# Towards Pose Robust Face Recognition

Dong Yi, Zhen Lei and Stan Z. Li
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
dyi, zlei, szli@cbsr.ia.ac.cn

## Abstract

*Most existing pose robust methods are too computational complex to meet practical applications and their performance under unconstrained environments are rarely evaluated. In this paper, we propose a novel method for pose robust face recognition towards practical applications, which is fast, pose robust and can work well under unconstrained environments. Firstly, a 3D deformable model is built and a fast 3D model fitting algorithm is proposed to estimate the pose of face image. Secondly, a group of Gabor filters are transformed according to the pose and shape of face image for feature extraction. Finally, PCA is applied on the pose adaptive Gabor features to remove the redundances and Cosine metric is used to evaluate the similarity. The proposed method has three advantages: (1) The pose correction is applied in the filter space rather than image space, which makes our method less affected by the precision of the 3D model; (2) By combining the holistic pose transformation and local Gabor filtering, the final feature is robust to pose and other negative factors in face recognition; (3) The 3D structure and facial symmetry are successfully used to deal with self-occlusion. Extensive experiments on FERET and PIE show the proposed method outperforms state-of-the-art methods significantly, meanwhile, the method works well on LFW.*

## 1. Introduction

Comparing with other biometrics, the most superiority of face biometric is its non-intrusive nature. Therefore, face is one of the most suitable biometrics for surveillance applications. Superiority is always followed by disadvantage. In typical surveillance scenarios, people are usually walking free, and they are impossible to always keep their faces frontal or looking to the cameras. This leads to a problem in face recognition, unconstrained face recognition. Most face images captured by surveillance systems are non-ideal, because they are often affected by many factors: pose, illumination, expression, occlusion, distance, weather and so on. This paper will mainly focus on the pose problem while considering the other factors together.

From the early stages of face recognition research to now [31], pose variation was always considered as an important problem. The problem gained great interest in the computer vision and pattern recognition research community, and many promising methods have been proposed to tackle the problem of recognizing faces in arbitrary poses, such as Illumination Cone Model (ICM) [9], EigenLight Field (ELF) [12], 3D Morphable Model (3DMM) [3], and so on. However, none of them is free from limitations and is able to fully solve the pose problem. As noted in a recent survey [28], the protocols for testing face recognition across pose are even not unified, which indicates we still have a long way to build a fully pose invariant face recognition system.

Existing methods can be divided into two categories: 2D methods and 3D methods (or their hybrid). Because the pose variation is essentially caused by the 3D rigid motion of face, 3D model based methods generally have higher precision than 2D methods. Due to lacking one degree of freedom, 2D methods often use some 2D transformations (*e.g.,* piecewise affine, thin plate splines [10]) to approximate the 3D transformation and compensate the error by some statistical learning strategies. The learning procedures are either conducted in image space or feature space. For example, Locally Linear Regression (LLR) [6] is to learn the relationship between the image patches with different poses. Once the regression function is learned, the similarity of image patches with different poses can be evaluated. Although some good results of 2D methods are reported, their inherent shortages are still constraining their performance, *e.g.,* hard to deal with self-occlusion.

3D methods are always based on a 3D face model, which may be a single model, or a deformable model in certain parametric forms. The flexibility and precision of the 3D face model is the core of 3D methods, therefore we usually call them as 3D model assisted methods. In typical face recognition applications, the enrolled face images (gallery) are usually captured under controlled environment, while

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.454

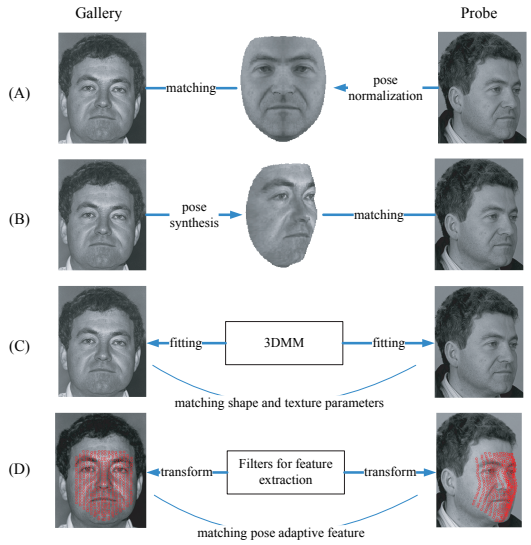


Figure 1. Four kinds of 3D methods for pose robust face recognition: (A) Pose normalization. (B) Pose synthesis. (C) Recognition by fitting. (D) Filter transformation.

