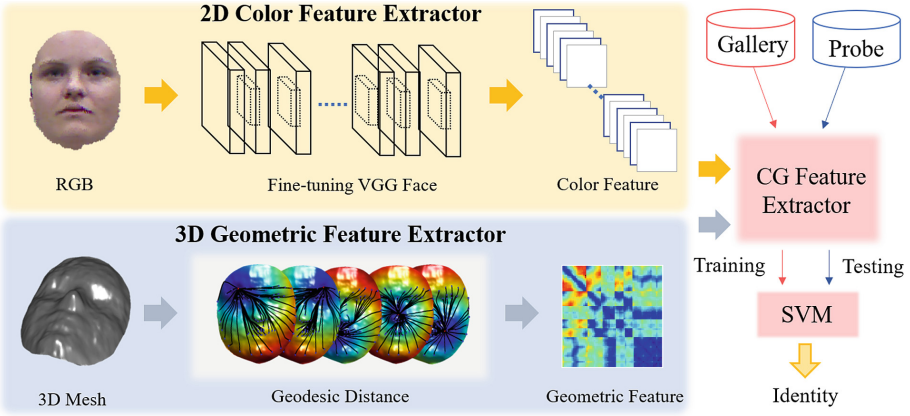

Fig. 5. Our SVM-based face recognition scheme using the Color and Geometric (CG) Feature Extractor.

According to [28], the bottom layers of the CNN typically extract the low-level features of an image, such as Gabor and SIFT. These low-level features can be extracted from an arbitrary generic natural image. In contrast, the outputs of the CNN top layers exhibit high-level characteristics that correspond to facial attributes such as poses and expressions and are thus critical to robust face recognition. Therefore, fine-tuning a pre-trained CNN on a new dataset is an efficient way to achieve high recognition accuracy but with less computational effort. VGG-Face [19] is a CNN-based 2D FR architecture trained on a dataset of 2.6M images and 2622 unique identities. It contains 22 layers, among which there are 13 convolutional layers, 5 max pooling layers, 3 fully connected layers, and a Softmax layer. The expected image resolution of the input layer is 224×224. We therefore take advantage of the VGG-Face to extract 2D color features from our input. To fit the input resolution of VGG-Face, our 2D color face image is resized to 224×224. We then transfer all the weights from VGG-Face to a CNN but remove all fully-connected layers and only keep the embedded $7 \times 7 \times 512$ features maps. We then flatten the feature maps to a 1D array and use it as our 2D color feature.

3D Geometric Feature. To exploit the depth information, we also extract 3D geometric features from our 3D face models. We choose the geodesic distances as our 3D geometric feature because the geodesic distance is the shortest distance between two points on a curved surface and is robust to expression changes [7,17]. To compute geodesic distances, we first construct triangular meshes using the face point cloud by Delaunay triangulation. We then use a smoothing filter to reduce the mesh noise. In our experiment, we implement the fast marching algorithm [15] to compute the geodesic distances on the face mesh from a source point. Facial landmarks are first detected on color images and then transferred to 3D meshes. By assigning k (in our paper, $k = 68$) facial landmarks on each

mesh, we generate a $k \times k$ matrix, where the (m, n) element indicates the geodesic distance from the m-th landmark to the n-th landmark. We resize this geodesic distance matrix into a 1D array and use it as our 3D geometric feature.

SVM Training. We finally concatenate the 2D color feature generated by pre-trained CNN and the 3D geometric feature computed with the geodesic distances and use it as our RGB-D face descriptor for classification. To perform face recognition, we train a support vector machine (SVM) as a classifier. Feature descriptors from the gallery set (frontal views) are used for training and the probe face by fusing multiple side views are used as testing input for recognition.

6 Experiments and Results

To validate our approach, we perform the experiment on the CurtinFace dataset [16] in which the RGB-D images are collected by Kinect. The dataset contains 52 subjects (42 males and 10 females), each has different poses, illuminations, and facial expressions. We use the frontal views as the gallery and side-view poses as probes. We use various poses of the same subject to emulate the images captured by our proposed multiple RGB-D camera system.

We compare our method with three state-of-the-art face frontalization algorithms: Li *et al.* [16], Hassner *et al.* [11], and Zhu *et al.* [35]. In order to prove the advantage of our CG feature extractor, we also compared it with a state-of-the-art 3D face recognition method, Emambakhsh *et al.* [7]. We perform both quantitative and qualitative comparisons on pose fusion and face recognition. All experiments are performed on a desktop PC with Intel Core i7-7700T CPU, 64 GB memory and two NVIDIA GeForce GTX 1060 6 GB GPUs. The varying pose fusion, UCS transformation, and geodesic distances are implemented through MATLAB R2018a. The SVM classifier and FR training are implemented with Sklearn, Tensorflow, and Keras.

