

## Automatic Multi-view Face Recognition via 3D Model Based Pose Regularization

Koichiro Niinuma, Hu Han, and Anil K. Jain  
Department of Computer Science and Engineering  
Michigan State University, East Lansing, MI, U.S.A.  
{niinumak, hhan, jain}@msu.edu

### Abstract

One of the major challenges encountered by face recognition lies in the difficulty of handling arbitrary poses variations. While different approaches have been developed for face recognition across pose variations, many methods either require manual landmark annotations or assume the face poses to be known. These constraints prevent many face recognition systems from working automatically. In this paper, we propose a fully automatic method for multi-view face recognition. We first build a 3D model from each frontal target face image, which is used to generate synthetic target face images. The pose of a query face image is also estimated using a multi-view face detector so that the synthetic target face images can be generated to resemble the pose variation of a query face image. Procrustes analysis is then applied to align the synthetic target images and the query image, and block based MLBP features are extracted for face matching. Experimental results on two public-domain databases (Color FERET and PubFig), and a Mobile face database collected using mobile phones show that the proposed approach outperforms two state-of-the-art face matchers (FaceVACS and MKD-SRC) in automatic multi-view face recognition. The proposed approach can also be easily extended to leverage existing face recognition systems for automatic multi-view face recognition.

### 1. Introduction

The goal of automated face recognition (AFR) is to automatically recognize a person from digital images or video sequences containing his face. AFR has attracted substantial attention in the past decades due to its wide applications in real-world scenarios [31] ranging from mobile phone authentication to surveillance. While AFR in controlled conditions, such as frontal or near-frontal poses, neutral expressions and near uniform illumination, has shown impressive performance, AFR in uncontrolled environments, such as arbitrary poses, non-uniform illumination, and partial occlusion, remains a challenging problem [31].

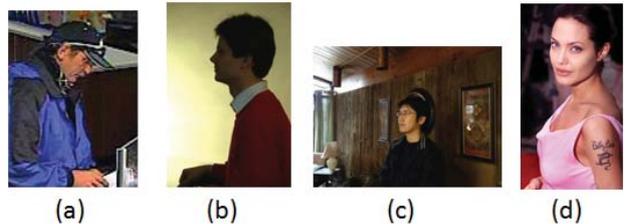


Figure 1. Scenarios and examples of images of face recognition in uncontrolled environments. (a) Profile face image which led to arrest<sup>1</sup>, (b) Non-frontal face image in FERET dataset [27], (c) Non-frontal face image in the Mobile dataset collected in our laboratory, (d) Non-frontal face image in PubFig dataset [17].

One typical application of AFR in uncontrolled conditions is identification or authorization of individuals with face images or videos captured by mobile devices, such as handheld terminals, mobile phones, or surveillance cameras (see Fig. 1). In these scenarios, there is a high possibility that the face images are captured without the cooperation of subjects. As a result, faces in the query images can be of arbitrary poses. Fig. 1 (a) shows an example where a profile face image led to the arrest of a robbery suspect. Despite the potential value of non-frontal face images in forensic applications, the arbitrary pose variations have become one of the primary stumbling blocks for most existing systems to perform face recognition automatically.

This paper proposes a fully automatic multi-view face recognition method that

1. does not necessitate manual landmark annotations or the assumption of known poses within a limited range,
2. achieves higher performance than two state-of-the-art face matchers in several scenarios with different pose variations,
3. and naturally facilitates the application of existing AFR systems in uncontrolled environments.

<sup>1</sup>[http://tonn.rssing.com/chan-1762138/all\\_p47.html](http://tonn.rssing.com/chan-1762138/all_p47.html)

Table 1. A comparison of existing methods for multi-view face recognition.