the quality of on-site face images (probe) are uncontrolled. In this situation, 3D methods can be divided into four categories depending on how to use of the 3D model:

1. Pose Normalization: Face images in the probe are normalized to frontal view based on the 3D model, and then match the normalized probe to the gallery [2].

2. Pose Synthesis: Use the 3D model to generate some virtual face images with various poses for the face images in the gallery, and then match the probe to the virtual face images [30, 28].

3. Recognition by Fitting: Fit all face images in the gallery and probe by the 3D model. The texture and shape parameters are used for face recognition [3].

4. Filter Transformation: Transform the filters according to the pose and shape of face image, and then use the pose adapted filters for feature extraction.

The processes of four kinds of 3D methods are shown in Fig. 1. As reported in existing papers, the first three kinds of methods all need several minutes to process a face image and their recognition rates are heavily dependent on the precision of the 3D model and optimization algorithm. On the contrary, filter transformation is efficient, because it doesn't need to fit the 3D model to face image in high precision and manipulate the texture of 3D model. Once having the pose and shape of face image, we can transform the filters to adapt to the face image. A early paper [14] has used this idea to achieve good results on the pose problem, in which Gabor filters were transformed according to the pose and normal direction of face surface to construct pose robust features. However, this idea rarely got the attention of face recognition community since then. Limited by the face recognition technologies at that time, the method in [14] is obscure and need many manual steps to construct the whole system. The main purpose of this paper is to propose a novel method to advance the 3D methods along the fourth category. The experiments will prove that our method is fast, flexible and effective. Compared to [14], our method is more systematic, complete and automatic.

The contributions of the paper are as follows.

1. We revisit the filter transformation based methods for the pose problem, which has been neglected for a long time. Inspired by this idea, many filters could be extended for the pose problem, such as Gabor, LBP, HOG and so on.

2. In the framework of filter transformation, we propose a novel pose robust face recognition method, which is both robust to pose variations and other negative factors in face recognition.

3. To meet the speed requirement of practical systems, we propose a fast 3D model fitting algorithm with acceptable precision for face recognition.

4. By experiments on FERET, we verify the importance of facial symmetry to deal with self-occlusion.

5. We improve the state-of-the-art recognition rate across pose on the FERET and PIE databases. Meanwhile, the proposed method achieves comparable results to the best methods on LFW.

## 2. Pose Adaptive Filter

The proposed method in this paper is belong to the fourth category: filter transformation, the main idea of which is transforming filter according to the pose and shape of face image and then using the transformed filter to extract pose robust features. As the filter varies with the change of pose, so we call the method as "Pose Adaptive Filter" (PAF).

The flow chart of PAF is shown in Fig. 2. First, we build a 3D deformable model and define many feature points on the 3D model. Given a 2D face image, we get its pose and shape by fitting the 3D model to the image and then project the defined 3D feature points to the image plane. Finally, pose robust features are extracted at the projected feature points by Gabor filters.

### 2.1. 3D Model and Feature Points Definition

Our 3D face model is similar to the shape part of classical 3D Morphable Model (3DMM) [3], and drops the texture part. The shape deformation is represented by PCA,

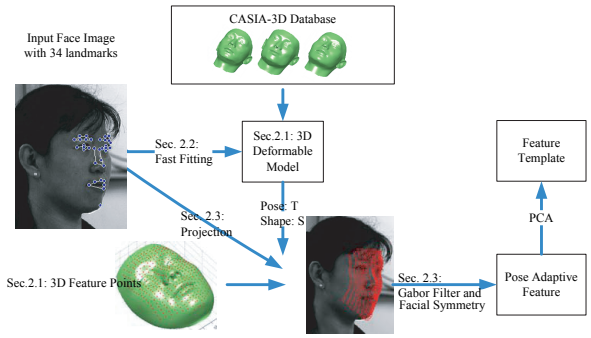


Figure 2. The flowchart of the proposed method: PAF.