Pose Fusion. We first demonstrate the effectiveness of our pose fusion algorithm. We perform a comparison with the three state-of-the-art face frontalization algorithms: Li *et al.* [16], Hassner *et al.* [11], and Zhu *et al.* [35]. Hassner *et al.* [11] and Zhu *et al.* [35] are designed for 2D face images. They first frontalize a 2D side-view face by first mapping it to a 3D frontal view average face model and then projecting the matched point cloud to xy-plane. Li *et al.* [16] took 3D face models as input and adopted the ICP algorithm to register the probe face to a mean face model. And they used the symmetric filling (SF) algorithm to obtain the complete face. We also compare it with our multiple pose fusion with SF (*i.e.*, we use our proposed method to frontalize the face and then flip the existing partial face to complete the entire face).

Our experiment is based on three sets of pose fusion: (1) $\pm\pi/2$, (2) $\pm\pi/3$, and (3) $\pm\pi/6$. The fusion results of all methods are shown in Fig. 6. We also use the fused faces of each method as input to our SVM-based face recognition algorithm to further validate the fusion quality. For the recognition task, our method achieves 32.69%, 70.05%, and 93.41% rank-1 accuracy, respectively, as

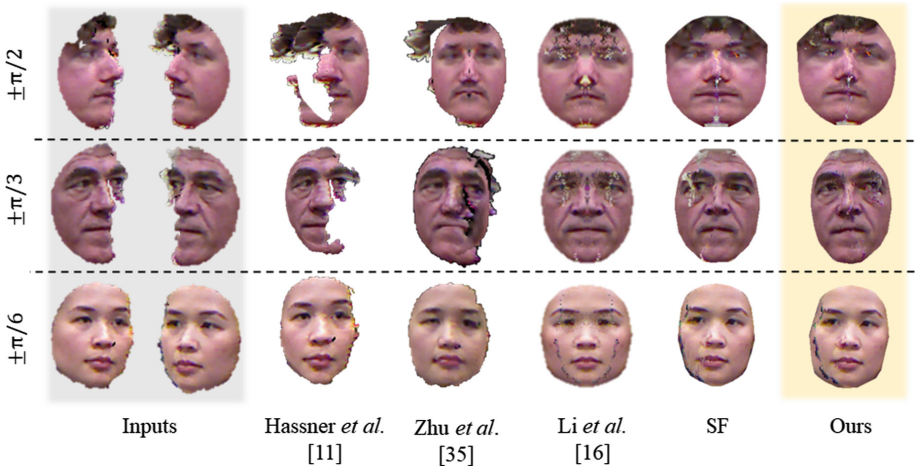

Fig. 6. Qualitative comparisons of the pose fusion results.

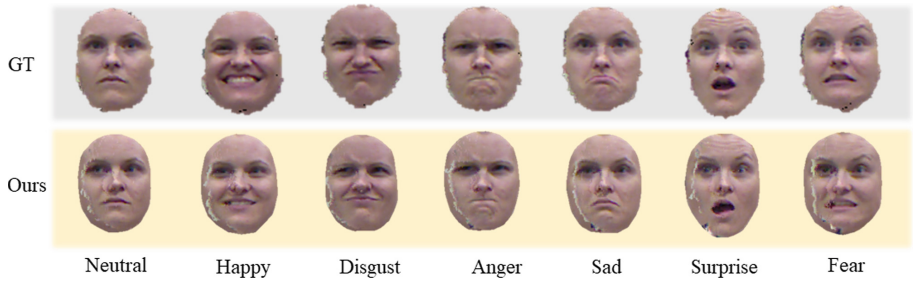

Fig. 7. Our fusion results for various expressions in comparison with the ground-truth (GT) frontal faces.

presented in Table 1. The rank-1 of 93.41% in the third pose set is the highest among all comparisons. It is important to note that 3D frontalization algorithms generally outperform 2D ones, due to the additional depth information and landmark detection failures on 3D faces by 2D approaches. Whereas our approach achieves the best performance especially in cases of large pose changes (for rotation angles larger than $\pi/4$), Li *et al.* [16] suffers from a low recognition rate due to the failure of ICP in presence of large pose variations. The SF results exhibit various artifacts (such as holes and duplicated regions) due to face asymmetry and misalignment. In contrast, our algorithm produces the most visually pleasing and accurate fusion results.