	Publication	Approach	Pose assumed to be known?	Manual annotation required for non-frontal face image?	Databases used (pose variations)	
Pose invariant feature extraction	Holistic	Sharma <i>et al.</i> [33]	Partial Least Squares, Bilinear Model, Canonical Correlation Analysis	Yes	Yes	FERET ( $\pm 60^\circ$ ) CMU-PIE ( $\pm 90^\circ$ ) Multi-PIE ( $\pm 90^\circ$ )
		Li <i>et al.</i> [19]	Partial Least Squares	Yes	Yes	Multi-PIE ( $\pm 90^\circ$ ) CMU-PIE ( $\pm 90^\circ$ )
		Fischer <i>et al.</i> [10]	Partial Least Squares	Yes	Yes	Multi-PIE ( $\pm 90^\circ$ )
		Prince <i>et al.</i> [29]	Tied Factor Analysis	Yes	Yes	FERET ( $\pm 90^\circ$ ) CMU-PIE ( $\pm 90^\circ$ ) XM2VTS ( $\pm 90^\circ$ )
		Li <i>et al.</i> [18]	Linear Regression	Yes	Yes	FERET ( $\pm 60^\circ$ ) CMU-PIE ( $\pm 90^\circ$ )
		Blanz and Vetter [6]	3D Morphable Model	No	Yes	FERET ( $\pm 60^\circ$ ) CMU-PIE ( $\pm 90^\circ$ )
		Wang <i>et al.</i> [36]	Orthogonal Discriminant Vector	No	No	FERET ( $\pm 25^\circ$ ) CMU-PIE ( $\pm 15^\circ$ ) Yale B, AR
Local	Kanade and Yamada [16]	Subregion Based Probabilistic Model	Yes	Yes	CMU-PIE ( $\pm 90^\circ$ )	
	Ashraf <i>et al.</i> [4]	Probabilistic Stack-flow	Yes	Yes	FERET ( $\pm 60^\circ$ )	
	Lucey and Chen [22]	Patch-whole Sparse Registration	No	Yes	FERET ( $\pm 60^\circ$ )	
	Castillo and Jacobs [7]	Stereo Matching	No	Yes	CMU-PIE ( $\pm 90^\circ$ )	
	Arashloo and Kittler [3]	Markov Random Field	No	No <sup>2</sup>	CMU-PIE ( $\pm 90^\circ$ ) XM2VTS	
Pose Normalization	To frontal	Liao <i>et al.</i> [21]	Multi-keypoint Descriptor	No	No	PubFig (Arbitrary)
		Chai <i>et al.</i> [8]	Linear Regression	Yes	No	CMU-PIE ( $\pm 45^\circ$ )
		Sarfraz and Hellwich [32]	Multivariate Regression	Yes	No	CMU-PIE ( $\pm 90^\circ$ ) FERET ( $\pm 60^\circ$ )
		Li <i>et al.</i> [20]	Morphable Displacement Field	Yes	No	FERET ( $\pm 60^\circ$ ) CMU-PIE ( $\pm 90^\circ$ )
		Teijeiro-Mosquera <i>et al.</i> [35]	Active Appearance Model	No	No	CMU-PIE ( $\pm 45^\circ$ )
	To non-frontal	Asthana <i>et al.</i> [5]	View Based Active Appearance Model	No	No	FERET ( $\pm 40^\circ$ ) CMU-PIE ( $\pm 45^\circ$ ) Multi-PIE ( $\pm 45^\circ$ ) FacePix ( $\pm 45^\circ$ )
		Ding <i>et al.</i> [9]	Random Forest Embedded Active Shape Model	No	No	FERET ( $\pm 60^\circ$ ) CMU-PIE ( $\pm 67.5^\circ$ ) CAS-PEAL ( $\pm 45^\circ$ )
		Prabhu <i>et al.</i> [28]	3D Generic Elastic Model	No	No	Multi-PIE ( $\pm 60^\circ$ ) Video Clips
		Han and Jain [12]	3D Modeling from two images	No	Yes	FERET ( $\pm 22.5^\circ$ )
		Our approach	3D Based Pose Regularization	No	No	FERET ( $\pm 90^\circ$ ) Mobile ( $\pm 90^\circ$ ) PubFig (Arbitrary)

## 1.1. Related Work

Over the past couple of decades, many methods have been proposed to handle the pose variation problem in AFR (see Table 1). These approaches for multi-view face recognition can be grouped into two main categories: (i) pose invariant feature extraction, and (ii) pose normalization. Approaches in the first category aim to provide a common representation which maximizes the correlation among subject’s face images with different poses. They can be further classified into (i) holistic representation, and (ii) local representation. For holistic representation, linear regression, partial least squares (PLS), Bilinear Model (BLM), Canonical Correlation Analysis (CCA), 3D Morphable Model, are widely used approaches [6, 10, 18, 19, 29, 33, 36] that project face images with different poses into latent spaces,

<sup>2</sup>A bounding box is required, but it is not clear if the bounding box is obtained manually or automatically.

where a pose-independent representation is obtained. The merit of these approaches is that the pose variation problem and feature representation are solved simultaneously. However, many holistic methods assume that the poses of face images are known. For example, the poses provided in the databases are directly used to build pose-specific models, and only the model covering the pose of a testing image is used for recognition. Additionally, holistic representation can easily be affected by face deformations due to large pose variations. By contrast, local representations that extract features from individual patches of a face are more robust to large pose variations. Markov Random Field [3], subregion based probabilistic model [16], probabilistic stack-flow [4], patch-whole sparse registration [22], and stereo matching [7] are representative approaches of this category. However, most local representation based approaches [3, 4, 7, 16, 22] require manual landmarks to establish the local patch correspondence between frontal and non-frontal face images.