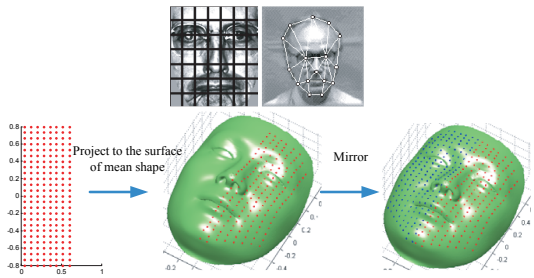


Figure 3. Left: Uniform grid used in LBP based face recognition. Right: User defined 2D feature points on face image. Bottom: Our feature points defined on the surface of 3D face model.

which is trained by the 4624 samples in the CASIA-3D FaceV1 Database [1]. Because the original 3D face have various poses and their cloud points are partial missing, we fit these faces by a generic 3D model with 33640 vertexes and 66750 triangles [5]. After this step, all 3D faces are repaired, aligned well and have the same parametric form. Apply PCA on the aligned 3D faces, we get a deformable 3D face model composed by the mean shape $\mathbf{m}$, eigenvalues $\sigma$ and eigen-shapes $\mathbf{w}$. We use the first 59 eigenshapes to reserve 95% energy.

For most face recognition methods, features are usually extracted on uniform grid [1] or feature points defined on image plane [26]. For in-plane rotation, the uniform grid or 2D feature points can easily adapt to the face image by a similarity transformation, but they cannot work for out-of-plane rotation. To deal with real 3D pose variations, we will define feature points on the surface of 3D face model, which is shown in Fig. 3.

Once having the 3D model, we define the 3D feature points by three steps.

1. Generate $25 \times 8 = 200$ points on image plane (*i.e.,* x-y plane).

2. Project the points from image plane to the surface of the mean shape $\mathbf{m}$ and remove the 24 points out of edges. As shown in Fig. 3, the projected 176 points are at the right half of the mean shape.

3. By mirroring the points at the right half according to facial symmetry, we get $176 \times 2 = 352$ feature points.

The 3D feature points can be represented by their coordinate $(x, y, z)$ or the index $\mathbf{J}$ of vertex on the mean shape. To better deal with the non-rigid deformation of face, we use vertex index in this paper. The advantages of the symmetric structure of the feature points will be illustrated in experiments, which is effective to deal with self-occlusion caused by pose variations. The readers who are interested in facial symmetry can refer to [18].

## 2.2. Fast 3D Model Fitting

3DMM [3] is the most popular model to estimate the pose, lighting, shape and texture parameters of face image. It first uses several landmarks to initialize the parameters and then optimizes them by non-linear optimization. But for a face image, 3DMM usually need several minutes to obtain good result. To appeal the time requirement of practical systems, we propose a fast algorithm to solve the pose and shape parameters, while neglect the other parameters. Compared to 3DMM, our algorithm has lower precision but is good enough for face recognition across pose.

Precise facial landmarking is the basis of 3D model fitting. In this work, we localize the facial landmarks by a three-view Active Shape Model (ASM) [7]. ASM is composed by three parts: PCA shape model, local experts and optimization strategy. For local experts, we use LBP [1] feature and Boosting classifier for each landmark. Based on the output of Boosting classifiers, we can get a confidence map for each landmark. These confidence maps are feed to a landmark mean-shift procedure [20]. Then we can get the positions of all facial landmarks. For robustness and efficiency, the process is repeated several times on two scales.

The training set of our landmarker is constructed from the MUCT database [15]. Subsets a, d and e, with small pose variations, are used to train the first view, and the other subsets (b and c) are used to train the second view. By mirroring the second view, we can get the third view. The final three-view landmarker can detect the landmarks well on face images from -60 to 60 degree. Note that we just use the 34 of 76 landmarks because those landmarks on the boundary of face are unstable to pose variations [2]. For more details about the training process and dataset, one can refer to [27].