Face Recognition. Next, we show that our pose fusion algorithm and the CG feature representation benefit 3D face recognition. We test face recognition w.r.t. pose and expression variations. For pose variations, our experimental setup is the same as the pose fusion experiments.

Table 1. Recognition rate (%) w.r.t. Pose variations

Modality	Method	$\pm\pi/2$	$\pm\pi/3$	$\pm\pi/6$
2D	Hassner *et al.* [11]	7.69	23.07	50
	Zhu *et al.* [35]	7.69	15.38	46.15
	Ours (2D)	**32.69**	67.03	92.86
3D	Li *et al.* [16]	11.53	28.84	42.30
	SF	26.92	59.61	92.31
	Ours (2D+3D)	**32.69**	**70.05**	**93.41**

Table 2. Recognition rate (%) w.r.t. Expression variations

Modality	Method	Neutral	Happy	Disgust	Anger	Sad	Surprise	Fear
2D	Hassner *et al.* [11]	50.00	50.00	55.77	51.92	51.92	40.38	55.77
	Zhu *et al.* [35]	46.15	25.00	32.69	23.84	34.61	26.92	25.00
	Ours (2D)	**100**	88.46	**92.31**	86.54	**98.08**	**90.38**	94.23
3D	Emambakhsh *et al.* [7]	57.69	63.46	63.46	55.77	59.62	59.62	78.85
	Li *et al.* [16]	42.31	36.54	34.62	30.77	38.46	30.77	36.54
	SF	92.31	84.82	80.76	75.00	92.31	86.53	90.38
	Ours (2D+3D)	**100**	**92.21**	88.46	**88.46**	**98.08**	**90.38**	**96.15**

Recall that Hassner *et al.* [11] and Zhu *et al.* [35] are applied 2D face images as input. For fair comparison, we only use the 2D color features of the gallery faces for SVM training when comparing with these two methods. The recognition rate of all methods are presented in Table 1. We can see that our SVM-based face recognition scheme with 2D color and 3D geometric features achieves a high recognition rate when the input face is of high quality (*e.g.*, when the rotation angles are small). The overall FR performance is improved by taking 3D information into account. In presence of extremely large pose variation (*e.g.*, $\pm\pi/2$), the recognition rate is downgraded due to fusion errors. However, our method still achieves the highest rate in this extreme case.

For expression variations, we pick 7 expression categories: neutral, happy, disgust, anger, sad, surprise, and fear. Recognition is performed in each category with the gallery always being the neutral expression. For each expression category, we use poses of $\pm\pi/6$ for face fusion and then take the fused face as the probe. Our fusion results in comparison to the ground-truth (GT) frontal faces are shown in Fig. 7. To illustrate the robustness of our CG feature extractor with respect to expression variations, we also compare with Emambakhsh et al. [7], which is the state-of-the-art approach for handling various expressions. Emambakhsh et al. [7] use features in the nose region for recognition and is therefore robust to expression variations. However, due to the low-quality of our RGB-D data, high resolution features in the nose region are largely missing. Their FR performance is therefore downgraded. All aforementioned methods are tested in this experiment as well. The recognition rates are presented in Table 2. We can

see that our method achieves the highest recognition rate for all expressions and is thus robust to expression variations.

7 Conclusion

In this paper, we presented an efficient 3D face recognition algorithm that is robust in an uncontrolled environment by fusing multiple pose-varying RGB-D images. We first frontalized different poses and then fused them to obtain a front face model with a complete set of facial landmarks. To compensate for the pose and scale variations, we introduced the UCS transformation to normalize the gallery face and probe face. To perform face recognition, we extracted 2D color and 3D geometric features and used them to train a robust SVM classifier. Through experiments and comparisons with the state-of-the-art methods, we showed that our method can achieve the highest face recognition rate and is robust to pose and expression variations. Due to its robustness and efficiency, our technique can be implemented as a practical system for deployment in an uncontrolled environment.

References

1. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D face recognition: a survey. Pattern Recogn. Lett. **28**(14), 1885–1906 (2007)
2. Aldrian, O., Smith, W.A.P.: Inverse rendering of faces with a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1080–1093 (2013)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**, 239–256 (1992)
4. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Conference on Computer Graphics and Interactive Techniques. ACM Press (1999)
5. Corneanu, C.A., et al.: Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans. Pattern Anal. Mach. Intell. **38**(8), 1548–1568 (2016)
6. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. ACM Trans. Intell. Syst. Technol. (TIST) **7**(3), 37 (2016)
7. Emambakhsh, M., Evans, A.: Nasal patches and curves for expression-robust 3D face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(5), 995–1007 (2017)
8. Gilani, S.Z., Mian, A., Shafait, F., Reid, I.: Dense 3D face correspondence. IEEE Trans. Pattern Anal. Mach. Intell. **40**(7), 1584–1598 (2017)
9. Gökberk, B., et al.: 3D shape-based face representation and feature extraction for face recognition. Image Vis. Comput. **24**(8), 857–869 (2006)
10. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3D face recognition. Int. J. Comput. Vis. **90**(3), 331–349 (2010)
11. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: IEEE Conference on CVPR (2015)
12. Kakadiaris, I.A., et al.: Three-dimensional Face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 640–649 (2007)

