Rendering or normalization? An analysis of the 3D-aided pose-invariant face recognition

Yuhang Wu, Shishir K. Shah, Ioannis A. Kakadiaris
Computational Biomedicine Lab
Department of Computer Science, University of Houston, Houston, TX, USA
ywu36@uh.edu {sshah,ikakadia}@central.uh.edu

Abstract

In spite of recent progress achieved in near-frontal face recognition, the problem of pose variations prevalent in 2D facial images captured in the wild still remains a challenging and unsolved issue. Among existing approaches of pose-invariant face recognition, 3D-aided methods have been demonstrated effective and promising. In this paper, we present an extensive evaluation of two widely adopted frameworks of 3D-aided face recognition in order to compare the state-of-the-art, identify remaining issues, and offer suggestions for future research. Specifically, we compare the pose normalization and the pose synthesis (rendering) based methods in an empirical manner. The database (UHDB31) that we use covers 21 well-controlled pose variations, half of which show a combination of yaw and pitch. Through the experiments, we present the advantages and disadvantages of these two methods to provide solid data for future research in 3D-aided pose-invariant face recognition.

1. Introduction

Recently, several methods have reached an exceptionally high face verification rate (99.63 %) [22] on the 'Labeled Faces in the Wild' database [14]. Even though significant progress has been made in this near-frontal face recognition challenge with deep neural networks and massive training data, pose-invariant face recognition required by real-world applications remains largely unsolved [7]. Fundamental research is still critical and essential for addressing other more challenging and demanding tasks (*e.g.*, unconstrained video-based face verification and face recognition in a surveillance stream). In this paper, we present an analysis of 3D-aided face recognition, one of the powerful strategies for pose-invariant face recognition [7].

In 3D-aided face recognition, a 3D facial model is exploited as prior information to assist recognition. The method is intrinsically designed to minimize the intraclass variations through unsupervised geometric transformation before feature extraction. This differs from the supervised methods (*e.g.*, metric learning or subspace analysis) that search for an implicit discriminative space after feature extraction. Compared to the 2D-based methods, the 3D-aided approach has the potential to perform better in unrestricted scenarios because the geometric shape transformation is more robust to the artifacts attributed to illumination and occlusion. Meanwhile, it preserves the local consistency in the transformed space and simultaneously suppresses the intraclass variations.

Most of the methods in this domain [25, 5, 12, 11, 13, 23, 19] fall roughly into two categories based on how they exploit 3D information: normalization-based methods [25, 5, 13, 23, 11] and rendering-based methods [12, 19]. The overview of these methods can be viewed in Figs. 1 and 2. In the normalization-based methods, a 3D model is used to normalize all the faces in the gallery and probe to a frontal pose. In rendering-based methods, instead of normalization, with the help of the 3D model, the facial images in the gallery are rotated to approximate the pose of the probe image. We compare normalization-based methods and rendering-based methods through experiments in a comprehensive facial database (UHDB31) and provide informative suggestions for future work on 3D-aided face recognition.

2. Related work

Pose alignment exploits geometric knowledge of a human face to transform either the gallery or probe, or both, to achieve correspondence. There are a few 2D-based methods for pose-alignment. Liu *et al*. [17] proposed an algorithm that aligns two images by searching for a dense scene correspondence through SIFT-flow. Qiu *et al*. [21] proposed a domain adaptation approach that models images under different poses as linear combinations of multiple parametric functions. Given a gallery or a probe image, a frontal facial