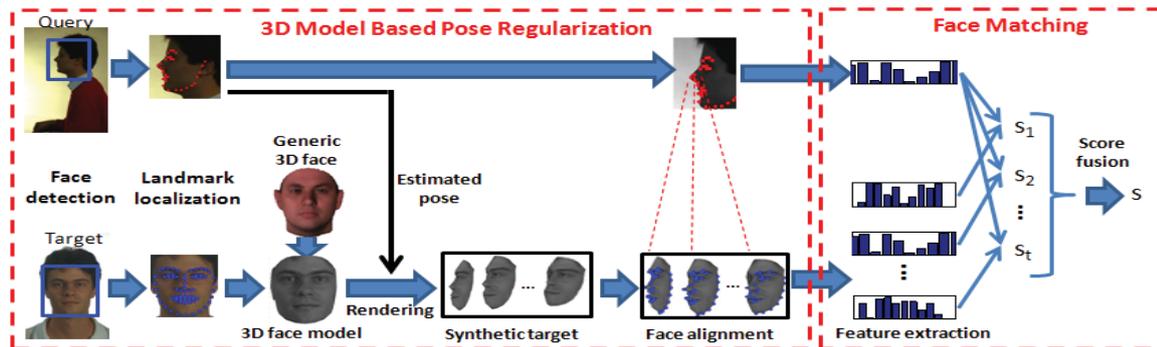


Figure 2. An overview of the proposed approach for multi-view automated face recognition.

Approaches for pose invariant feature extraction usually involves the design of new feature representation and matching methods. By contrast, pose normalization approaches which transform face images with different poses into face images with the same pose, make it possible to directly use existing feature representation and matching methods for multi-view face recognition. Since face recognition techniques for frontal or near-frontal poses have been widely studied, a natural approach is to transform non-frontal face images into frontal images. Linear and multivariate regressions, Active Shape Model (ASM), and Active Appearance Model (AAM) are representative approaches [5, 8, 9, 20, 32, 35] that are used to recover frontal face images from non-frontal views. However, as observed in [5], the recovered frontal face images can be bad due to the self-occlusion under large poses. Li *et al.* [20] avoided the matching of corrupted facial regions in the recovered frontal images by generating occlusion masks. Instead of recovering a frontal image and dropping its corrupted facial regions, a different approach is to generate non-frontal views from frontal images, so that the generated non-frontal views are able to resemble the poses in testing face images. This idea was explored by Park *et al.* [25] using 3D face data. However, 3D sensing is still expensive and the acquisition time can be slow. Also, 2D images constitute the legacy databases; subjects may not be available to provide their 3D images. Under these circumstances, 3D face models reconstructed from frontal face images can be the substitutions for real 3D faces. 3D Morphable Model and 3D generic elastic model (3D GEM), are typical approaches [12, 28] used for generating non-frontal images from frontal views.

Despite various studies in pose invariant feature extraction and pose normalization, most face recognition systems cannot perform fully automatic multi-view face recognition due to the requirements of manual landmark annotations and assumption of known poses. These constraints limit the application of these systems in real scenarios.

<sup>3</sup>For face images with large pose variations, one of the two eyes is

## 1.2. Proposed Method

The proposed method presents a new fully automatic multi-view face recognition method via 3D model based pose regularization, and extends existing face recognition systems into multi-view scenarios. Fig. 2 illustrates the proposed approach which consists of two main modules: (i) Pose regularization based on 3D model, and (ii) Face matching with block based multi-scale LBP (MLBP) features. Unlike previous pose normalization approaches, where non-frontal face images were transformed into frontal images, the proposed 3D model based pose regularization method generates synthetic target images to resemble the pose variations in query images. We should point out that generating non-frontal views from frontal face images is much easier and more accurate than recovering frontal views from non-frontal face images. This is because it is difficult to automatically detect accurate landmarks under large pose variations which are required to build a 3D face model. Additionally, since many areas of a face are significantly occluded under large pose variations, it is problematic to recover the frontal view for the occluded facial regions.

The proposed pose regularization approach is similar to the novel view rendering based on 3D GEM [28], but the proposed method uses a simplified 3D Morphable Model [6]. Additionally, instead of aligning the synthetic target images and testing face images based on eye positions<sup>3</sup>, we perform face alignment using Procrustes analysis under large pose variations. Moreover, our face matching method with blocked MLBP features provides better robustness against face illumination and expression variations<sup>4</sup>. Finally, we show the expansibility of the proposed approach by replacing our MLBP based face matcher with two state-of-the-art face matching systems.

invisible. Under these circumstances, face alignment based on two eyes no longer works.

