

Research Article

Pose-Invariant Face Recognition via RGB-D Images

Gaoli Sang, Jing Li, and Qijun Zhao

State Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University, Chengdu, Sichuan 610064, China

Correspondence should be addressed to Jing Li; lijing712@scu.edu.cn

Received 12 July 2015; Revised 4 October 2015; Accepted 7 October 2015

Academic Editor: Manuel Grana

Copyright © 2016 Gaoli Sang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three-dimensional (3D) face models can intrinsically handle large pose face recognition problem. In this paper, we propose a novel pose-invariant face recognition method via RGB-D images. By employing depth, our method is able to handle self-occlusion and deformation, both of which are challenging problems in two-dimensional (2D) face recognition. Texture images in the gallery can be rendered to the same view as the probe via depth. Meanwhile, depth is also used for similarity measure via frontalization and symmetric filling. Finally, both texture and depth contribute to the final identity estimation. Experiments on Bosphorus, CurtinFaces, Eurecom, and Kiwi databases demonstrate that the additional depth information has improved the performance of face recognition with large pose variations and under even more challenging conditions.

1. Introduction

Face recognition has been attracting considerable attention from researchers due to its wide variety of applications, such as homeland security, video surveillance, law enforcement, and identity management. The face recognition task can be performed nonintrusively, without user's awareness or explicit cooperation, thanks to the development of image sensing techniques. During the past decades, performance of face recognition is improving [1–5], but large pose variation problem still remains unsolved [6–8].

2D Based Method. Many researchers explored the large pose face recognition problem in recent years, and the popular 2D image based method achieved significant improvement. Deep learning based method has attracted researchers' attention because of its boosted performance. Zhu et al. [4] learned face identity preserving features by using a deep neural network to transform a nonfrontal face to frontal. In [5], a new deep neural network named multiview perceptron (MVP) was designed to learn features with respect to identity and view using millions of face images. The features learned from MVP achieve satisfactory performance on the Multi-PIE dataset. However, both methods required an extremely large dataset to train. Our method, on the other hand, attempts to solve large pose variation problem using additional depth

information. The complementary 3D data provides a different view to solve large pose variation problem.

Passive 3D Based Method. 3D face models can be achieved from 2D face images. Reference [9] reconstructed a 3D shape model from multiple 2D face images and generated a set of densely sampled 2D face images as templates for pose-invariant recognition. Reference [10] proposed a fully automatic system that matched the reconstructed faces under frontal view. It can handle continuous pose variation up to ± 45 in yaw and ± 30 in pitch angles. Yet face recognition under large pose variation is still challenging. One issue lies in the chicken and egg problem. On the one hand, accuracy of the reconstructed model depends on precise facial landmark localization. On the other hand, precise facial landmark still remains unsolved under large pose variations.

Active 3D Based Method. With the development of active 3D sensing techniques, it is possible to capture 3D face models without considering the pose. There are roughly two kinds of 3D sensors; one is depth cameras like Microsoft Kinect [11], affordable and fast, but is of low resolution, low precision, and low reliability. The other is 3D scanners like Minolta [12], producing high quality model but usually slow and expensive. Li et al. [11] utilized low-quality RGB-D data from a consumer level sensor to handle face recognition problem under various

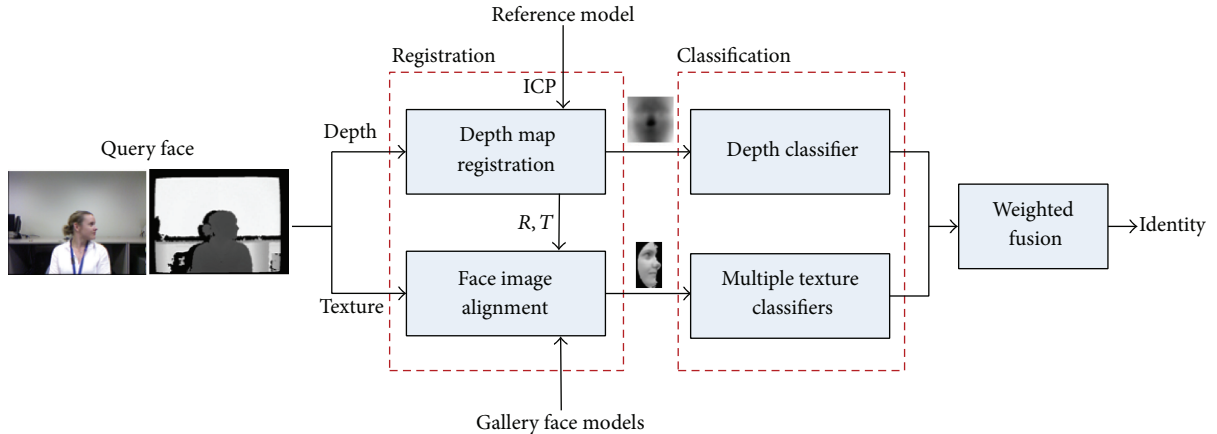


FIGURE 1: Framework of the proposed method. The input depth map D is first frontalized and registered to a predefined reference model. Its RGB counterpart I is then aligned to the database facing the same direction as the query. Finally, different classifiers are used for both texture and depth, and the final result is fused at score level. Here, ICP denotes the iterative closest points algorithm, and R and T are the rotation and translation matrixes estimated by ICP for registering the query face.