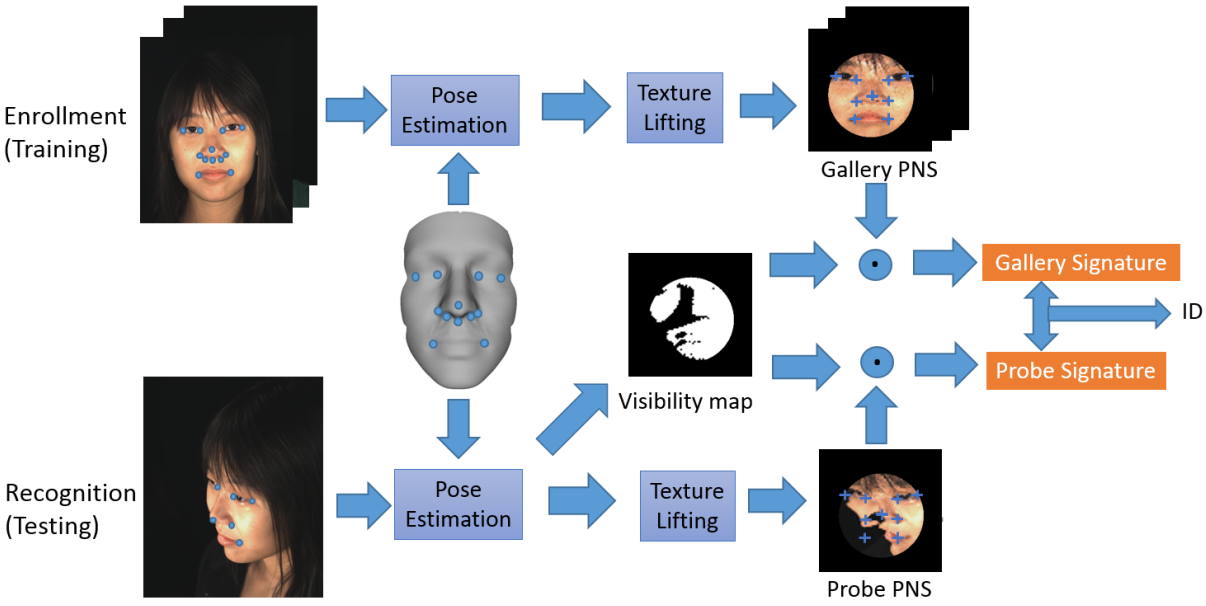

Figure 1. Depiction of the overview of a normalization-based method. The blue dots depict the landmarks used for pose estimation. The blue crosses depict the positions used for extracting local features.

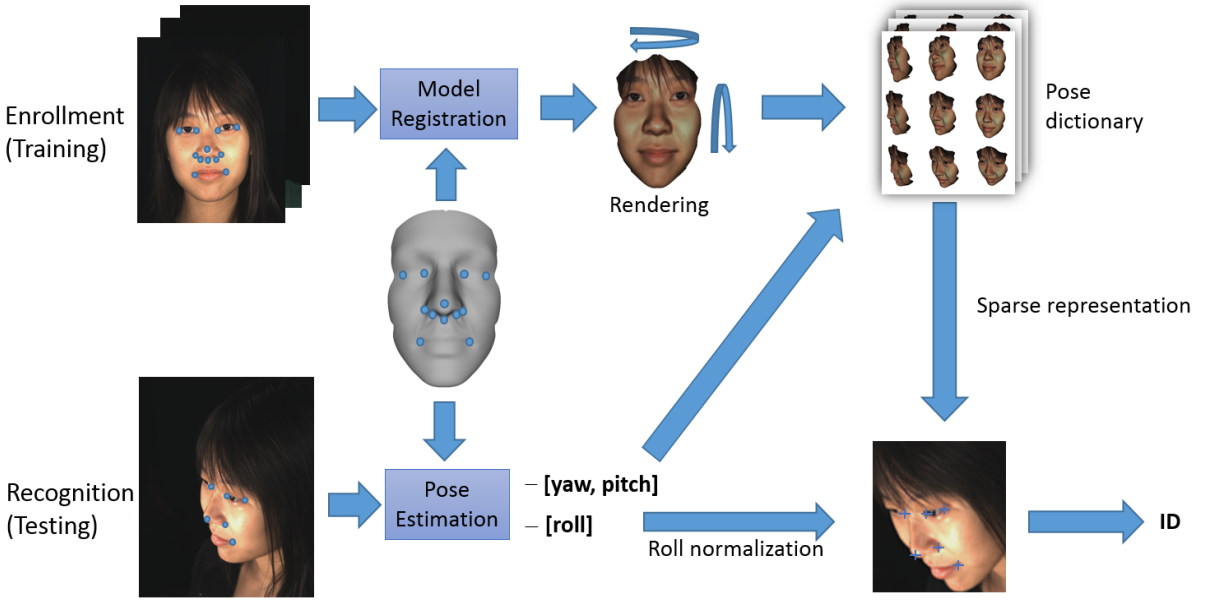

Figure 2. Depiction of the overview of a rendering-based method. The blue dots depict the landmarks used for pose estimation. The blue crosses depict the landmark positions used for extracting local features.

image can be synthesized once the sparse coefficients are computed. Even though they have outperformed previous state-of-the-art methods, they are limited to near-frontal face recognition. They have not provided a good solution for when the yaw or pitch of the head pose is large. Recently, 3D-based methods have drawn more attention, including both normalization-based methods and rendering-based methods. One common point of normalization-based methods [25, 5, 13, 23, 11] is that they normalize the head pose of both gallery and probe images to frontal view; the pipeline is shown in Fig. 1. Though broadly applied, this method may introduce artifacts when the facial pose is large and can incur self-occlusions, causing loss of discriminative information. Another problem is that the method is

not robust to pose estimation errors caused by inaccurate landmarks or incorrect 2D-3D projection matrix estimates. An alternative solution [12, 19] rotates each image in the gallery to synthesize and build multiple complementary dictionaries of various facial poses (Fig. 2). Because the number of samples in a dictionary may compensate for inaccurate pose estimation, this method also attained competitive results on several databases.