Given a face image, landmarks $\mathbf{x}$ on the face image, and their corresponding vertex index $\mathbf{I}$ on the 3D model, we can solve the pose $T$ of face and the shape parameter $\alpha$ by optimizing the following problem.

$$min_{T,\alpha}||\mathbf{x} - T(\mathbf{m}(\mathbf{I}) + \sum_{i=1}^{n} \alpha_i \mathbf{w}_i(\mathbf{I}))||_2^2 + \lambda \alpha^T \Sigma^{-1} \alpha, \quad (1)$$

[1]CASIA-3D FaceV1: http://biometrics.idealtest.org/

where $(\mathbf{m}, \mathbf{w}, \sigma)$ is the 3D deformable model. $\Sigma$ is a diagonal matrix composed by the eigen-values $\sigma$ and $n$ is the number of eigen-shapes. The first term of Equ.(1) is data cost of the landmarks and the second term is regularization term to constrain the shape parameter. While using the weak perspective projection model, $T = s[R, t]$ is a $2 \times 4$ matrix composed by scale $s$, rotation $R$ and translation $t$. Because $T$ and $\alpha$ are coupled by matrix product, we can not get a close-form solution. When $T$ or $\alpha$ is fixed, the objective function is quadratic for another variable. Therefore, we can solve the problem by updating $T$ and $\alpha$ in an alternative way:

1. Let $\alpha = 0$, solve the problem $min_T||\mathbf{x} - T\mathbf{m}(\mathbf{I})||_2^2$ by least squares. At this time, we assume T is a $2 \times 4$ affine matrix.

2. For rigid motion, $T(1, 1:3)$ and $T(2, 1:3)$ must be orthogonal. Hence, we use the method in [4] to orthnormalize these two vectors and can decompose the matrix $T$ to $s[R, t]$.

3. And then $\alpha$ can be updated by solving the problem $min_\alpha||\mathbf{x} - T\mathbf{m}(\mathbf{I}) - \sum_{i=1}^{n} \alpha_i T\mathbf{w}(\mathbf{I})||_2^2 + \lambda\alpha^T\Sigma^{-1}\alpha$, which is a least squares problem too.

4. Repeat step 1 to 3 until $T$ and $\alpha$ are convergent (3-5 iterations are enough in practice).

After getting $T$ and $\alpha$, we say $T$ is the pose of the face image, and its corresponding 3D shape $S$ can be reconstructed by Equ.(2).

$$S = \mathbf{m} + \sum_{i=1}^{n} \alpha_i \mathbf{w}_i \tag{2}$$

### 2.3. Pose Adaptive Feature Extraction

Mapping the face image to the vertexes of $S$, we can get a 3D face with texture, using which we could synthesize face images with new poses. Because the fitting algorithm is coarse and the reconstructed 3D face is far from perfect, we don't use this model to generate face images in PAF. $S$ is just used as a mid-man to extract pose adaptive features.

As described in Section 2.1, the 3D feature points are defined on the surface of 3D face and denoted by the vertex index $\mathbf{J}$. By projecting $S$ to the image plane, we can get the 2D coordinates of the feature points $TS(\mathbf{J})$ on the face image. The process is shown in Fig. 4, from which we can see feature points always have fixed semantic meaning for face images with various poses, *i.e.,* feature points on the nose are always on the nose, and those on the mouth are always on the mouth.

At these feature points, we extract local features by a Gabor wavelet described in [17].

$$\Psi(\mathbf{k}, \sigma) = \frac{k^2}{\sigma^2} e^{\frac{k^2}{-2\sigma^2}x^2} \{e^{i\mathbf{kx}} - e^{-\frac{\sigma^2}{2}}\} \tag{3}$$

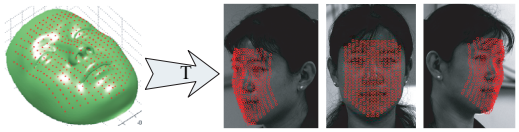


Figure 4. 2D feature points are obtained by projecting the reconstructed 3D shape S to image plane. Three face images of the same subject are shown and their feature points are marked by red dots.

