

A Novel Approach to Mugshot Based Arbitrary View Face Recognition

Dan Zeng, Shuqin Long, Jing Li, and Qijun Zhao*

National Key Laboratory of Fundamental Science on Synthetic Vision,
College of Computer Science, Sichuan University, Chengdu 610065, P. R. China

(Received November 18, 2015 : revised February 11, 2016 : accepted March 4, 2016)

Mugshot face images, routinely collected by police, usually contain both frontal and profile views. Existing automated face recognition methods exploited mugshot databases by enlarging the gallery with synthetic multi-view face images generated from the mugshot face images. This paper, instead, proposes to match the query arbitrary view face image directly to the enrolled frontal and profile face images. During matching, the 3D face shape model reconstructed from the mugshot face images is used to establish corresponding semantic parts between query and gallery face images, based on which comparison is done. The final recognition result is obtained by fusing the matching results with frontal and profile face images. Compared with previous methods, the proposed method better utilizes mugshot databases without using synthetic face images that may have artifacts. Its effectiveness has been demonstrated on the Color FERET and CMU PIE databases.

Keywords : Mugshot-based face recognition, Arbitrary view face recognition, Three-dimensional face reconstruction

OCIS codes : (100.5010) Pattern recognition; (100.0100) Image processing; (100.3008) Image recognition, algorithms and filters; (100.6890) Three-dimensional image processing

I. INTRODUCTION

Mugshot face images are widely used for identity recognition in forensic and security applications. They usually consist of frontal and profile face images, which are routinely collected by police. The frontal and profile face images provide complementary information of the face, and are thus believed to be useful for pose-robust face recognition if both of them are effectively utilized. However, most existing automated face recognition methods [1-3] are devised for the scenario of only frontal face images in the gallery. To recognize non-frontal faces, they usually adopt one of the following three ways: (i) Normalizing the non-frontal probe face to frontal pose and matching it to the enrolled frontal faces [4-6], (ii) Generating synthetic face images from the frontal faces in the gallery according to the pose of the probe face, and then comparing the probe face with these synthetic face images [7, 8], (iii) Extracting pose-adaptive features directly from the gallery and probe face images for comparison [9, 10].

When both frontal and profile face images are available in the gallery, in order to utilize the profile face images to better recognize arbitrary view faces, Zhang *et al.* [11] first reconstructed 3D face shapes from the frontal and profile face images, and then recovered facial textures by fitting the Phong reflection model based on the obtained 3D face shapes and the assumed lighting direction prior. They generated synthetic face images of various poses for each enrolled subject to enlarge the gallery. Given a new face image, they first estimate its head pose, and then match it with the gallery face images that have the same pose. A main drawback of this method lies in that the facial textures in the synthetic face images are computed by an assumed model, which may, however, cause artifacts especially when the frontal and profile face images have varying shading effects. Recently, Han *et al.* [12] proposed to directly map the textures in frontal face images onto the reconstructed 3D face shapes based on Delaunay triangulation of the 2D facial fiducial points. Although they avoided using synthetic

*Corresponding author: qjzhao@scu.edu.cn

Color versions of one or more of the figures in this paper are available online.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

textures, they simply used the profile face images for adjusting the depth information of 3D face shapes without further using the texture of the profile face images.

This paper proposes a novel approach to better use mugshot face images to deal with arbitrary view face recognition. Unlike previous approaches, we directly compare probe face images to both frontal and profile gallery face images with assistance of 3D face shape models reconstructed from the gallery mugshot face images. This way, it gets rid of potential artifacts in synthetic face images, and meanwhile effectively exploits the information in both frontal and profile face images. Experiments performed on two challenging public face databases, Color FERET [13] and CMU PIE [14], have proven its effectiveness.

The remainder of the paper is organized as follows. Section II introduces in detail the proposed mugshot based arbitrary view face recognition approach. Section III reports the experimental results on the Color FERET [13] and CMU PIE [14] databases along with some discussion on the results. Finally, a conclusion is drawn in Section IV.

II. METHODS

An overview of the proposed approach is depicted in Fig. 1. As can be seen, it includes two phases: enrollment and matching.