To the best of our knowledge, this is the first paper to compare these two types of 3D-aided methods as well as investigate their robustness to pose estimation errors.

3. Methods

We analyze two types of 3D-aided face recognition methods under the widely adopted closed-set identification protocol. In our experiments, we assume that there is only one controlled 2D image available for each subject in the gallery. In testing, the input is the 2D facial images with moderate or extreme (*e.g.*, face profile images) pitch or yaw rotations.

3.1. Normalization

We employ the method proposed in [25] by Toderici *et al*. (without light normalization) to obtain a frontal image of the face in a pose normalized space (PNS) with the help of an annotated face model [2, 27], which could either be a personalized or a generic 3D facial model. In this paper, we briefly summarize the pipeline of this method. More details can be found in [3, 25].

After pre-processing the 3D raw data using the method proposed by Kakadiaris *et al*. [2], an Annotated Facial Model is registered to each individual's 3D data to obtain a personalized 3D mesh model. In training, after annotating all of the landmarks in both 3D and 2D faces, a transformation of the mesh model to the 2D gallery image is estimated. Then, the personalized AFM is registered to the image accordingly. Using the fitted AFM, the facial texture from the 2D image is lifted and the pose normalized face representation is generated. In testing, the same process is repeated. Because the subject's identity is unknown, there are two options to obtain the texture of the probe image: generating a texture image using a generic (mean) 3D model or using a personalized 3D model. The generated textures are dot multiplied by the visibility map [8] and then the similarity of the texture between the probe and each subject in the gallery is computed in order to infer the subject's identity (Fig. 1).

There are several works on normalization-based methods [12, 8] that use a reconstructed 3D model from a 2D gallery image to create the PNS. In this paper, we simplify the problem and seek to understand how much improvement could be attained if a personalized 3D model is employed instead of a "mean" model. Any reconstructed model's accuracy will lie between these two extreme cases. Since constructing a 3D facial model from a single image is another research direction that is beyond the scope of this paper, we will not discuss it further in this paper.

3.2. Rendering

For rendering-based methods, following the framework of Moeini *et al*. [19], the toolbox provided in Hassner *et al*. [10] is employed for the AFM rendering as shown in Fig. 2. After estimating the facial pose and establishing the dense correspondence between meshes, the textured AFM in the gallery is rotated to generate multiple synthetic facial images that cover the 2D (pitch and yaw) pose space - either a generic AFM or a personalized AFM can be used in this process. The pose space is discretized between [-92°,92°] with a step of 4°. In testing, the pitch, yaw, and roll of the probe image are estimated using facial landmarks at the beginning, then the input image is rotated so that the eyes are parallel with the x-axis of the image (this is slightly different from [19] in which the author built a 3D dictionary covering all pitch, yaw and roll, which is not essential and is computationally expensive). Then, the nearest neighbors for the estimated pitch and yaw angle among the discretized pose spaces are retrieved along with a 7×7 window that covers the 48 neighbors of the estimated pose. The synthetic facial images within these selected pose space form a dictionary comprising $7\times7\times$N columns (N is the number of subjects) and M rows (M is the length of feature vector). Sparse coding is performed to determine the identity of the probe using the minimum reconstruction error criterion [26].

4. Experimental design

We designed three experiments to analyze the performance of both rendering-based methods and normalization-based methods for 3D-aided face recognition. In the first experiment, we seek to understand whether each method can still perform well if only a 3D mean model is provided instead of a personalized model. The statistics will indicate an upper and an lower bound of the performance using a reconstructed 3D facial model. In the second experiment, we explore the robustness of the two methods. Since the landmark annotation is more likely to be imperfect in most of the real-world applications, we investigate to what extent the face recognition performance will degrade when the pose estimation has some errors. In the last experiment, we explore whether building a pose dictionary is a promising way to untangle the pose error and whether the sparse representation is a potential direction to boost the performance.