The wavelet is a plane wave with wave vector $\mathbf{k}$, restricted by a Gaussian envelope, the size of which relative to the wavelength is parameterized by $\sigma$. The second term in the brace removes the DC component. Following the popular way, we sample the space of wave vectors $\mathbf{k}$ and scale $\sigma$ in 8 orientations and 5 resolutions, thus giving $5 \times 8 = 40$ complex values for each feature point. Because the phase information is sensitive to mis-alignment, we drop the phase and use the amplitude as feature for face recognition.

Merging the feature values at all feature points together and grouping them by left and right halves, we get two feature vectors with $176 \times 40 = 7040$ dimensions for each face image. To reduce the dimensionality of feature and remove the redundant information, PCA is used to learn a low dimensional subspace. In training phase, the left and right halves are considered as two samples. In testing phase, only the lesser-occluded half is used for matching. By using z-buffer algorithm [25], the occluded area can be easily got based on the pose $T$ and 3D shape $S$. In the PCA subspace, the similarity of feature vectors are evaluated by Cosine metric.

In summary, the proposed PAF deals with pose variations from four aspects: holistic rigid transformation, non-rigid shape deformation, local Gabor filtering, and "half face" selection by facial symmetry. By combining rigid, non-rigid transformations and local Gabor feature, PAF is robust to pose variations and other factors. And the facial symmetry can help it robust to self-occlusion.

## 3. Experiments

The advantages of the proposed method are illustrated on three popular databases: FERET [19], PIE [22] and LFW [13]. FERET is mainly used for algorithm development, on which we compare some methods with PAF and evaluate the performance improvement caused by each step. As the most widely used database in the pose problem, PIE is further used for comparing PAF with state-of-the-art methods. Furthermore, we use LFW to evaluate the performance of PAF under unconstrained environments.

### 3.1. Data Description

FERET has been used to evaluate the robustness of face recognition system to pose in FRVT 2000 [16]. There are 4 sub-experiments named as T1, T2, T3 and T4. In the ex-

periments, 200 frontal images are used as gallery, and face images with pose variations are used as probe. The poses of T1, T2, T3, T4 are (bf: -15, be: 15), (bg: -25, bd: 25), (bh:-40, bc: 40), (bi: -60, bb: 60) degree respectively. The face images in the training set are all fontal.

Because most of existing methods have reported their results on PIE, the comparison with state-of-the-art methods is further performed on the expression subset of PIE, with frontal pose as gallery and the remaining 12 poses as probes. Gallery includes 68 subjects and 1 image / subject. Probes are divided into 12 sub-experiments by their pose, and each probe includes 68 subjects and 96 images. We denote each sub-experiment by its code of pose: c01, c05, c07, c09, c11, c14, c22, c25, c29, c31, c34 and c37. To illustrate the generalization ability of the proposed method, FERET is still used as training set.

Good pose robust face algorithms should perform well against not only pose variations but also other factors. LFW is the best database to evaluate the overall performance of face recognition algorithms under unconstrained environments. The restricted protocol is used in this paper, and the ROC and accuracy are reported by 10-fold cross validation. In the training phase, the label of training pairs are ignored, all the samples are used in an unsupervised way by PCA.

## 3.2. Face Recognition across Pose

### 3.2.1 Performance Analysis of PAF

To evaluate the contributions of each step of PAF, the method and three baselines are tested on FERET.

1. 2L: Two landmarks based alignment (the center of two eyes and the center of mouth), and the full face is used for face recognition.

2. 2L-Half: Two landmarks based alignment, and the lesser-occluded half of face is used for face recognition.

3. PN: All face images are normalized to frontal pose by the 3D model in Subsection 2.1 and the algorithm described in Subsection 2.2 and the missing texture is filled by facial symmetry. Then face recognition is performed on the normalized face images.

4. The proposed method (PAF): The filters are adapted to the pose of face images, and the pose adaptive features are extracted for face recognition.

For fair comparison, the above methods both use Gabor filter and PCA for feature extraction. PN and PAF are also only use the lesser-occluded half of face. 2L vs. 2L-Half is used to evaluate the contribution of facial symmetry. 2L-Half vs. PAF is used to evaluate the contribution of the 3D model. PN vs. PAF is used to compare the "pose normalization" and "filter transformation".