During enrollment, mugshot face images are captured. Here, we assume that three face images of each subject are included, i.e., frontal, left and right profile face images. This is a conventional configuration in forensic mugshot databases. Based on these three face images, the 3D face shape is reconstructed for each subject by using the method

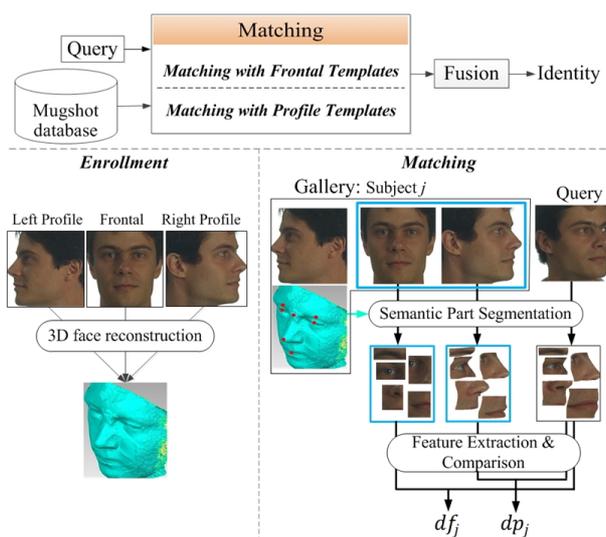


FIG. 1. Overview of the proposed approach. df_j and dp_j denote the distances of the query to the frontal and profile templates of the j^{th} subject, respectively ($j = 1, 2, \dots, L$, and L is the total number of enrolled subjects).

in [15]. Motivated by the component-based face recognition method in [16], we define on the frontal faces a set of local patches (called semantic parts in this paper) around the following seven facial fiducial points: two eyebrow centers, two eye centers, nose tip, nose root, and mouth center (see Fig. 1 and Fig. 2). When a face changes its head pose, these patches will deform according to the pose angles. The deformed patches of a subject under different poses can be easily computed by transforming his/her 3D face shape, and features are extracted from semantic parts on the frontal and profile face images as enrolled templates.

Given a new query face image to be recognized, its head pose is estimated and fiducial points are located. To compare it with a subject in the gallery, its semantic parts are segmented according to its head pose and the subject's 3D face shape. Features are then extracted from the semantic parts, and compared with the enrolled templates. Matching results with the enrolled frontal and profile templates are fused to give the final recognition result, i.e., the identity of the query face image. Next, we introduce in detail the matching steps.

2.1. Semantic Part Segmentation

Segmenting the semantic parts (i.e., deformed local patches around fiducial points) is the key of the proposed approach to handle arbitrary view face images. An essential difficulty in recognizing arbitrary view face images is due to the non-linear deformation occurring to 2D face images when rotating the face in 3D space [10]. To compute the distance between two face images that have different pose angles, it is important to find their corresponding semantic parts because comparison between different semantic parts is nonsense.

In this paper, we consider the following semantic parts. For each enrolled subject, we first project its 3D face shape to 2D plane via orthogonal projection to get a frontal face image. We define ten local patches w.r.t. the seven fiducial points on the 2D frontal face image (See Fig. 2(a)). These patches are then mapped back to the 3D face shape to get the semantic parts in 3D space. Without loss of generality, we assume that the inter-ocular distance is about 90 pixels, and sizes of the semantic parts are then scaled as follows:

- Eyebrow - 20×50 pixels

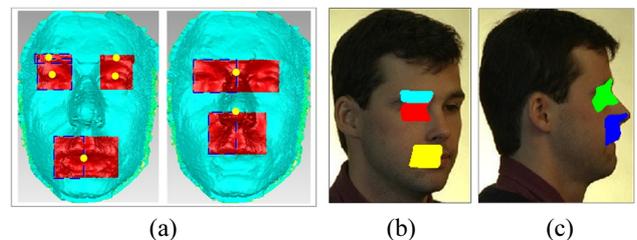


FIG. 2. Illustration of semantic parts. (a) shows the used ten semantic parts. (b) and (c) show the deformed local patches of some semantic parts under different pose angles.

- Eye - 40×45 pixels
- Mouth - 55×50 pixels
- Nose root - 55×50 pixels
- Nose tip - 60×50 pixels

In order to get the semantic parts on an arbitrary view 2D face image to be matched with the subject, the above defined semantic parts have to be projected onto 2D face images. This is done by estimating a 3D-to-2D projection based on the corresponding fiducial points in the 3D face shape and the 2D face image. Let $\{p_i = (x_i, y_i, z_i) | i = 1, 2, \dots, n\}$ and $\{q_i = (u_i, v_i) | i = 1, 2, \dots, n\}$ be, respectively, the 3D coordinates of the fiducial points in the 3D face shape and their 2D coordinates on the corresponding 2D face image (n is the total number of available fiducial points). The projection can be then obtained by fulfilling the following optimization objective,

$$\min_T \sum_{i=1}^n |q_i - Tp_i|^2, \quad (1)$$

where the projection $T = s[R, t]$ is a 2×4 matrix composed by three parameters, i.e., scale s , rotation R and translation t . See Fig. 2.