In our experiments, two kinds of features were used, namely the global descriptors on the ROI and the local features on the landmarks. Following the recent trend of using binary features to describe the face for computational efficiency, we use four binary features and two classical

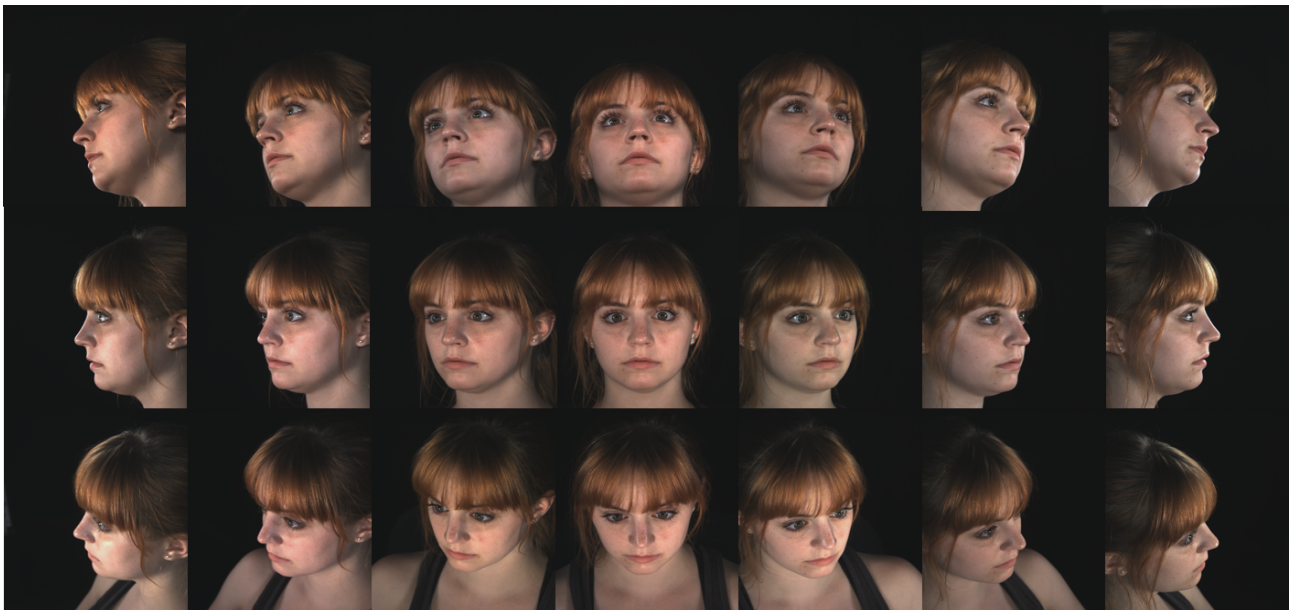

Figure 3. Example 21 views captured by our 21-pod system.

features for evaluation. The HOG [6], LBP [20], and BSIF [15] features are extracted from the global ROI. The BRISK [16], FREAK [4], and SIFT [18] features are extracted locally. All these features have been widely used in 2D face recognition and have shown competitive performance. In this paper, we do not aim to benchmark each feature's performance. For a fair comparison, we use the cosine distance as the similarity metric for HOG and SIFT, and use hamming distance for all the binary features.

4.1. Database

Most of the previous experiments have been conducted on Multi-PIE [9] or LFW [14]. However, these two databases lack the ground truth 3D data needed for a comprehensive analysis. To fill this gap, Kakadiaris *et al*. made available the UHDB11 database [24] which includes personalized 3D data for 3D-aided face recognition. In this paper, we use a new database (UHDB31) comprising 3D facial models and facial images from 21 viewpoints for each of the 77 subjects to evaluate these two 3D-aided face recognition methods. All the data were captured by 21 3dMD^{TM}[1] high resolution cameras properly set-up in a semi-sphere configuration as shown in Fig. 4. An example of 21 facial images of a subject captured using the system is shown in Fig. 3. The average size of the face ROI is around 800×600 pixels. In enrollment, each subject is required to look directly towards the camera located in the center of the system. The system captures all 21 views of the person simultaneously. Concentrating on pose estimation, we control the illumination conditions and use direct ambient illumination during data acquisition. The 3D model derived from the data obtained using the central camera is employed as the ground-truth personalized 3D model for each identity.

Figure 4. Depiction of our 21-pod used to capture the UHDB31 database.

4.2. Implementation details

We manually annotated 12 landmarks on each of the 2D images and the 3D models as shown in Figs. 1 and 2. The pose is estimated based on the visible landmarks among the 12 annotated positions (marked by annotators) in all views. A visibility map is used when comparing two pose-normalized space. The local features on PNS are extracted from nine fixed locations in the PNS space (indicated in Fig. 1). In rendering-based methods, the face in the probe is resized to the similar size as the rendered image before computing the local features on the landmarks.