<sup>4</sup>Following the discussions in [13], additional face preprocessing methods might be integrated with MLBP to further improve the robustness.

## 2. 3D Model Based Pose Regularization

As shown in Fig. 2, to perform pose regularization, we first build a 3D model from each frontal target face image, which is used to generate synthetic target face images. The pose of a query face image is also estimated so that the generated synthetic target face images are able to resemble the pose variation of a query face image.

### 2.1. 3D Modeling from A Frontal Image

In this work, we utilize a simplified 3D Morphable Model [6] without the texture fitting due to its robustness and computational efficiency. We derive our 3D shape model from the USF Human ID 3-D database [2], which includes 3D face shape and texture of 100 subjects captured with a 3D scanner. The original 3D face includes 75,972 vertices, but for efficient computation, we interactively select 76 vertices based on the 76 keypoints defined in an open source Active Shape Model (Stasm [23]). Given the 100 3D faces, the 3D shape of a new face can be represented using a PCA model

$$\mathbf{S} = \bar{\mathbf{S}} + \sum_{k=1}^K \alpha_k \mathbf{W}_k, \quad (1)$$

where  $\mathbf{S}$  is the shape of a new 3D face,  $\bar{\mathbf{S}}$  is the average 3D shape of 100 3D faces from the USF Human ID 3-D database,  $\mathbf{W}_k$  is the shape eigenvector corresponding to the  $k$ -th largest eigenvalue, and  $\alpha_k$  is a coefficient for the  $k$ -th shape eigenvector.

A 2D face image is a projection of a 3D face onto a 2D plane under a set of transformations such as translation, rotation, scaling, and projection. Based on such a face imaging process, the shape of a 3D face can be recovered from its 2D projection (facial landmarks in a 2D face image) by minimizing the following cost function [26]

$$e(\mathbf{P}, \mathbf{R}, \mathbf{T}, s, \{\alpha_k\}_{k=1}^K) = \|\mathbf{P}_{2D} - s \cdot \mathbf{PRTS}\|_{L^2}, \quad (2)$$

where  $\mathbf{P}_{2D}$  is a set of facial landmarks that are detected using Stasm,  $\mathbf{P}$  is an orthogonal projection from 3D to 2D, and  $\mathbf{R}, \mathbf{T}, s$  are the rotation, translation and scaling operations for the 3D face shape  $\mathbf{S}$ , respectively.

We directly use the input frontal face image as the texture corresponding to the frontal 3D facial shape. When a novel view of the face is generated, we directly map the frontal face image to a novel view based on Delaunay triangulation of the 2D facial landmarks. Compared with a statistical face texture model used in 3D Morphable Model, texture mapping in our simplified 3D Model is more efficient. Additionally, texture mapping retains detailed and realistic features that are important for face recognition.

### 2.2. Generating Synthetic Target Images

The recovered 3D facial shape  $\mathbf{S}$  from (1) and (2) is with a frontal pose. By transforming  $\mathbf{S}$  using different transla-

tion, rotation, scaling, and projection transformations, we can easily generate novel synthetic target images from a target face image. Figs. 3 (a) and (b) show two target face images and their synthetic images under 19 novel views ( $\pm 90^\circ$  with an interval of  $10^\circ$ ) using our 3D face model.

However, the synthetic face images should not be generated arbitrarily. In fact, to reduce the pose difference between a target face image and a query face image, the synthetic target face images should be generated to resemble the pose of a query face image. Although in some public-domain face databases (e.g. FERET) the poses for individual face images are available, in many other multi-view face databases, the poses are not known. Under these circumstances, automatic pose estimation from arbitrary face images is necessary in order to perform fully automatic face recognition. In our approach, we utilize a mixture of tree-structured part models (MTSPM) [37] to estimate the pose from a single query face image<sup>5</sup>. With the MTSPM, we are also able to detect a set of facial landmarks, which makes it possible for us to perform alignment between the synthetic target images and the query image. Based on the pose estimation for a query image, only synthetic target face images with similar poses will be generated for face matching.

However, generating synthetic target images online would increase the computational cost of face matching too much. In our approach, we adopt a more efficient strategy. Specifically, after we build a 3D model from each target face image, 19 synthetic target images are generated for each target image offline. Upon the matching of a target image to a query image, only synthetic images with similar poses to the query image will be selected for matching<sup>6</sup>. In our experiments, five synthetic target images are typically used for face matching (see the red rectangles in Fig. 3). This strategy makes it possible for our system to perform large scale face recognition without increasing the computational cost significantly.