poses. Both texture image and depth maps were transformed to the frontal view for similarity calculation. However, symmetric filling on texture images involved in the frontalized face images often results in artifacts that may degrade the matching performance. Rather than frontalization on the 2D face image, [13] generated multiple face images under some predefined poses from 3D face models in the gallery, and the query face compared to all these projected images. Experiments show that recognition rate was higher than frontalization based method. Reference [14] proposed a resurfacing approach to tackle the challenge case with depth map corrupted by quantization noise, where the performance was highly dependent on precision of landmark localization and face alignment. Reference [15] used data acquired by high-quality 3D scanner and demonstrated that recognition rate is significantly higher by using both color and depth maps than that using depth alone. Reference [16] also combined color and depth for face recognition based on high-quality 3D face models captured using Inspeck Mega Capturor II 3D sensor. To deal with pose variation, the texture images are deformed according to rotation from depth maps and achieved 69.1% rank-1 recognition rate on Bosphorus database.

In this paper, we aim at solving large pose face recognition problem using RGB-D face images, by fusing both texture and depth into one framework via Joint Bayesian classifiers. Input of our algorithm is a RGB-D face image facing any direction (-90 degree to 90 degree), and output is identity of that person by similarity estimation between the input and that in the gallery.

The remainder of this paper is arranged as follows. The proposed method is shown in Section 2. Section 3 presents experimental results and relevant discussions. And concluding remarks are drawn in the end.

2. Proposed Method

Suppose $Q = \{I, D\}$ denotes the query face, where I is the input 2D face image from arbitrary view and D its

corresponding depth map, with the assumption that I and D are semantically registered. Suppose G is the gallery set, where each identity corresponds to a RGB-D pair of the frontal view. In order to locate the most similar face, our face recognition system (1) calculates similarity between Q and each sample in G , and (2) the final recognition result is obtained using some pretrained classifiers. For similarity calculation, we separately estimate similarity values from texture images and depth maps. Nevertheless, different sources of similarity estimation should undergo the same preprocessing step: to register the query face Q to each of the samples in G .

For depth map registration, we rotate the input profile face to the front via a 3D reference model P . And for texture image alignment, as its frontalized face is prone to artifacts, we on the other hand rotate the gallery to the input view. We then extract Histogram of Oriented Gradients (HOG) [18] features from the registered faces, which forms the basis for calculating similarity. For classification, we use different classifiers on texture and depth images separately. The final recognition result is obtained using some weighting scheme. Figure 1 shows the framework of our algorithm.

2.1. Query-Gallery Face Registration

2.1.1. Depth Map Registration. Given a query face $Q = \{I, D\}$ facing any direction, we first estimate the pose-invariant frontal representation of the depth map, which can further be used for face recognition.

Several methods [19, 20] on face frontalization are discussed. In this paper, the frontalization task is solved by registering each input face to a reference model. Motivated by the method in [11], we build the reference model using standard face models with neutral expression from the USF [21] database. The reference face model is constructed by aligning all the faces in USF, resampling them on a uniform grid, and then taking their mean. The complete reference face has 128×128 points.

Suppose $[R, T]$ represents rigid transformation between the query Q and reference model P , where R is the rotation matrix and T is translation vector. Thus, Q and P are registered via

$$P = RQ + T. \quad (1)$$

We use iterative closest point (ICP) [22] to solve (1). ICP is an accurate technique for registration. To avoid its sensitivity to initial condition in classical ICP, here we incorporate nose tip localization into ICP estimation. The nose tip positions are used for an initial and coarse matching between each query face and the reference face model. In this paper, we simply shift the original of the local coordinate system on the 3D face model to its nose tip position, so the two models can share the same coordinate system. At the end of the ICP registration, each query face and the reference model are registered. And we can obtain the pose normalized query face model and the corresponding rotation matrix R and transform vector T .

When the nose tip position is obtained, face cropping (both texture and depth) can also be easily done in 3D. The points that are more than 80 mm away from the nose tip are removed. Thus, the irrelevant data is abandoned. Note that the cropped faces are used in all our experiments.

After ICP registration, some data may be missing due to self-occlusion from nonfrontal query faces. Here, we estimate the values of the missing data through facial symmetry [11].

In order to reduce the noise and obtain a more reliable model, we smooth the point cloud (XYZ) using a publicly available code [23]. For each face, 128×128 points are normalized from its minimum X and Y to the maximum X and Y values.

2.1.2. Texture Image Alignment. Reference [13] enlarged the gallery by rendering face images under different predefined poses and compared the query face with all templates in the gallery. However, density of the sampled poses matters. Too sparse sampling intervals will not help recognizing arbitrary view face images, whereas too dense sampling intervals will lead to high storage requirement. In this paper, we rotate each of the gallery face to the same view as the query and thus avoid facing the above dilemma.