The local features are extracted from the visible landmarks among all the nine annotated positions (indicated in Fig. 2). When extracting the global ROI features, in either the normalization-based or rendering-based methods, the ROI size is set to be 90×90. For computing the features, we adopt the default parameter settings specified in the original papers.

4.3. Is a personalized 3D model necessary?

In this experiment, we concentrate on comparing the performance of normalization-based methods and rendering-based methods with either a mean or a personalized 3D model. The results are summarized in Table 1.

We observe that using a personalized 3D model indeed improves the performance of rendering-based methods. However, it does not necessarily help normalization-based methods, especially when local features such as SIFT and BRISK are used, in which case better performance is achieved using a mean 3D model. This was a surprising and interesting finding since it is usually assumed that exploiting the personalized 3D model helps improve the performance monotonically. Examining these two methods in detail provides us with the necessary insights. Rendering-based methods discriminate between different subjects based on both facial appearance and facial geometry (contour shape). This information are better preserved by using a personalized 3D model and is better encoded by features computated on a local patch. Specifically, the personalized 3D shape layout around the contours is preserved and generalized very well through the HOG feature when the 3D model is rendered into different aspects; this may help explain why HOG outperforms the other features. When the mean 3D model is used, in rendering-based methods, the discriminative information related to facial geometry is neutralized. On the contrary, in normalization-based methods, the 3D model is used to lift the local texture from the 2D image and transform it into the PNS, where facial geometry is explicitly normalized to achieve pixel-wise correspondence across shape and pose variations. As a result, the personalized 3D information might not contribute too much in this process. This also might explain why the fine-grained texture descriptors (*e.g.*, BSIF and LBP) perform better on PNS.

For keypoint descriptors, including SIFT, BRISK, and FREAK, we observe that using SIFT features extracted from the fiducial points on PNS provides the best result. This is because the normalization-based methods are better at achieving pixel-wise correspondence across different viewpoints and, at the same time, preserving local texture information. In rendering-based methods, however, we observe that the keypoint descriptors perform poorly. There are two reasons. First, in sparse sampling, the dictionary in rendering-based methods might not be robust enough to handle the variations exhibited in the probe image, including not only facial pose, but also the scale. If the face scale in the dictionary of rendering-based methods does not match with the face scale of the probe image exactly, which happens frequently in dealing with facial images from different viewpoints, the face recognition performance will drop significantly. The second reason might be that the information captured on the fiducial landmarks is insufficient to represent the subject's identity with so much information (*e.g.*, the contour of the face, the skin color, the hair style) being discarded. Based on the experimental results, we conclude that dense sampling of local features is critical to handling the large pose problem in rendering-based methods. Due to the poor performance of keypoint descriptors on rendering-based methods, we do not use them in the following experiments.

Table 1. Identification rates (%) for normalization-based methods and rendering-based methods using a personalized (P) or a mean (M) 3D model.

Feature	Normalization-based		Rendering-based	
	P	M	P	M
HOG	51.8	53.2	**83.0**	**77.3**
BSIF	73.2	71.2	65.7	62.9
LBP	65.9	57.4	62.1	59.3
SIFT	**75.8**	**76.8**	1.0	0.0
BRISK	64.4	68.0	1.6	0.4
FREAK	56.2	55.0	0.9	0.0

Figures 6(a) and 6(d) depict the best face identification accuracy of each of the 21 views, which are Mean 3D + SIFT (normalization-based method) and Personalized 3D + HOG (rendering-based method). We observe that even for extreme facial poses, such as profile view, the accuracy of rendering-based methods is still above 85 %, which demonstrates its robustness against facial pose. In contrast, the normalization-based method will generate unreal artifacts as shown in Fig. 1 which may detrimental to recognition. We also observed that it is more difficult to recognize the probe when s/he is looking down rather than looking up.

4.4. Robustness to pose estimation error

In the second experiment, we evaluate the robustness of 3D-aided methods to pose estimation error. We select the two best performing protocols from the last experiments to explore. A generic model is adopted for normalized-based methods, whereas a personalized model is employed in rendering-based methods. In both protocols, the pose is estimated based on the sparse 3D landmarks on a generic 3D model. We gradually add 3°, 6° and 9° (maximum) random noise around all the pitch, yaw and roll angles of the estimated pose. The PNS/pose dictionary is obtained/retrieved under the inaccurate projection matrix. Figure 5 depicts the