### 2.3. Face Alignment

By building a 3D model and generating synthetic target face images to resemble the poses of query images, we are able to reduce the pose disparity between them. However, face alignment is still necessary for the following feature extraction and face matching steps. Holistic face alignment based on two eyes (e.g. Inter-Pupil Distance (IPD)) has been a widely used approach for frontal or near-frontal face images [28]. However, IPD based face alignment becomes problematic for non-frontal poses. Under large pose variations one of the two eyes is often not visible, and even

<sup>5</sup>In [37], the authors discussed the computational cost of MTSPM, and stated that pose estimation with MTSPM could be performed in real-time.

<sup>6</sup>Since the estimated pose by MTSPM is prone to error, instead of using only one synthetic image, multiple synthetic images with similar poses will be used for matching.

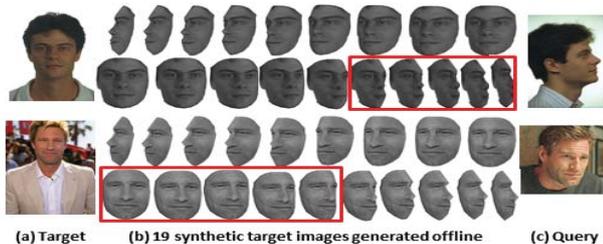


Figure 3. Examples of (a) original target face images, (b) synthetic target images generated offline using 3D face models, and (c) query images. The red rectangles indicate the online selection of synthetic target images based on the pose estimation from query images.

when both eyes are visible in non-frontal images, IPD based alignment can lead to an artificial increase in the overall size of the face image.

In our approach, we apply Procrustes analysis [11] to align the synthetic target images and a query image based on the facial landmarks from a 3D face model and keypoints that are detected by MTSPM. Although the numbers of keypoints defined in a 3D face model and MTSPM are different, the keypoint sequence in each model is fixed. This makes it possible for us to manually establish the keypoint correspondence between two models. We have manually identified 19 landmarks in MTSPM that have corresponding landmarks in a 3D model. The Procrustes analysis is performed based on 19 corresponding landmark pairs.

### 3. Face Matching

Given the aligned synthetic target face images and a query face image, we extract MLBP [24] features for face matching. In our experiments, we use MLBP features which are a concatenation of LBP histograms with 8 neighbors sampled at different radii  $R = \{1; 3; 5; 7\}$ . We first divide a holistic face image ( $256 \times 192$ ) into 768 sub-regions ( $8 \times 8$  non-overlapped blocks). Then, MLBP features are extracted from individual blocks and concatenated together to represent a face.

Given two MLBP histograms  $\mathbf{x}$  and  $\mathbf{y}$  with  $n$  dimensions which are extracted from two face images, chi-squared distance  $\chi^2$  is calculated as a measure of similarity between two face images:

$$\chi^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{(x_i + y_i)/2} \quad (3)$$

where  $x_i$  and  $y_i$  are the features for  $i$ -th bin. Since we have multiple synthetic target images, multiple distances are calculated. The final distance between a target and a query is calculated by finding the minimum of these distances.

<sup>7</sup>The LFW database [14] also includes arbitrary pose variations, but the

## 4. Experiments and Results

### 4.1. Databases and Baselines

Two public-domain databases (Color FERET [27] and PubFig [17]<sup>7</sup>) and a Mobile dataset collected in our laboratory using mobile phones are used to evaluate the performance of the proposed approach for fully automatic multi-view face recognition. The Color FERET database includes facial images with multiple poses from 994 subjects. We use one frontal image (fa) per subject as the target, and images with 6 non-frontal poses (ql, qr, hl, hr, pl, pr) as the query. The FERET database has advanced the development of multi-view face recognition systems in the past ten years. However, the FERET database is collected under a well controlled scenario. For example, the participants are required to rotate the head and body to pre-designed directions<sup>8</sup>, and the background and illumination in face images are nearly uniform. To replicate the scenarios of face recognition from images or videos captured using mobile devices, we have collected a Mobile dataset consisting of 112 subjects using iPhone 4S. For each subject, one or two frontal face images<sup>9</sup> and around 10 non-frontal face images were captured at several locations inside a building (see some example images in Fig. 7). Compared with the FERET database, the Mobile dataset has less subjects but more challenging background and illumination variations, as well as motion blurs due to the movement of the hand. The PubFig database [17] contains 200 famous personalities collected from the Internet, where 60 subjects are designed for algorithm development, and the remaining 140 subjects are designed for algorithm evaluation. Since our method is a non-learning based approach, we directly evaluate our method using the 140 subjects from the evaluation set. One frontal face image per subject is used as the target set, and 513 non-frontal images with arbitrary poses are used as the query set.