2D face images are rendered from gallery faces facing the same direction as the probe by using pairs of rotation matrix R and transform vectors T obtained via ICP.

In our method, one frontal RGB-D face for each subject is needed in the gallery. Each of the samples in G can be transformed into G' facing the same view as the probe by

$$G' = R^{-1}G - R^{-1} \cdot T, \quad (2)$$

where R^{-1} denotes the inverse of the rotation matrix R .

We transform all of the face models in G through (2) and render texture images with the same view as the query image using weak perspective projection [24]. Similarity between query face and rendered one is then estimated accordingly.

So, once the rotation matrix R and transform vectors T of a query are determined via ICP obtained, the rendered face images are automatically registered from their corresponding 3D counterparts.

2.2. Classification. We employ HOG for both texture image and depth map feature extraction. HOG is proven to be robust against geometric and has been used successfully in many applications such as object detection and recognition. In general, we consider that the appearance and shape information of an image can be well described by gradient or the direction of the edge density distribution. HOG is defined as the histogram of image gradients over a combination of positions, orientations, and scales. Moreover, in order to improve performance when using HOG, the locality histogram is made to normalize the contrast in the larger scope of image.

Compared with other classifiers such as Support Vector Machine, neural network, k -Nearest Neighbor, and Classification Trees, Joint Bayesian [25] is easier to train and does not require retraining when new persons are enrolled.

In our proposed framework, we employ Joint Bayesian algorithm for face recognition. Specifically, Joint Bayesian classification is applied on both texture and depth maps. We design to learn multiple pose-specific classifiers for texture images, in order to handle various poses. Consider a continuous pose set ranging $[-90, 90]$ degree, each classifier is trained by a fixed degree partitioning (i.e., an interval of 10 degree). As for depth maps, all the frontal view depth maps in the training set are used to train a depth classifier. The match scores of texture and depth are fused via weighted summation as follows:

$$S = \lambda_1 * S_{\text{depth}} + \lambda_2 * S_{\text{texture}}, \quad (3)$$

where S_{depth} and S_{texture} represent scores from depth maps and texture images, respectively. λ_1 and λ_2 are weighting parameters, which are determined based on the reliability of texture and depth. For instance, RGB-D images acquired by low-resolution sensors usually have less distinguishable depth maps than texture images, and hence the weight for depth score tends to be smaller than that for texture score. Finally, the query is assigned with the label of the class with highest similarity score.

3. Experiments

We test our method on four publicly available datasets, including (1) high-quality RGB-D data from Bosphorus database and (2) some more challenging datasets that contain low-quality depth capture and illumination changing along with various expressions (e.g., CurtinFaces) and occlusions (e.g., Eurecom and Kiwi). We compare our method with the prior-arts methods including PGM [16], PGDP [17, 26], DCT+SRC [11], COV+LBP [13], and RCRC [14]. We also evaluate in which way the 3D information can be fully explored and works better for 3D face recognition. Two experiments are designed; one is profile-to-profile depth comparison and the other is frontal-to-frontal mode. To further demonstrate robustness of our proposed method, we also investigate how the landmark location accuracy affects recognition rate of the proposed method.

Unless otherwise indicated, all the experiments HOG feature extraction was run on 9×9 patches and the gradient

orientations between $[0; 2\pi)$ are evenly divided into 18 bins. As the texture and depth classifiers are designed separately, the final score is obtained by linearly fusing them using manually set weighting parameters 0.6 and 0.4, respectively.

According to our experimental results, ICP still works when out-of-plane rotation is 90 degrees. But we have to carefully choose the distance tolerance when establishing closest point correspondences. In our experiments, we empirically set it to 3 millimeters.

3.1. Evaluation on Bosphorus Dataset. The Bosphorus database contains 4666 textured depth maps of 105 persons in various poses, expressions, and occlusion conditions. The faces in the Bosphorus database are acquired using Inspeck Mega Capturor II 3D sensor that captures high-precision face models with accuracy of $x = 0.3$ mm, $y = 0.3$ mm, and $z = 0.4$ mm. For each 3D model, a subset of 24 facial landmarks (contains nose tip) is provided in the database. For each subject, there are various head poses including seven yaw angles, four pitch angles, and two cross rotations which incorporate both yaw and pitch. In order to ensure efficient and effective performance of the proposed method, a preprocessing step that downsample the high-precision face data is needed in our experiment.

We compare the performance of the proposed method with a set of approaches. PGM [16] transformed the texture map of a face model through its shape data via the Patch Geodesic Distance into a transformed texture map to tackle the missing data problem due to pose and expression changes. Both texture and spatial distance contribute to the final similarity measure between the query and the gallery face models. To handle the missing data problem caused by pose and expression changes, PGDP [17] transformed the texture map with pose variations through Patch Geodesic Distances into the view in-variance texture. Then shape adjusted texture patches are encoded into local patterns for face recognition. In our method, similarity measure on texture images is calculated among pairs of images that face the same direction as the query face. For each subject, one frontal neutral face model is used as the gallery set. All the frontal view face models for each subject are used to train 13 texture (including seven yaw angles, four pitch angles, and two cross rotations) and depth Joint Bayesian classifiers. 13 images with yaw, pitch, and cross rotation incorporate both yaw and pitch for each subject are used as the probe set. Figure 2 shows the preprocessing results of some examples of a subject in the probe set. In this experiment, we set that λ_1 equals λ_2 (0.5) under the observation that depth maps are as distinguishable as texture images.