Table 1 shows the results of above four methods on the FERET database. When on experiments with little pose (T1), the difference of all methods are not notable. But as the pose increases, the gaps between these methods are becoming large and the advantage of PAF is significant. Taking T4 as an example, the "half" trick improves the recognition rate from 29.0% to 75.0%, which illustrates the importance of facial symmetry in pose problem. PN applies pose normalization in the image space using the same 3D model and algorithm with PAF, therefore, PN would be expected to have comparable performance with PAF. But PAF is consistently better than PN in all experiments, the main reason may be that PN is too heavily rely on the precision the 3D model. As the precision of the 3D model and the fitting algorithm improves, we think the difference between PN and PAF will decrease.

Table 1 also list three latest methods for reference, in which the automatic pose normalization proposed in [2] is very similar to our PN baseline. We can see the performance of [2] and [20] are comparable to PN and worse than PAF.

### 3.2.2 Comparisons with Other Methods

To compare with other reported methods comprehensively, we evaluate PAF on PIE for each pose. For PAF, frontal images (c27) are used as gallery and other poses are used as probes. For other methods, their gallery and probes are list in Table 2. The training set is remained as the training set of FERET, *i.e.,* the PCA projection matrix of FERET is used in the following experiments. Note that the pose parameters in the second column come from [12] and [28]. Since the authors of ELF didn't provide specific numbers, we estimate the recognition rates from Fig. 6 in [11].

From Table 2 we can see that PAF is almost better than other methods across all poses. Especially for large poses, the superiority of PAF is significantly, even when some other methods use multiple images in the gallery. The last column of Table 2 shows that PAF performs well in most experiments on PIE except for the hardest cases: c22 and c34. This phenomenon indicates that we should focus more on large pose variations ($> 45$ degree) in the future. Furthermore, the mean recognition rate of PAF on 12 poses is 95.31%, which is higher than 93.9% the latest reported result in [21]. Although the performance of CLS [21] in c22 and c34 are higher than our method, but CLS is only evaluated on 34 subjects. On the contrary, our method is evaluated in a more hard setting, training on FERET and testing on all 68 subjects of PIE.

## 3.3. Unconstrained Face Recognition

As the most challenging database in face recognition community, LFW nearly contains all typical variations of face image. Therefore, we use LFW to evaluate the over-

Table 1. The rank-1 recognition rate of the four methods on the FERET database. Some state-of-the-art methods are list for comparison.

| Experiment | T1 (be: 15, bf: -15) | T2 (bd: 25, bg: -25) | T3 (bc: 40, bh: -40) | T4 (bb: 60, bi: -60) |
|---|---|---|---|---|
| 2L | 96.25% | 86.75% | 60.25% | 29.0% |
| 2L-Half | 96.50% | 96.0% | 90.0% | 75.0% |
| PN | 98.25% | 97.25% | 92.50% | 80.75% |
| PAF | 99.25% | 98.50% | 98.0% | 93.75% |
| APN [2] | be: 97.5%, bf: 98.5% | bd: 97.5%, bg: 98.0% | bc: 91.9%, bh: 90.5% | N/A |
| Sarfraz [20] | be: 98.6%, bf: 100% | bd: 97.0%, bg: 89.7% | bc: 89.0%, bh: 92.4% | bb: 82.5%, bi: 79.2% |
| CLS [21] | be: 95.0%, bf: 96.0% | bd: 90.0%, bg: 94.0% | bc: 82.0%, bh: 85.0% | bb: 70.0%, bi: 79.0% |

Table 2. Recognition rate comparison of the proposed PAF and the reported methods: ELF [11], PDM [10], AA-PCA [29], AA-LBP [29], APN [2] and CLS [21].