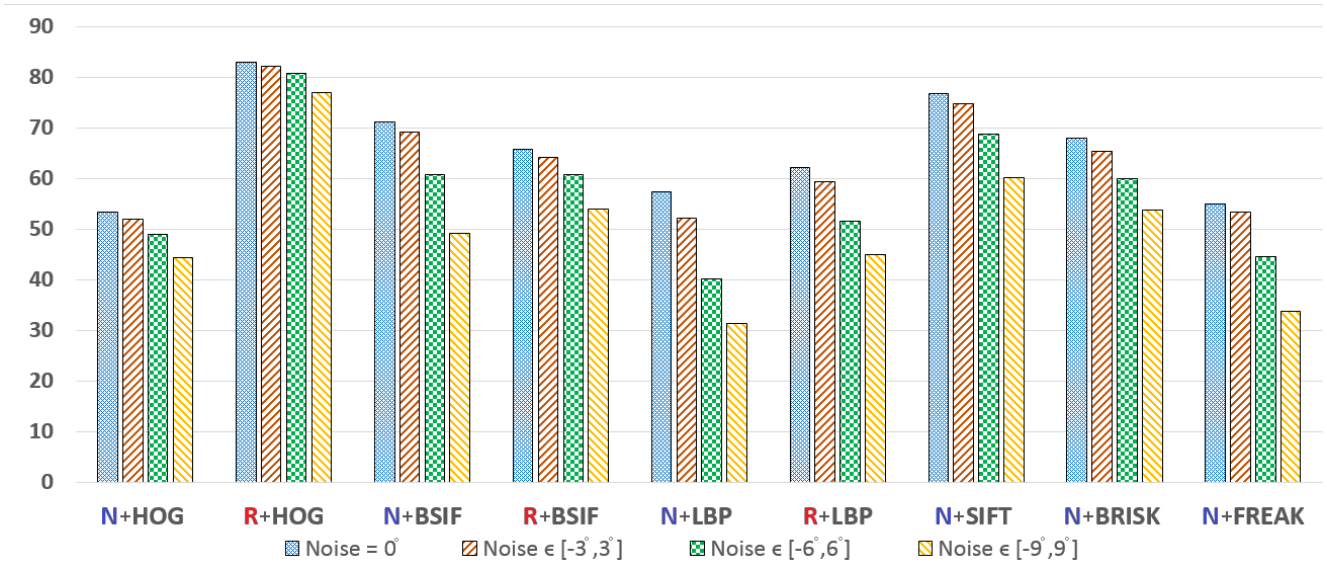

Figure 5. The degradation of identification rate (%) with different pose errors. The characters 'N' in front of the feature names indicate the normalization-based methods. The characters 'R' indicate the rendering-based methods.