Table 1 summarizes the recognition rates on different methods with respect to different rotation angles on the Bosphorus database. Obviously, the proposed method is much better than other two methods. The proposed method achieves a 98.7% recognition rate for the $+45^\circ$ rotation to the right, while PGM and PGDP have the recognition rate of 39% and 36.2%. For the -45° rotation to the left, the recognition rate for the proposed method is 98.6% and PGM and PGDP are 38.1% and 37.1%, respectively. Again, for the right-downwards rotation, the recognition rate of the

TABLE 1: Recognition rates for different pose rotations on the Bosphorus database.

Pose	PGM [16]	PGDP [17]	Proposed method		
			Depth	Texture	Fusion
$+10^\circ$ right	92.3%	96.1%	90.0%	100%	99.9%
$+20^\circ$ right	88.6%	91.4%	89.5%	99.7%	99.9%
$+30^\circ$ right	80.0%	80.1%	89.3%	99.6%	99.8%
$+45^\circ$ right	39.0%	36.2%	88.8%	97.7%	98.7%
-45° left	38.1%	37.1%	88.6%	98.1%	98.6%
$+90^\circ$ right	N/A	N/A	84.2%	91.4%	95.1%
-90° left	N/A	N/A	84.5%	92.3%	95.7%
Upwards (slight)	87.6%	96.2%	88.9%	93.4%	97.3%
Upwards (strong)	79.0%	91.4%	87.1%	92.0%	96.6%
Downwards (slight)	87.6%	86.7%	87.1%	93.0%	97.8%
Downwards (strong)	69.5%	70.0%	85.4%	92.7%	94.8%
Right-upwards	63.8%	64.7%	85.6%	91.2%	95.6%
Right-downwards	34.3%	35.2%	84.3%	89.0%	92.5%
Average	69.1%	71.4%	87.1%	93.4%	96.7%

N/A denotes that the measure is not available in the paper where the method was proposed (the same for the other tables in this paper).

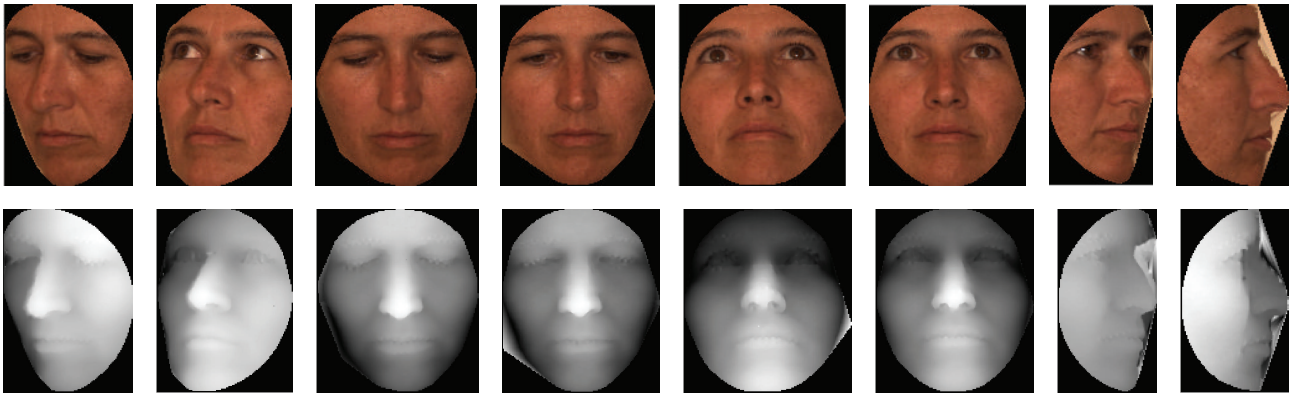
proposed method still outperforms PGM and PGDP by 58.2% and 57.3%, showing its consistent superior performance in accuracy. Moreover, For the $+90^\circ$, -90° rotation, the proposed method achieves as much as 95.1% and 95.7% recognition rate, while PGM and PGDP have not reported.

The algorithm has been compared with other state-of-the-art methods testing on this database. Reference [26] transformed the query 3D face from many different poses to frontal pose by using the Hausdroff Distance metric and then extracted the corresponding normal values for face recognition. Reference [26] achieved the average rank-1 recognition rate that is 66%. It should be noted that [26] is based on 3D surface information, while the proposed method achieves an average accuracy of 87.1% on depth maps, outperforming the rank-1 recognition rate of [26] by 21.1%.