13. Kan, M., Shan, S., Chen, X.: Multi-view deep network for cross-view classification. In: IEEE Conference on CVPR, pp. 4847–4855 (2016)
14. Kim, D., Hernandez, M., Choi, J., Medioni, G.: Deep 3D face identification. In: IEEE International Joint Conference on Biometrics (IJCB) (2017)
15. Kimmel, R., Sethian, J.A.: Computing geodesic paths on manifolds. Proc. Nat. Acad. Sci. **95**(15), 8431–8435 (1998)
16. Li, B.Y.L., et al.: Using kinect for face recognition under varying poses, expressions, illumination and disguise. In: IEEE Workshop on Applications of Computer Vision (WACV), pp. 186–192 (2013)
17. Li, X., Zhang, H.: Adapting geometric attributes for expression-invariant 3D face recognition. In: IEEE Conference on Shape Modeling and Applications (2007)
18. Mian, A.S., Bennamoun, M., Owens, R.: Keypoint detection and local feature matching for textured 3D face recognition. Int. J. Comput. Vis. **79**(1), 1–12 (2008)
19. Parkhi, O.M., Vedaldi, A., Zisserman, A., et al.: Deep face recognition. In: British Machine Vision Conference, vol. 1, p. 6 (2015)
20. Passalis, G., et al.: Using facial symmetry to handle pose variations in real-world 3D face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10), 1938–1951 (2011)
21. Patil, H., Kothari, A., Bhurchandi, K.: 3-D face recognition: features, databases, algorithms and challenges. Artif. Intell. Rev. **44**(3), 393–441 (2015)
22. Phillips, P.J., et al.: Overview of the face recognition grand challenge. In: IEEE Conference on CVPR, vol. 1, pp. 947–954 (2005)
23. Queirolo, C.C., Silva, L., Bellon, O.R., Segundo, M.P.: 3D face recognition using simulated annealing and the surface interpenetration measure. IEEE Trans. Pattern Anal. Mach. Intell. **32**(2), 206–219 (2009)
24. Ranjan, R., et al.: Deep learning for understanding faces: machines may be just as good, or better, than humans. IEEE Signal Process. Mag. **35**(1), 66–83 (2018)
25. Sang, G., Li, J., Zhao, Q.: Pose-invariant face recognition via RGB-D images. Comput. Intell. Neurosci. **2016**, 13 (2016)
26. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(6), 1113–1133 (2015)
27. Soltanpour, S., Boufama, B., Wu, Q.J.: A survey of local feature methods for 3D face recognition. Pattern Recogn. **72**, 391–406 (2017)
28. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: IEEE Conference on CVPR, pp. 1701–1708 (2014)
29. Tang, X., Chen, J., Moon, Y.S.: Accurate 3D face registration based on the symmetry plane analysis on nose regions. In: European Signal Processing Conference, pp. 1–5 (2008)
30. Wang, M., Deng, W.: Deep face recognition: a survey. arXiv preprint arXiv:1804.06655 (2018)
31. Wang, N., Gao, X., Tao, D., Yang, H., Li, X.: Facial feature point detection: a comprehensive survey. Neurocomputing **275**, 50–65 (2018)
32. Yi, D., Lei, Z., Li, S.Z.: Towards pose robust face recognition. In: IEEE Conference on CVPR (2013)
33. Yu, Y., Mora, K.A.F., Odobez, J.: HeadFusion: 360° head pose tracking combining 3D morphable model and 3D reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. **40**(11), 2653–2667 (2018)
34. Zafeiriou, S., Zhang, C., Zhang, Z.: A survey on face detection in the wild: past, present and future. Comput. Vis. Image Underst. **138**, 1–24 (2015)

35. Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z.: High-fidelity pose and expression normalization for face recognition in the wild. In: IEEE Conference on CVPR, pp. 787–796 (2015)
36. Zulqarnain Gilani, S., Mian, A.: Learning from millions of 3D scans for large-scale 3D face recognition. In: IEEE Conference on CVPR, pp. 1896–1905 (2018)