| | ELF | PDM | AA-PCA | AA-LBP | APN | CLS | PAF |
|---|---|---|---|---|---|---|---|
| Gallery / Probe | c05, c27, c29 | c27 | c22, c27 | c22, c27 | c27 | c27 | c27 |
| c02, yaw: 44 | 68% | 72.5% | 67% | 95% | N/A | 88.0% | 97.92% |
| c05, yaw: 16 | N/A | 100% | 100% | 100% | 100% | 100% | 100% |
| c07, yaw: 0, pitch: -13 | 98% | N/A | 100% | 100% | 98.5% | 100% | 100% |
| c09, yaw: 0, pitch: 13 | 94% | N/A | 100% | 100% | 100% | 100% | 100% |
| c11, yaw: -32 | 83% | 94.12% | 95% | 100% | 98.5% | 100% | 100% |
| c14, yaw: -46 | 76% | 62.5% | 79% | 91% | N/A | 97% | 98.96% |
| c22, yaw: 62 | 38% | N/A | N/A | N/A | N/A | 79.0% | 72.92% |
| c25, yaw: 44, pitch: 11 | 44% | N/A | 61% | 89% | N/A | 85.0% | 100% |
| c29, yaw: -17 | N/A | 98.53% | 100% | 100% | 100% | 100% | 100% |
| c31, yaw: -47, pitch: 11 | 70% | N/A | 67% | 80% | N/A | 91.0% | 98.96% |
| c34, yaw: -66 | 50% | N/A | 65% | 73% | N/A | 85.0% | 75% |
| c37, yaw: 31 | 89% | 97.06% | 98% | 100% | 97% | 100% | 100% |

Table 3. Mean classification accuracy and standard error on LFW View 2.

| Method | Mean | Std |
|---|---|---|
| BIF [8] | 88.13% | 0.0058 |
| I-LQP [24] | 86.20% | 0.0046 |
| PAF | 87.77% | 0.0051 |

all performance of PAF under unconstrained environments besides of pose. All face images in LFW are processed by the assembly line described in Section 2, and then the results are reported according to the restricted protocol. Because PAF is an unsupervised method and uses outside data for alignment, we compare its performance to the methods belonging to the same categories: the best unsupervised method I-LQP [23] and the best method using outside data for alignment BIF (Brain-Inspired Features) [8]. The mean classification accuracy of the methods on View 2 are shown in Table 3, from which we can see that the proposed PAF is comparable to BIF and outperforms the best unsupervised method I-LQP.

### 3.4. Computational Cost Analysis

Speed is a major contribution of this paper. For $480\times640$ image, the mean processing time is 104.2ms on a laptop with P4 CPU@2.0GHz. The details are as follows. 1) Face detection: 12.5ms, 2) Landmarking: 60.9ms, 3) 3D model fitting: 10ms, 4) Gabor filtering: 18.7ms, 5) PCA projection: 2.1ms. Compared to other 3D model based methods, the biggest advantage is the speed of step 3), which just needs 10ms to process an image while other methods usually need several minutes. Due to the simplicity of LBP, our landmarker in step 2) is also faster than other multi-view landmarkers.

## 4. Conclusions

Pose is a challenging and unsolved problem in face recognition. And the pose problem is usually coupled with other factors to jointly affect the performance of practical face recognition systems. To build a fast and pose robust face recognition system, this paper proposed a method PAF to transform filters according to the pose of face image and extract pose adaptive features. Compared to existing 3D model based methods, PAF applied 3D transformation in the filter space instead of the image space, which was faster and less affected by the precision of the 3D model. Experiments on three popular databases illustrated the advantages of PAF from various aspects. From the results on FERET and PIE, we can see PAF outperforms other compared pose

robust methods significantly. Furthermore, PAF achieves comparable performance with state-of-the-art methods on LFW, which shows PAF can work well under unconstrained environments.

## Acknowledgements

This work was supported by the NSFC Project #61070146, #61105023, #61103156, #61105037, #61203267, National IoT R&D Project #2150510, National Science and Technology Support Program Project #2013BAK02B01, CAS Project KGZD-EW-102-2, EU FP7 Project #257289 (TABULA RASA), and AuthenMetric R&D Funds.