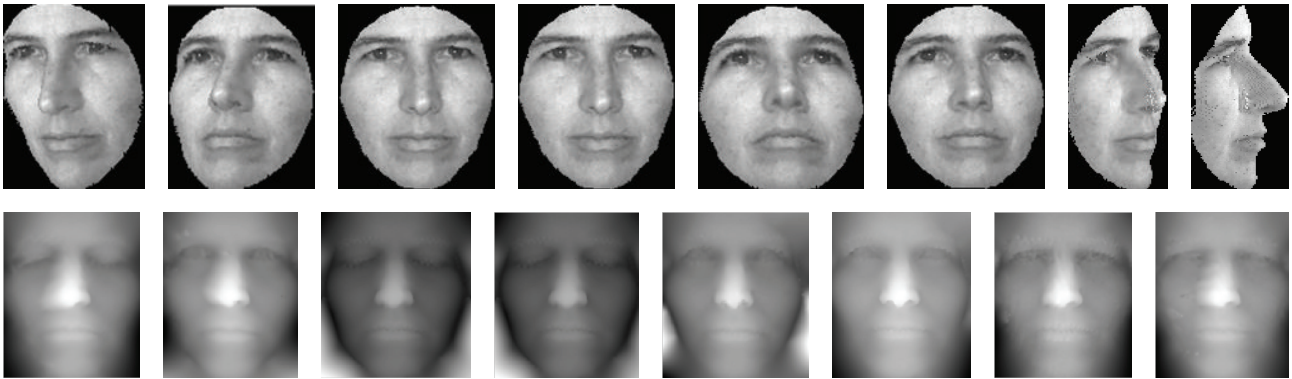
3.2. Evaluation on More Challenging Datasets. In this section, we evaluate our method using challenging data varying in quality, pose, expression, illumination, and occlusion conditions.

3.2.1. Expression Variations on CurtinFaces Dataset. The CurtinFaces dataset contains more than 5000 images of 52 subjects with variations in poses, expressions, illumination, and sunglasses disguise acquired by Microsoft Kinect. Thus the face models in CurtinFaces are of low quality, and recognition using these data is challenging.

For a fair comparison with [11, 13, 14], we use the same subset that contains 7 poses ($L90^\circ, L60^\circ, L30^\circ, 0^\circ, R30^\circ, R60^\circ, R90^\circ$) and 7 expressions (neutral, happy, disgust, anger, sad, surprise, and fear) per pose in the CurtinFaces database. In our experiment, 12 pairs of frontal views with various expressions and illuminations of each subject (see Figure 3) are used to train 9 texture classifiers



(a) The figure in the first line shows the probe texture images and the second line shows the corresponding depth maps



(b) The figure in the first line shows texture images at the same view as the probe faces rendered by face models in the gallery; the second line shows probe depth maps after frontalization

FIGURE 2: Samples of test images of one subject with poses variations.



FIGURE 3: Samples of a subject used for training.

(at about 0° , $\pm 30^\circ$, $\pm 60^\circ$, $\pm 90^\circ$ yaw rotation, and $\pm 60^\circ$ pitch rotation) and 1 depth classifier. 45 images variations in poses and expressions remaining per subject are used for testing. We compare the proposed method with the method in [13] which matches the texture maps with similar pose of the query face and the method in [11] matches the texture and the depth map with the gallery faces in a frontal view. For our method, one pair of RGB-D face under frontal view is used as the gallery for each subject. The recognition rate is reported in Table 2 for different poses under different expressions.

As shown in Table 2, we report the results achieved by texture, depth map, and fusion of texture and depth map using the proposed method. It is obvious that fusion of texture and depth information performs better than only using texture or depth map. Our method performs better with a certain degree of improvement, especially when the query faces at profile view, the recognition rates of texture and depth improve significantly compared with previous works. Exactly, compared with existing methods, the proposed method achieves 1.1% improvement than the second best method [14]

TABLE 2: Recognition rates of different methods on the pose-expression subset of the CurtinFaces database.

Pose	COV+LBP [13]	DCT+SRC [11]			RSRC [14]	Proposed method		
		Depth	Texture	Fusion		Depth	Texture	Fusion
Frontal	N/A	100%	100%	100%	100%	100%	100%	100%
$\pm 30^\circ$ yaw	94.2%	88.3%	99.8%	99.4%	99.4%	88.5%	99.5%	99.5%
$\pm 60^\circ$ yaw	84.6%	87.0%	97.4%	98.2%	98.2%	87.7%	97.5%	98.4%
$\pm 90^\circ$ yaw	75.0%	74.0%	83.7%	84.6%	93.5%	78.5%	94.0%	95.1%
$\pm 60^\circ$ pitch	N/A	81.6%	89.1%	92.8%	N/A	82.5%	94.3%	96.7%

For COV+LBP, a subset of pose-variation of the CurtinFaces is used.

For RSRC, we report the average of the best results at the same yaw pose obtained by the method.