2.2. Feature Extraction and Comparison

Extracting features from an arbitrary view 2D face image consists of two steps. First, visible semantic parts are segmented according to the head pose of the 2D face image. For example, for frontal and nearly-frontal 2D face images, all semantic parts are visible and will be used for feature extraction and comparison; for left-profile and nearly-left-profile 2D face images, only the semantic parts on the left side of the face are visible, and thus to be considered. Second, once the semantic parts have been segmented from the 2D face image, certain features are extracted from them. In this paper, we extract local binary pattern (LBP) features [17] for the sake of simplicity, but other more elaborated features could be used also.

To compute the distance between two face images, their corresponding visible semantic parts are first determined. The distances between these semantic parts are then computed. In this paper, we employ the chi-square distances for LBP features of the semantic part defined as

$$\chi_w^2(x, \xi) = \sum_{j,i} \frac{(x_{i,j} - \xi_{i,j})^2}{x_{i,j} + \xi_{i,j}}, \quad (2)$$

in which x and ξ are the normalized enhanced histograms of the corresponding semantic part to be compared, indices i and j refer to the i^{th} bin in the histogram corresponding to the j^{th} local region in the semantic part. The size of the local region is 5×5 pixels. The distance between the two face images is finally defined as the average of the distances

between all visible corresponding semantic parts on them, i.e.,

$$d = \frac{1}{N_p} \sum_{i=1}^{N_p} w_i * \chi_w^2(x_i, \xi_i), \quad (3)$$

where N_p is the total number of semantic parts, and w_i is the weight of the i^{th} semantic part.

2.3. Fusion of Frontal and Profile Comparisons

If the probe face image is frontal or near frontal (i.e., facing $[-15, 15]$ degrees), it is directly compared with the enrolled frontal face templates. Otherwise, it will be compared with both frontal and profile templates, and matching results will be fused.

Let df_j and dp_j denote its distance to the j^{th} subject's frontal and profile templates, respectively $j = 1, 2, \dots, L$ and L is the total number of enrolled subjects). Among $\{df_1, df_2, \dots, df_L\}$, we find the first two smallest distances, denoted as df_1 and df_2 . Similarly, the first two smallest distances among $\{dp_1, dp_2, \dots, dp_L\}$ are dp_{p1} and dp_{p2} . The confidence of frontal and profile comparisons is then defined as $S_f = df_{p2} - df_{p1}$ and $S_p = df_{p2} - df_{p1}$, respectively. The identity of the probe face is finally decided as,

$$\text{Identity} = \begin{cases} \text{frontal}_{id}, & (S_p / S_f < \delta) \\ \text{profile}_{id}, & (S_p / S_f \geq \delta) \end{cases} \quad (4)$$

where frontal_{id} is the identity according merely to the matching result with the frontal faces, and profile_{id} is the identity according merely to the matching result with the profile faces. Here, δ is a pre-specified confidence threshold ($\delta = 8$ in this paper).

III. RESULTS

In this section, we evaluate the proposed approach on the Color FERET [13] and CMU PIE [14] databases. Color FERET database contains face images of 768×512 pixels with multiple pose variations (see Fig. 3). Among them, we manually remove the face images whose poses are inaccurately labelled, resulting in 461 subjects used in the experiments. CMU PIE database consists of 884 face images of 68 persons under 13 various poses. The size of these face images is 486×640 pixels. For all the face images, facial fiducial points are manually marked, and they are then normalized to 233×187 pixels by affine transformation based on the fiducial points.

In order to evaluate the effectiveness of comparing deformed semantic parts, we also implement a method that uses the same local features as the proposed method, but extracts features from regular (rather than pose-deformed) local patches (baseline) [17]. In addition, a commercial off-the-shelf face

TABLE 1. Rank-1 and Rank-10 recognition accuracy of different methods: Using regular patches [17] (Baseline, for short), VeriLook [18] and Proposed deformed patches (Proposed, for short) on the Color FERET database [13] at various yaw angles

Yaw	Baseline [7]		VeriLook [8]		Proposed	
	Rank-1	Rank-10	Rank-1	Rank-10	Rank-1	Rank-10
re:-75°	30.0%	93.3%	18.9%	40.0%	82.2%	96.7%
hr:-67.5°	35.6%	89.2%	64.0%	73.2%	84.4%	97.6%
rd:-45°	50.0%	90.0%	67.0%	83.0%	85.0%	99.0%
qr:-22.5°	42.5%	71.5%	96.4%	97.4%	100%	100%
rc:-15°	75.7%	90.8%	98.7%	99.3%	100%	100%
rb:+15°	68.8%	95.5%	99.1%	100%	100%	100%
ql:+22.5°	47.7%	93.3%	97.6%	98.2%	100%	100%
ra:+45°	58.0%	94.0%	57.0%	72.0%	88.0%	100%
hl:+67.5°	33.0%	84.0%	40.5%	57.5%	76.0%	94.5%