The proposed approach is fully automatic in performing multi-view face recognition. For fair comparison, a state-of-the-art system (Multi-keypoint descriptor based sparse representation (MKD-SRC) [21]) reviewed in Table 1 that is also fully automatic in multi-view face recognition is used as the baseline<sup>10</sup>. Additionally, we also compare the proposed approach with a Commercial-Off-The-Shelf (COTS) face matching system (FaceVACS [1]).

While most existing systems are evaluated under small yaw rotations, we evaluate the proposed approach under three scenarios: (i) Small yaw rotations (Typically two eyes are visible.); (ii) Large yaw rotations, and (iii) Arbitrary pose variations. We also investigated the extensibility of

images of many subjects are captured in the same environment.

<sup>8</sup>[http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html)

<sup>9</sup>Only one frontal face image per subject is used as a target.

<sup>10</sup>We would like to provide comparisons with more existing systems, but most are unavailable.

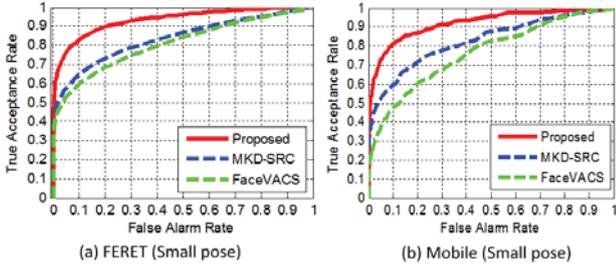


Figure 4. Face verification under the scenario of small yaw rotations on the (a) FERET and (b) Mobile databases.

the proposed method by incorporating pose regularization in our approach with two state-of-the-art matchers (MKD-SRC and FaceVACS).

For databases with limited subjects and images, face verification experiments are more convincing than face identification experiments. Thus, in all our experiments below, face verification is performed under three scenarios, and the performance is reported in terms of receiver operating characteristic (ROC) curve.

## 4.2. Small Yaw Rotations

Face images with small yaw rotations are commonly used by existing multi-view face matching systems in their evaluations (see Table 1). In this section, we also provide an evaluation of the proposed approach under small yaw rotations. Considering that there is no standard definition for small or large yaw rotations, we regard the yaw rotations where two eyes can be automatically detected by FaceVACS as small poses in our experiments, and vice versa. Face images with small yaw rotations from the FERET and Mobile datasets are used in these experiments.

Figures 4 (a) and (b) show the performance of the proposed approach, MKD-SRC, and FaceVACS on the FERET and Mobile databases, respectively. As we discussed above, the FERET database is collected under a more controlled condition than our Mobile database. From Fig. 4 (a), we can find that in a more controlled condition (the FERET database), the state-of-the-art multi-view face matcher MKD-SRC shows no significant advantages than the COTS system FaceVACS, which is mainly designed for near-frontal face recognition. By contrast, the proposed approach achieves more than 15% higher face verification rate than MKD-SRC and FaceVACS at 0.1 False Alarm Rate (FAR). However, while the proposed approach and MKD-SRC are found to be robust to the challenges in the Mobile database, 10% performance degradation can be observed for FaceVACS at 0.1 FAR. The comparison between Fig. 4 (a) and (b) reveals that both the proposed approach and MKD-SRC are more robust to background and illumination variations, as well as motion blurs in the Mobile database

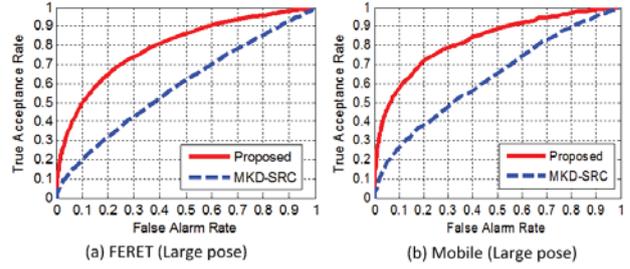


Figure 5. Face verification under the scenario of large yaw rotations on the (a) FERET and (b) Mobile databases.

than FaceVACS, but the proposed approach is more effective than MKD-SRC in handling small pose variations.

## 4.3. Large Yaw Rotations

As shown in Fig. 1 (a), a  $90^\circ$  profile face image captured by surveillance cameras can be the important clue to find the potential suspect. This suggests that in real applications we must also consider the scenario of large pose variations. In this section, we provide a complementary evaluation for that in Section 4.2 by using face images with large yaw rotations. We should point out that under the scenario of large yaw rotations, FaceVACS is no longer available as a baseline because no faces can be enrolled.