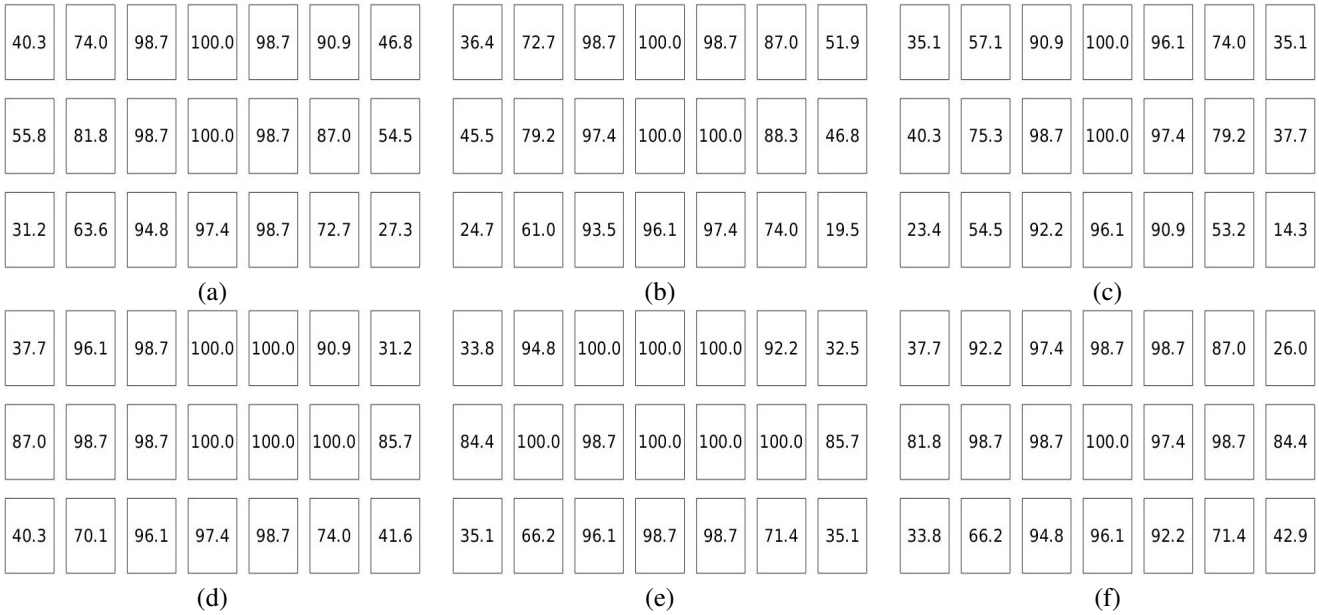

Figure 6. Identification rates (%) of two selected methods under the 21 views shown in Fig. 3. The number in each rectangle represents the identification rate of the corresponding image of Fig. 3. (a-c) correspond to the result of the normalization-based method: SIFT + mean 3D model under the pose error: 0°, 3°, 6°, respectively; (d-f) correspond to the result of the rendering-based method: HOG + Personalized 3D model under the pose error: 0°, 3°, 6°, respectively. The average accuracy over all 21 views are as follows: (a) 76.8 %; (b) 74.7 %; (c) 68.7 %; (d) 83.0 %; (e) 82.1 % and (f) 80.7 %.

tendency of performance degradation of different features under the normalization-based method and rendering-based method. Comparing the results of feature HOG, BSIF and LBP, we observe that the normalization-based method is more sensitive to the pose estimation error than the rendering-based method. We also tracked the performance of two selected features (Fig. 6). When the pose has an average 6° error around pitch, yaw and roll (which might be caused by one or two inexact landmark detections) the average accuracy decrease of normalization-based methods (SIFT) was 10.5 %. However, the rendering-based methods (HOG) has just a 2.8 % degradation.

4.5. Dose sparse coding provide more robustness?

In the third experiment, we aim to test whether sparse coding is able to boost the face recognition performance even in the presence of pose estimation error. In rendering-based methods, the original method is based on sparse coding. To build a method without sparse coding, we simply select one image in the pose dictionary that has the minimal pitch and yaw distance from the probe image and compute the cosine similarity between the probe and each of N subjects in the gallery (N is the number of subjects in gallery). The results are presented in Table 2.

The experimental results in Table 2 demonstrate that in rendering-based methods, the sparse coding boosts the performance when the pose estimation is inaccurate, the average accuracy degradation when using sparse coding is 9.1 %, compared to 19.2 % without sparse coding. The sparse coding also improved the result when there is no noise being added.

Table 2. Identification rates (%) with the rendering-based methods where 0° and 6° represents the pose error. Deg. represents the degradation (%) of identification rate when the error goes up.

Feature	Without sparse coding			With sparse coding		
	0°	6°	Deg.	0°	6°	Deg.
HOG	64.1	55.8	12.9	**83.0**	**80.7**	**2.8**
BSIF	53.8	42.6	20.8	**65.7**	**60.8**	**7.5**
LBP	41.9	31.9	23.9	**62.1**	**51.5**	**17.1**

5. Discussion

Even though better performance could be achieved through extracting higher-level features from images, the simplest features disclose something more elegant and intrinsic. Most of the work in 3D-aided face recognition adopted the framework of normalized-based method and attained superior performance. Our experiments suggest that the rendering-based method might be a better choice when a face has a large deviation from the frontal pose.

From this experimental result, we conclude that rendering-based methods rely on both the facial geometry and facial texture to discriminate different identities, while normalization-based methods use mainly the facial texture on PNS. This might provide some suggestions for future research in both directions. On one hand, explicit 3D facial shape reconstruction might contribute more to rendering-based methods than to normalization-based methods since the facial geometry information could be better utilized in rendering-based methods. On the other hand, a more robust and accurate texture descriptor should be more suitable for normalization-based methods than rendering-based methods, since it is able to better capture facial texture and, at the same time, not be affected by the misalignment of face patches caused by facial pose variations. To the best of our knowledge, there is no research in 3D-aided face recognition that combines the personalized shape (rendering-based methods) and canonical texture information (normalization-based methods) together to boost the performance yet. This may be a promising research direction.

We also observed that pose estimation is critical in 3D-aided face recognition. In both rendering-based methods and normalization-based methods, the accuracy degrades significantly when pose estimation error increases, which is common in applications. As a result, better landmark detectors and pose estimation algorithms are critically needed.