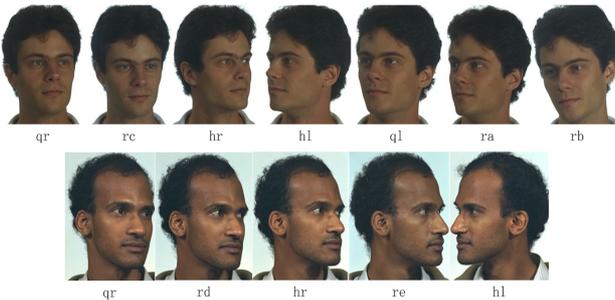


FIG. 3. Example multi-view face images of two subjects in the Color FERET database [13].

matcher, VeriLook [18], is also considered in the comparison experiments.

Table 1 reports the rank-1 and rank-10 recognition accuracy of different methods on Color FERET at various yaw angles. It can be seen that the proposed method works consistently well under varying yaw angles. When yaw angles are larger than 60 degrees, the proposed method still achieves about 80% rank-1 recognition accuracy, whereas the accuracy of the counterpart methods drastically decreased to about 60% or even lower than 30%.

We also compare the proposed method with the one in [11]. It is the only published research work we are aware of that reported mugshot-based face recognition accuracy on the CMU PIE database. For a fair comparison, following the protocol in [11], we choose pose c27 as the frontal face, and c22 and c34 as the left profile face and right profile face, respectively. The remaining 10 poses of 68 persons are considered as probe images. The results are shown in Fig. 4, which again proves the effectiveness of the proposed method in recognizing face images of large pose angles. The rank-1 accuracy of the proposed method is better than the one in [11] especially for large poses (i.e., c31, c14, c25, c02). The recognition rate of the method

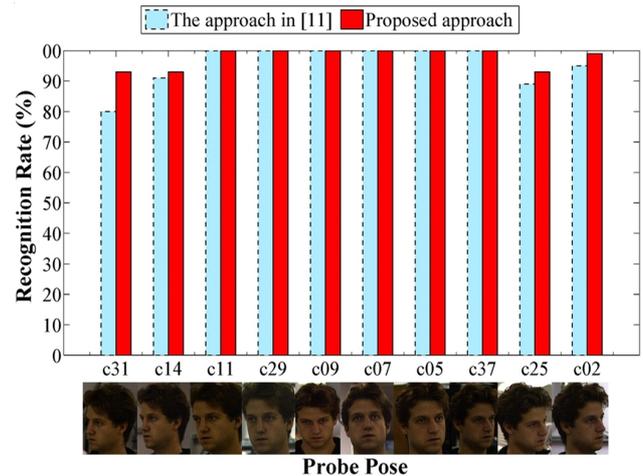


FIG. 4. Rank-1 recognition accuracy of our proposed method and the method in [11] on the CMU PIE database [14].

in [11] for pose c31 is 80%, whereas our recognition rate reaches 93%. The rank-1 accuracy of the proposed method is firmly higher than the other one across poses. The mean accuracy among various poses of probe images by our method and the method in [11] are 97.8% and 95.5%, respectively. The standard deviations across various probe poses by them are 3.32% and 6.86%, respectively.

The superior performance of the proposed method is likely due to several factors. (i) The semantic parts are much more robust to distortions caused by pose variation. (ii) The semantic parts provide pixel-level matching rather than matching local patches around facial fiducial points. (iii) The proposed method can enlarge the inter-class variation thanks to the specific-subject 3D face model to segment the semantic parts. Figure 5 shows some example off-angle probe face images and the results by the proposed and baseline methods.