Figures 5 (a) and (b) show the performance of the proposed approach and MKD-SRC on the FERET and Mobile databases, respectively. While the state-of-the-art system MKD-SRC gets around 20% verification rates at 0.1 FAR under large yaw rotations, the proposed approach achieves much better performance (50%). We do notice that there is performance degradation compared with that under the scenario of small pose variations in Section 4.2. We attribute the performance degradation as follows. A manual inspection of the images shows that 90% of the query images are  $\pm 90^\circ$  profile images. Under small poses, the eyes and eyebrows have been identified to have the most discriminative ability [34]; however, in  $90^\circ$  profile face images, these discriminative features are no longer available. The lack of discriminative facial features under large pose variations has already motivated the exploration of new features, such as profile curves [15] and ears [30], to represent profile faces.

## 4.4. Arbitrary Pose Variations

Unlike the FERET and Mobile databases, face images in the PubFig dataset are taken in completely uncontrolled situations with non-cooperative subjects. Thus, there are arbitrary variations in pose, illumination, expression, etc. in the PubFig database. In this experiment, in order to use FaceVACS as a baseline, only face images that can be enrolled by FaceVACS are used as the query set. Such a query set may not contain faces with large pose variations, but it

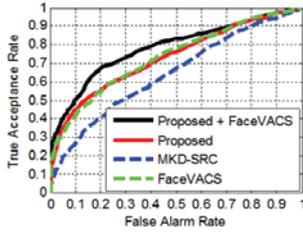


Figure 6. Face verification under the scenario of arbitrary poses on the PubFig database.

covers yaw, pitch, and roll rotations. Therefore, the PubFig dataset provides a challenging scenario for multi-view face recognition systems. Fig. 6 shows the performance of the proposed approach, MKD-SRC and FaceVACS. The proposed approach achieved much better performance than MKD-SRC (45% vs 30% at 0.1 FAR), but nearly the same performance as the COTS matcher FaceVACS. We attribute this to the other types of variations such as illumination, expression and aging. We are able to achieve a leading performance on PubFig by a score level fusion of the proposed approach and FaceVACS.

Fig. 7 shows examples where the proposed approach leads to correct matching of query and target images. We can find that pose regularization in the proposed method greatly reduces the pose gap between target and query images, and therefore improves the face matching accuracy.

#### 4.5. Extensibility of the Proposed Approach

The framework of the proposed approach makes it easy to leverage the other state-of-the-art face matchers, e.g. MKD-SRC and FaceVACS. In this experiment, we replace our MLBP based matcher with MKD-SRC and FaceVACS to evaluate the extensibility of the proposed approach. Since FaceVACS cannot enroll face image under large poses, we preform the experiments under the scenarios of small and arbitrary pose variations. The performance of MKD-SRC and FaceVACS without and with the proposed pose regularization is shown in Table 2. We can find that our pose regularization method greatly improves the performance of two state-of-the-art face matchers, which reveals the good extensibility of the proposed approach.

### 5. Conclusions

We have addressed the problem of fully automatic face recognition across multiple views. The proposed approach consists of 3D model based pose regularization and MLBP based face matching. 3D face models are built from frontal target face images, and used to generate synthetic target images to resemble the poses in query face images. Synthetic target face images and query face images are further aligned using Procrustes analysis, and blocked based MLBP

Table 2. Face verification performance of MKD-SRC and FaceVACS at 0.1 FAR without and with the proposed pose regularization (without/with).

Matchers	Small Pose		Arbitrary Pose
	FERET	Mobile	PugFig
MKD-SRC	65.8%/77.0%	58.7%/70.5%	26.7%/30.6%
FaceVACS	60.4%/81.8%	48.3%/75.6%	41.5%/48.0%

features are extracted for face matching. The proposed approach requires neither a priori pose of the face image, nor manual landmark annotations, and thus is able to perform fully automatic multi-view face recognition. Our experimental results on the FERET, Mobile, and PubFig databases show that the proposed approach outperforms two state-of-the-art face matchers, and facilitates the application of existing AFR systems into multi-view scenarios.

We are in the process of expanding the capabilities of the proposed approach to handle other challenges such as occlusion, illumination, and expression, and exploring supplementary features, e.g. profile curves and ears, to aid in face recognition under large poses. We also plan to improve the 3D modeling accuracy by building 3D models for individual demographic groups, such as age, gender and race.