Even though we may end up with a system without accurate head pose estimation, there are still many ways to boost performance. In rendering-based methods, we observe that the features extracted from the global ROI outperform the keypoint descriptors. This demonstrates that dense sampling of features within the whole face ROI is beneficial to achieve better results. In addition, dictionary construction and sparse coding in rendering-based methods provides us another feasible solution to boost robustness. This method could utilize the previous knowledge of 3D shape and simulates possible misalignment through generating a massive number of reference images.

6. Conclusion

In this paper, we evaluated two widely-used frameworks for 3D-aided face recognition. We conclude that the rendering-based methods perform better than normalization-based methods when a face has a large deviation from frontal pose. We observed that the personalized 3D shape information is more important to rendering-based methods than to normalization-based methods. In addition, normalization-based methods can achieve better alignment of facial texture. Both systems contribute to discriminating. We also demonstrated that accurate pose estimation is one of the critical challenges in 3D-aided face recognition.

7. Acknowledgment

We would like to thank Pengfei Dou for his valuable suggestions. This research was funded in part by the US Army Research Lab (W911NF-13-1-0127) and the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

References

[1] 3dMD. 3dMD: 3D Imaging Systems and Software, Nov. 2012.

[2] I. A. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional

face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

[3] I. A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S. K. Shah, and T. Theoharis. 3D-2D face recognition with pose and illumination normalization. *arXiv preprint*, 2015.

[4] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast retina keypoint. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517, Providence, Rhode Island, June 16-21 2012.

[5] O. Aldrian and W. A. Smith. Inverse rendering of faces with a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1080–1093, 2013.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, San Diego, CA, June 20–26 2005.

[7] C. Ding and D. Tao. A comprehensive survey on pose-invariant face recognition. *arXiv preprint:1502.04383*, 2015.

[8] P. Dou, L. Zhang, Y. Wu, S. K. Shah, and I. A. Kakadiaris. Pose-robust face signature for multi-view face recognition. In *Proc. 7^{th} IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, Arlington, VA, September 8-11 2015.

[9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.

[10] T. Hassner. Viewing real-world faces in 3D. In *Proc. IEEE International Conference on Computer Vision*, pages 3607 – 3614, Sydney, Australia, December 1-8 2013.

[11] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 4295–4304, Boston, MA, June 7-12 2015.

[12] G. Hsu and H. Peng. Face recognition across poses using a single 3D reference model. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 869–874, Portland, Oregon, June 25-27 2013.

[13] G. Hu, Chi-Ho-Chan, F. Yan, W. Christmas, and J. Kittler. Robust face recognition by an albedo based 3D morphable model. In *Proc. 2^{nd} International Joint Conference on Biometrics*, Clearwater, FL, September 29 - October 2 2014.

[14] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst, MA, October 2007.

[15] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *Proc. 21^{st} International Conference on Pattern Recognition*, pages 1363–1366, Tsukuba Science City, Japan, November 11-15 2012.

[16] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. In *Proc. IEEE International Conference on Computer Vision*, pages 2548–2555, Tsukuba Science City, Japan, November 11-15 2011.

[17] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2011.

[18] D. Lowe. Object recognition from local scale-invariant features. In *Proc. IEEE International Conference on Computer Vision*, pages 1150–1157, Kerkyra, Greece, September 1999.

[19] A. Moeini, H. Moeini, and K. Faez. Unrestricted pose-invariant face recognition by sparse dictionary matrix. *Image and Vision Computing*, 36:9–22, 2015.

[20] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[21] Q. Qiu, V. M. Patel, P. Turaga, and R. Chellappa. Domain adaptive dictionary learning. In *Proc. European Conference on Computer Vision*, pages 631–645, Firenze, October 2012.

[22] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, June 7-12 2015.

[23] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, June 24-27 2014.

[24] G. Toderici, G. Evangelopoulos, T. Fang, T. Theoharis, and I. A. Kakadiaris. UHDB11 database for 3D-2D face recognition. In *Proc. 6^{th} Pacific-Rim Symposium on Image and Video Technology*, pages 73–86, Guanajuato, Mexico, October 28 - November 1 2013.

[25] G. Toderici, G. Passalis, S. Zafeiriou, G. Tzimiropoulos, M. Petrou, T. Theoharis, and I. A. Kakadiaris. Bidirectional relighting for 3D-aided 2D face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 2721–2728, San Francisco, CA, June 13-18 2010.

[26] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[27] X. Zhao, S. K. Shah, and I. A. Kakadiaris. Illumination normalization using self-lighting ratios for 3D-2D face recognition. In *Proc. European Conference on Computer Vision Workshop*, pages 220–229, Firenze, Italy, October 7-13 2012.