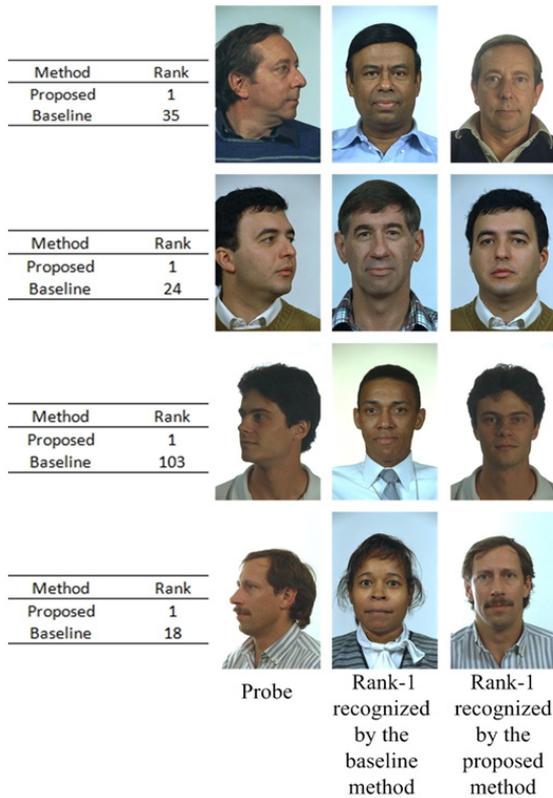


FIG. 5. Example off-angle probe face images from the Color FERET database [13] and the results obtained by the proposed and the baseline methods. The first to third columns show, respectively, the probe face images, the rank-1 gallery face images identified by the baseline method, and the rank-1 gallery face images identified by the proposed method. The table gives the rank of the true gallery face images for these probe images obtained by the two methods.

IV. CONCLUSION

This paper proposed a method to better utilize mugshot databases to recognize arbitrary view face images. It directly matches probe face images to frontal and profile gallery face images based on their corresponding semantic parts. Compared with previous methods, it not only makes better use of both frontal and profile face images, but also avoids synthesizing novel view face images, which often contain artifacts. Evaluation results on public databases have demonstrated the effectiveness of the proposed method in recognizing large off-angle face images. A major limitation of the proposed method is its relatively high computational cost. According to our experiments using Matlab R2012b on an Intel core i5 2.60-GHz desktop computer with 4-GB memory, the proposed method takes, on average, around 9 seconds to recognize a probe face from a background database of 68 subjects. In the future, we will improve the efficiency, investigate automatic facial fiducial point detection methods, and explore more semantic parts and advanced local features.

ACKNOWLEDGMENT

This work has been supported by the National Key Scientific Instrument and Equipment Development Projects of China (No. 2013YQ49087904).

REFERENCES

1. C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," arXiv preprint arXiv:1502.04383 (2015).
2. X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition* **42**, 2876-2896 (2009).
3. 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* **101**, 1-15 (2006).
4. X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Transactions on Image Processing* **16**, 1716-1725 (2007).
5. A. Astnana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," in *Proc. IEEE International Conference on Computer Vision* (Barcelona, Spain, Nov. 2011), pp. 937-944.
6. X. Shao, X. Zhou, C. Cheng, and T. X. Han, "3D face reconstruction and dynamic feature extraction for pose-invariant face recognition," in *Proc. International Symposium on Computer, Communication, Control and Automation* (Singapore, Singapore, Dec. 2013), pp. 119-122.
7. A. Franco, D. Miao, and D. Maltoni, "2D face recognition based on supervised subspace learning from 3D models," *Pattern Recognition* **41**, 3822-3833 (2008).
8. J. Heo and M. Savvides, "Gender and ethnicity specific generic elastic models from a single 2D image for novel 2D pose face synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**, 2341-2350 (2012).
9. A. Li, S. Shan, X. Chen, and W. Gao, "Maximizing intra-individual correlations for face recognition across pose differences," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition* (Miami Beach, Florida, USA, Jun. 2009), pp. 605-611.
10. D. Yi, Z. Lei, and S. Z. Li, "Towards pose robust face recognition," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition* (Portland, Oregon, USA, Jun. 2013), pp. 3539-3545.
11. X. Zhang, Y. Gao, and M. K. H. Leung, "Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases," *IEEE Transactions on Information Forensics and Security* **3**, 684-697 (2008).
12. H. Han and A. K. Jain, "3D face texture modeling from uncalibrated frontal and profile images," in *IEEE International Conference on Biometrics: Theory, Applications and Systems* (Doubletree Hotel, Crystal City, USA, Sept. 2012), pp. 223-230.
13. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms,"

- IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 1090-1104 (2000).
14. T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition* (Washington D.C., USA, May 2002), pp. 46-51.
 15. J. Li, S. Long, D. Zeng, and Q. Zhao, "Example-based 3D face reconstruction from uncalibrated frontal and profile images," in *Proc. IEEE International Conference on Biometrics* (Phuket, Thailand, May 2015), pp. 193-200.
 16. K. Bonnen, B. F. Klare, and A. K. Jain, "Component-based representation in automated face recognition," *IEEE Transactions on Information Forensics and Security* **8**, 239-253 (2013).
 17. T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 2037-2041 (2006).
 18. NeuroTechnology, <http://www.neurotechnology.com>.