FIGURE 4: An example of a subject from Eurecom database with each nose tip is labelled. The rows from left to right are neutral, left profile and right profile, light on, and occlusion eyes; the bottom rows are occlusion mouth, occlusion paper, smile, and open mouth.

at the profile view. Compared to the method in [13], the proposed method yields about 20.1% improvements at the profile view. No obvious drop in performance is observed across vast pose variations, even under the challenging case of profile view. Thanks to the ICP-based pose estimation method, our method projects gallery faces to the same view as the query, thus producing more accurate similarity measure. Moreover, our method aligns face images through 3D point clouds rather than several facial landmarks [14].

Furthermore, it can be observed from Tables 1 and 2 that the recognition rates for the depth only under the same view achieved by the Bosphorus database are higher than the results achieved under the corresponding head rotations by low quality data in the CurtinFaces dataset; the reason for this phenomena is that (1) in terms of depth precision high precision models provide more robust and accurate projection for texture rendering, thus producing more realistic rendered view compared with its low precision counterparts; (2) as in this paper, we did not handle nonrigid deformations (i.e., expression in CurtinFaces), and these data are treated as additional noise in our case; thus its recognition rate drops compare with that from noise-free data (i.e., Bosphorus).

3.2.2. Occlusion Variations on Eurecom and Kiwi Databases.

The Eurecom database contains 936 images of 52 subjects and the images are captured with variations in pose, illumination, and occlusion by Microsoft Kinect. Figure 4 shows an example of 9 images of a subject in the Eurecom database. In this experiment, we use frontal faces except Occlusion Mouth and Occlusion Paper of each subject in Session 1 as training

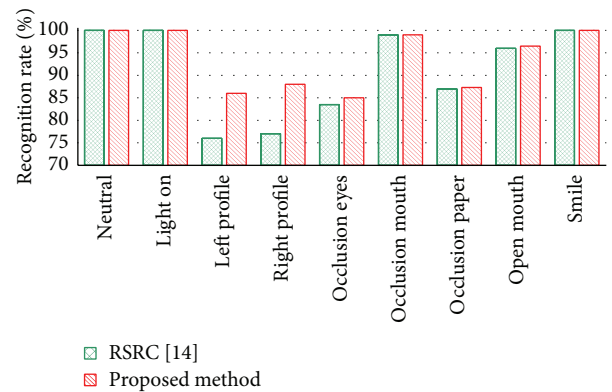


FIGURE 5: Performance comparison of the method RSRC by Hsu et al. and the proposed method on the Eurecom database with 9 conditions.

set to train three texture (frontal, left profile, and right profile) and one depth Joint Bayesian classifiers. The rest four images (left profile, right profile, occlusion mouth, occlusion paper) along with faces in Session 2 as the probe set.

The recognition performance of the proposed method and the benchmark method are shown in Figure 5 under different conditions. By comparing the results of RSRC [14] and the proposed method with different conditions, it can be seen that the recognition rate of left and right profiles are significantly better than the benchmark method.

The Kiwi Kinect Head Pose database (abbreviated as Kiwi) contains over 15k images of 20 people (6 females and 14

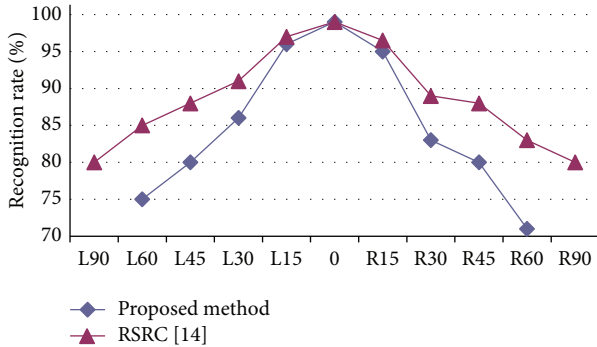


FIGURE 6: Performance comparison of different methods on the Kiwi database with different yaw angles ranging from -90 (L90) to $+90$ (R90) degrees.

males, 4 subjects recorded twice by with and without glasses) turning their head and captured at around 1 meter away from the sensor (Microsoft Kinect). The head pose range covers about $\pm 90^\circ$ yaw, $\pm 60^\circ$ pitch, and $\pm 50^\circ$ roll rotations. For a fair comparison with the methods RSRC [14], we use 10 pairs of frontal RGB-D faces of each subject to train 13 texture classifiers in which pose range covers about $\pm 90^\circ$ yaw rotations with an interval of 15 degree and one depth classifier.

The recognition performance of our method and RSRC are shown in Figure 6 with different yaw angles. The performances of the proposed method and RSRC are similar when the yaw angle $\leq 15^\circ$, although the former appears slightly better, while the recognition rate of RSRC degrades substantially when the yaw angle $> 15^\circ$ because the performance of RSRC depends heavily on accuracy of facial landmark localization. However, facial landmark location is nontrivial for large pose face images due to self-occlusion.