## References

[1] T. Ahonen, A. Hadid, and M.Pietikainen. "Face recognition with local binary patterns". In *Proceedings of the European Conference on Computer Vision*, pages 469–481, Prague, Czech, 2004. 3

[2] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. V. Rohith. "Fully automatic pose-invariant face recognition via 3d pose normalization". In *Proceedings of IEEE International Conference on Computer Vision*, pages 937–944, 2011. 2, 3, 5, 6

[3] V. Blanz and T. Vetter. "Face recognition based on fitting a 3d morphable model". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003. 1, 2, 3

[4] A. M. Bruckstein, R. J. Holt, T. S. Huang, and A. N. Netravali. "Optimum fiducials under weak perspective projection". In *Proceedings of IEEE International Conference on Computer Vision*, pages 67–72, 1999. 4

[5] J. D. Bustard and M. S. Nixon. "3D morphable model construction for robust ear and face recognition". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2582–2589, 2010. 3

[6] X. Chai, S. Shan, X. Chen, and W. Gao. "Locally linear regression for pose-invariant face recognition". *IEEE Transactions on Image Processing*, 16(7):1716–1725, 2007. 1

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active shape models: Their training and application". *CVGIP: Image Understanding*, 61:38–59, 1995. 3

[8] D. Cox and N. Pinto. "Beyond simple features: A large-scale feature search approach to unconstrained face recognition". In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 8–15, 2011. 6

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. "From few to many: Illumination cone models for face recognition under variable lighting and pose". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001. 1

[10] D. González-Jiménez and J. L. Alba-Castro. "Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry". *IEEE Transactions on Information Forensics and Security*, 2(3-1):413–429, 2007. 1, 6

[11] R. Gross, I. Matthews, and S. Baker. "Eigen light-fields and face recognition across pose". In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 3–9, 2002. 5, 6

[12] R. Gross, I. Matthews, and S. Baker. "Appearance-based face recognition and light-fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):449–465, 2004. 1, 5

[13] 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, October 2007. 4

[14] T. Maurer and C. V. D. Malsburg. "Single-view based recognition of faces rotated in depth". In *International Workshop on Automatic Face and Gesture Recognition*, pages 176–181, 1995. 2

[15] S. Milborrow, J. Morkel, and F. Nicolls. "The MUCT landmarked face database". *Pattern Recognition Association of South Africa*, 2010. http://www.milbo.org/muct. 3

[16] NIST. Face Recognition Vendor Test (FRVT) 2000. http://www.nist.gov/itl/iad/ig/frvt-2000.cfm. 4

[17] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg. "The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test", 1998. 4

[18] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris. "Using facial symmetry to handle pose variations in real-world 3d face recognition". *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(10):1938–1951, 2011. 3

[19] P. J. Phillips, H. Wechsler, J. Huang, , and P. Rauss. "The feret database and evaluation procedure for face recognition algorithms". *Image and Vision Computing J*, 16(5):295–306, 1998. 4

[20] M. Saquib Sarfraz and O. Hellwich. "Probabilistic learning for fully automatic face recognition across pose". *Image and Vision Computing*, 28(5):744–753, May 2010. 3, 5, 6

[21] 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". *Comput. Vis. Image Underst.*, 116(11):1095–1110, Nov. 2012. 5, 6

[22] T. Sim, S. Baker, and M. Bsat. "The CMU pose, illumination, and expression database". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003. 4

[23] S. ul Hussain and B. Triggs. "Visual recognition using local quantized patterns". In *Proceedings of the European Conference on Computer Vision*, pages 716–729, 2012. 6

[24] S. ul Hussain, Wheeler, T. Napolon, and F. Jurie. "Face recognition using local quantized patterns". In *Proc. British Machine Vision Conference*, volume 1, pages 52–61, 2012. 6

[25] A. Watt. *3D Computer Graphics*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 1993. 4

[26] L. Wiskott, J. Fellous, N. Kruger, and C. v. d. Malsburg. "Face recognition by elastic bunch graph matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997. 3

[27] D. Yi, Z. Lei, Y. Hu, and S. Z. Li. Fast matching by 2 lines of code for large scale face recognition systems. *CoRR*, abs/1302.7180, 2013. 3

[28] X. Zhang and Y. Gao. "Face recognition across pose: A review". *Pattern Recognition*, 42(11):2876–2896, 2009. 1, 2, 5

[29] X. Zhang and Y. Gao. "Heterogeneous specular and diffuse 3-d surface approximation for face recognition across pose". *IEEE Transactions on Information Forensics and Security*, 7(2):506–517, 2012. 6

[30] 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(4):684–697, 2008. 2

[31] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. "Face recognition: A literature survey". *ACM Computing Surveys*, pages 399–458, 2003. 1