### References

- [1] FaceVACS Software Developer Kit Cognitec Systems GmbH, <http://www.cognitec-systems.de/>.
- [2] USF DARPA HumanID 3D Face Database. Courtesy of Professor Sudeep Sarkar, University of South Florida.
- [3] S. R. Arashloo and J. Kittler. Energy normalization for pose-invariant face recognition based on MRF model image matching. *IEEE Trans. PAMI*, 33(6):1274–1280, 2011.
- [4] A. B. Ashraf, S. Lucey, and T. Chen. Learning patch correspondences for improved viewpoint invariant face recognition. *Proc. IEEE CVPR*, pages 1–8, 2008.
- [5] 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. *Proc. ICCV*, pages 937–944, 2011.
- [6] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. PAMI*, 25(9):1063–1074, 2003.
- [7] C. D. Castillo and D. W. Jacobs. Wide-baseline stereo for face recognition with large pose variation. *Proc. IEEE CVPR*, pages 537–544, 2011.
- [8] X. Chai, S. Shan, X. Chen, and W. Gao. Locally linear regression for pose invariant face recognition. *IEEE Trans. Image Processing*, 16(7):1716–1725, 2007.
- [9] L. Ding, X. Ding, and C. Fang. Continuous pose normalization for pose-robust face recognition. *IEEE Signal Processing Letters*, 19(11):721–724, 2012.
- [10] M. Fischer, H. K. Ekenel, and R. Stiefelwagen. Analysis of partial least squares for pose-invariant face recognition. *Proc. IEEE BTAS*, 2012.
- [11] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.



Figure 7. Examples of successfully matches using the proposed approach at 0.1 FAR: (a) Target images, (b) synthetic target images using the 3D model that lead to correct matching of the query and target images, and (c) query images.

- [12] H. Han and A. K. Jain. 3D face texture modeling from uncalibrated frontal and profile images. *Proc. IEEE BTAS*, 2012.
- [13] H. Han, S. Shan, X. Chen, and W. Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recogn.*, 46(6):1691–1699, 2013.
- [14] G. B. 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, 2007.
- [15] I. A. Kakadiaris, H. Abdelmunim, W. Yang, and T. Theoharis. Profile-based face recognition. *Proc. FGR*, pages 1–8, 2008.
- [16] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition. *Proc. IEEE CIRA*, pages 954–959, 2003.
- [17] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. *Proc. ICCV*, 2009.
- [18] A. Li, S. Chai, and W. Gao. Coupled bias-variance tradeoff for cross-pose face recognition. *IEEE Trans. Image Processing*, 21(1):305–315, 2012.
- [19] A. Li, S. Shan, X. Chen, and W. Gao. Cross-pose face recognition based on partial least squares. *Pattern Recogn. Lett.*, 32(15):1948–1955, 2011.
- [20] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan. Morphable displacement field based image matching for face recognition across pose. *Proc. ECCV*, 1:102–115, 2012.
- [21] S. Liao, A. K. Jain, and S. Z. Li. Partial face recognition: Alignment-free approach. *IEEE Trans. PAMI*, 35(5):1193–1205, 2013.
- [22] S. Lucey and T. Chen. A viewpoint invariant, sparsely registered, patch based, face verifier. *IJCV*, 80(1):58–71, 2008.
- [23] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. *Proc. ECCV*, 5305:504–513, 2008.
- [24] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. PAMI*, 24(7):971–987, 2002.
- [25] U. Park, H. Chen, and A. K. Jain. 3D model-assisted face recognition in video. *Proc. CRV*, pages 322–329, 2005.
- [26] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE Trans. PAMI*, 32(5):947–954, 2010.
- [27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. PAMI*, 22(10):1090–1104, 2000.
- [28] U. Prabhu, J. Heo, and M. Savvides. Unconstrained pose invariant face recognition using 3D generic elastic models. *IEEE Trans. PAMI*, 33(10):1952–1961, 2011.
- [29] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti. Tied factor analysis for face recognition across large pose differences. *IEEE Trans. PAMI*, 30(6):970–984, 2008.
- [30] A. Ross and A. Abaza. Human ear recognition. *IEEE Computer*, 44(11):79–81, 2011.
- [31] S. Z. Li and A. K. Jain (eds.). Handbook of face recognition, second edition. *Springer*, 2011.
- [32] M. S. Sarfraz and O. Hellwich. Probabilistic learning for fully automatic face recognition across pose. *Image and Vision Computing*, 28(5):744–753, 2010.
- [33] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs. Robust pose invariant face recognition using coupled latent space discriminant analysis. *CVIU*, 116(11):1095–1110, 2012.
- [34] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE*, 94(11):1948–1962, 2006.
- [35] L. Teijeiro-Mosquera, J. L. Alba-Castro, and D. Gonzalez-Jimenez. Face recognition across pose with automatic estimation of pose parameters through AAM-based landmarking. *Proc. ICPR*, pages 1339–1342, 2010.
- [36] J. Wang, J. You, Q. Li, and Y. Xu. Orthogonal discriminant vector for face recognition across pose. *Pattern Recogn.*, 45(12):4069–4079, 2012.
- [37] X. Zhu and D. Ramanan. Face detection, pose estimation and landmark localization in the wild. *Proc. IEEE CVPR*, pages 2879–2886, 2012.