3.3. Evaluation on Different Depth Comparison Mechanisms.

In this part, we design experiments to explore the appropriate approach of 3D information that works better. We evaluate two different schemes for depth comparison. One approach is to rotate query face to the frontal view and then match to the gallery faces in a frontal view (denoted as frontal-to-frontal), while the other approach is to rotate each gallery face to the same pose as the query face; then the query face matches gallery face at nonfrontal view (denoted as profile-to-profile). For this comparison, we use a subset that contains only yaw pose variations of CurtinFaces database to evaluate the two approaches. One frontal face for each subject is used as the gallery set. Frontal faces with expression (seven different expressions) and illuminations (five different illuminations) of each subject are used as training set.

Interestingly, as shown in Table 3, the recognition rate achieved at frontal view is significantly better than nonfrontal view, especially under large pose variations. This is probably because the profile depth maps that are rotated from low-resolution frontal ones are seriously distorted. Furthermore, some information is lost due to self-occlusion when rotating. However, this is not the case for texture images, where profile-based comparison performs better, as the transformation

TABLE 3: Recognition rates when two different schemes are used for head rotations of depth map.

Pose	Depth map	
	Frontal-to-frontal	Profile-to-profile
$\pm 30^\circ$ yaw	89.4%	86.5%
$\pm 60^\circ$ yaw	87.5%	72.1%
$\pm 90^\circ$ yaw	78.8%	57.6%

from profile to frontal may produce artifacts, that is, illumination change.

3.4. Robustness on Nose Tip Localization. The proposed method required the nose tip for face cropping and rough alignment. In the above experiments, we suppose the nose tip is provided by database or marked manually. While in real-world applications, when using automatically algorithms to detect nose tip, there will be inevitably some errors in detecting nose tip locations. In order to evaluate the impact of nose tip location on face recognition accuracy, we perform proposed method using noisy location of nose tip on the CurtinFaces database. Specifically, we disturb the ground truth nose tip by randomly selecting the nose tip within a radius of 3, 5, 7, and 9 mm of the ground truth location. We conduct two series of experiments on the subset that contains only yaw pose variations of CurtinFaces database: (1) training and testing both using nose tip provided by database or manually marked (denoted as groundtruth data); (2) training using nose tip provided by database or manually marked and testing using data with noisy nose tip location (denoted as noisy data). Table 4 presents the recognition rates of the proposed method testing with ground truth and noisy nose tip location.

As shown by Table 4, our proposed method is not sensitive to the noisy nose tip location. The larger the radius we test is, the stronger the noise is. As can be seen, the recognition rate is comparably the same as intensity of the noise growth. When the radius is as large as 9 mm, the recognition rates only decline about 0.9%. The results prove that our proposed method can perform well even when the nose tip location is not as accurate.

3.5. Time Complexity. Table 5 lists the average training and test time of our proposed method using a 64-bit MATLAB implementation on a PC with Intel i3 CPU and 4 GB memory. No extra effort was made for code optimization. In the training phase, most of the time is due to the 2D face rendering process. In other words, rendering 2D face images from each of the gallery faces to the same direction as the probe is the most time-consuming step in the proposed method.

In terms of computational complexity, the proposed method takes more time than the DCT+SRC method to recognize a single query face mainly due to 2D face rendering process. On average, the computational complexity of our algorithm is 9 s for recognizing a single query face from a database of 52 persons, while the time consumption on

TABLE 4: Recognition rates testing with groundtruth and noisy nose tip location.

Pose	Groundtruth data	Noisy data			
		R = 3 mm	R = 5 mm	R = 7 mm	R = 9 mm
$\pm 30^\circ$ yaw	99.1%	98.1%	99.1%	99.1%	98.1%
$\pm 60^\circ$ yaw	98.1%	98.1%	98.2%	97.2%	97.2%
$\pm 90^\circ$ yaw	95.2%	95.2%	95.2%	94.2%	94.2%

TABLE 5: Time performance comparison of the proposed and DCT+SRC methods on the CurtinFaces database (average time in seconds).

	Training (ours)	Testing	
		Single query [11]	Single query (ours)
Face cropping	89	0.061	0.127
ICP registration	3230	3.467	3.459
Symmetric filling	917	0.989	0.941
Resampling	174	0.477	0.002
2D face rendering	37440	—	4.024
Classification (depth)	13	0.026	0.036
Classification (texture)	15	0.084	0.057
Fusion	—	0.017	0.024
Total	44562	5.114	8.748

verification is only about 4.5 s. If the proposed method could be implemented using C/C++ with code optimization and parallel computation, we expect an improvement over its efficiency.

4. Conclusions

In this paper, we presented a pose-invariant face recognition method using depth and texture. We evaluated our proposed method on four public available face databases, and the results demonstrated that the proposed method can not only handle large pose variation problem, but also achieve promising results under challenge conditions like expression and occlusion.

A limit of the proposed method is its high computational cost when identifying a person in a large gallery. In this sense, it is more applicable for face verification (i.e., one-to-one matching). In our future work, we are going to investigate the possibility of speeding it up via parallel computing. In addition, we are also going to exploit other feature representation methods to further improve the proposed method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

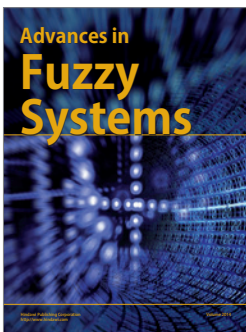
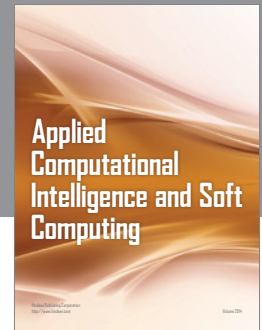
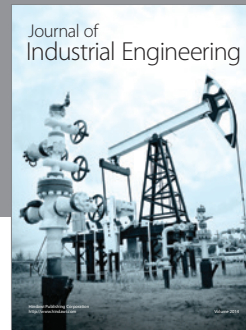
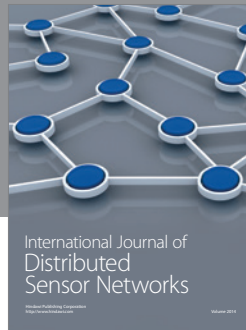
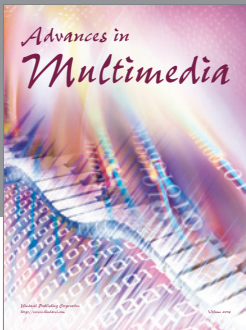
Acknowledgment

This work is supported by the National Key Scientific Instrument and Equipment Development Project of China (no. 2013YQ49087903).

References

- [1] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [3] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [4] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '13)*, pp. 113–120, IEEE, Sydney, Australia, December 2013.
- [5] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning multi-view representation for face recognition," Technical Report, 2014, <http://arxiv.org/abs/1406.6947>.
- [6] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques," *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41–68, 2009.
- [7] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.
- [8] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: a survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.
- [9] M. Ishimoto and Y.-W. Chen, "Pose-robust face recognition based on 3D shape reconstruction," in *Proceedings of the 5th International Conference on Natural Computation (ICNC '09)*, pp. 40–43, Tianjin, China, August 2009.

- [10] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. V. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 937–944, Barcelona, Spain, November 2011.
- [11] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '13)*, pp. 186–192, IEEE, Tampa, Fla, USA, January 2013.
- [12] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D face recognition database," in *Proceedings of the IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI '10)*, pp. 97–100, IEEE, Austin, Tex, USA, May 2010.
- [13] C. Ciaccio, L. Wen, and G. Guo, "Face recognition robust to head pose changes based on the RGB-D sensor," in *Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '13)*, pp. 1–6, Arlington, Va, USA, October 2013.
- [14] G.-S. J. Hsu, Y.-L. Liu, H.-C. Peng, and P.-X. Wu, "RGB-D-based face reconstruction and recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2110–2118, 2014.
- [15] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multimodal 2d and 3d biometrics for face recognition," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG '03)*, pp. 187–194, IEEE, Nice, France, October 2003.
- [16] F. Hajati, A. A. Raie, and Y. Gao, "2.5 D face recognition using patch geodesic moments," *Pattern Recognition*, vol. 45, no. 3, pp. 969–982, 2012.
- [17] S. Gheisari, S. Javadi, and A. Kashaninya, "3D face recognition using patch geodesic derivative pattern," *International Journal of Smart Electrical Engineering*, vol. 2, no. 3, pp. 127–132, 2013.
- [18] O. L. Junior, D. Delgado, V. Gonçalves, and U. Nunes, "Trainable classifier-fusion schemes: an application to pedestrian detection," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems (ITSC '09)*, pp. 1–6, IEEE, St. Louis, Mo, USA, October 2009.
- [19] B. Raducanu and F. Dornaika, "Pose-invariant face recognition in videos for human-machine interaction," in *Computer Vision—ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part II*, vol. 7584 of *Lecture Notes in Computer Science*, pp. 566–575, Springer, Berlin, Germany, 2012.
- [20] T. Shan, C. B. Lovell, and S. Chen, "Face recognition robust to head pose from one sample image," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, pp. 515–518, Hong Kong, China, August 2006.
- [21] USF DARPA human-Id 3D Face Database, <http://www.lib.usf.edu/hgsc/collections/databases/>.
- [22] ICP2, <http://staffhome.ecm.uwa.edu.au/~00053650/code.html>.
- [23] <http://mathworks.com/matlabcentral/fileexchange/8998>.
- [24] A. M. Bruckstein, R. J. Holt, T. S. Huang, and A. N. Netravali, "Optimum fiducials under weak perspective projection," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 223–244, 1999.
- [25] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: a joint formulation," in *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part III*, vol. 7574 of *Lecture Notes in Computer Science*, pp. 566–579, Springer, Berlin, Germany, 2012.
- [26] P. Bagchi, D. Bhattacharjee, and M. Nasipuri, "3d face recognition across pose extremities," in *Advanced Computing, Networking and Informatics—Volume 1*, vol. 27 of *Smart Innovation, Systems and Technologies*, pp. 291–299, Springer, Berlin, Germany, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

