

Illumination-Recovered Pose Normalization for Unconstrained Face Recognition

Zhongjun Wu^(✉), Weihong Deng, and Zhanfu An

Beijing University of Posts and Telecommunications, Beijing, China
wuzhongjun1992@126.com, {whdeng, anzhanfu}@bupt.edu.cn

Abstract. Identifying subjects with pose variations is still considered as one of the most challenging problems in face recognition, despite the great progress achieved in unconstrained face recognition in recent years. Pose problem is essentially a misalignment problem together with self-occlusion (information loss). In this paper, we propose a continuous identity-preserving face pose normalization method and produce natural results in terms of preserving the illumination condition of the query face, based on only five fiducial landmarks. “Raw” frontalization is performed by aligning a generic 3D face model into the query face and rendering it at frontal pose, with an accurate self-occlusion part estimation based on face borderline detection. Then we apply Quotient Image as a face symmetrical feature which is robust to illumination to fill the self-occlusion part. Natural normalization result is obtained where the self-occlusion part keeps the illumination conditions of the query face. Large scale face recognition experiments on LFW and MultiPIE achieve comparative results with state-of-the-art methods, verifying effectiveness of proposed method, with advantage of being database-independent and suitable both for face identification and face verification.

1 Introduction

Face recognition has been an active research area for its huge potential in real world applications, such as access control or video surveillance. The focus of face recognition study has shifted from constrained settings to unconstrained settings, as evidenced by the development of face databases, from lab databases, such as FERET [1], MultiPIE [2], to databases in the wild, such as LFW [3]. In unconstrained environment, the irregular conditions of pose, illumination, expression and resolution significantly affects the performance of face recognition system. Among these factors, pose is considered the most challenging one. An excellent solution towards pose variations brings benefit to other tasks such as feature extraction or facial attributes analysis.

Pose problem is essentially a misalignment problem caused by the rigid motion of 3D face structure, resulting in self-occlusion (loss of information) and loss of semantic correspondence [4]. Directly comparing two faces in different poses is difficult. The basic idea is to match pixels in 2D face images to the same semantic 3D facial points by face synthesis.

Existing pose-invariant face recognition methods can be broadly categorized into two families: 2D-based and 3D-based. In the first class, Li *et al.* in [5] represents a test image using some bases or exemplars and the coefficients can be regarded as one kind of pose-invariant features. Local linear regression (LLR) [6] learned appearance transformation between different poses. However, the performance is limited for the incapability of capturing 3D rotations as well as solving self-occlusion problem with using 2D warping. Recent years, deep learning models, such as FIP [7], SPAE [8], MVP [9], CPF [10] have been designed to learn non-linear transformation to convert a non-frontal face to a canonical (frontal) face or several target poses faces and get high recognition rates. But large and well-arranged data has to be prepared and the distribution of test data is usually different from the training data in real world application.

3D-based methods are usually based on a reference 3D face model or a deformable model with shape and illumination parameters, to handle 3D pose variations intuitively. 3D methods are divided into several categories as followed [11].

Recognition by fitting: 3DMM [12] is a powerful 3D representation for human face which fits parameters of 3D shape, pose and illumination and use them for recognition. But it is hard to implement it in practical system for high computational burden.

Pose synthesis: virtual face images under arbitrary poses can be generated using 3D models constructed from gallery images. Probe face is matched to the virtual images with similar pose to the probe. GEM [13] is an efficient 3D face modeling method, which estimates 3D shape by assigning generic face depth information directly to probe 2D images. However, GEM only deals with frontal face, requiring frontal faces for each identity, which is not always satisfied in unconstrained setting.

Pose normalization: 2D probe image is normalized to a canonical (frontal) view based on a 3D model to simplify unconstrained setting to constrained one in terms of pose variations. Asthana *et al.* [14] synthesized a frontal view of the input face by aligning an averaged 3D face model to it, using view-based AAM. But the self-occlusion part is unfilled. HPEN [15] fit the shape parameters of 3DMM and get a complete identity-preserving normalization results by filling the invisible region naturally. But it is based on 68 landmarks detection, where performance may drop due to unprecise localization. LFW3D [16] employed a generic 3D face model to “frontalize” non-frontal images and synthesized the occlusion part based on face symmetry with occlusion degree estimation. But the lighting conditions of output is not consistent with input face when lighting on both sides of face are different and unnatural results will be produced.

In this paper, we propose a continuous face pose normalization method which is identity-preserving and produces natural results in terms of illumination condition, based on only five fiducial landmarks. First, a generic 3D face model is aligned to the input face image based on the detected five landmarks. Then the face contour is detected for purpose of accurately estimating the self-occlusion

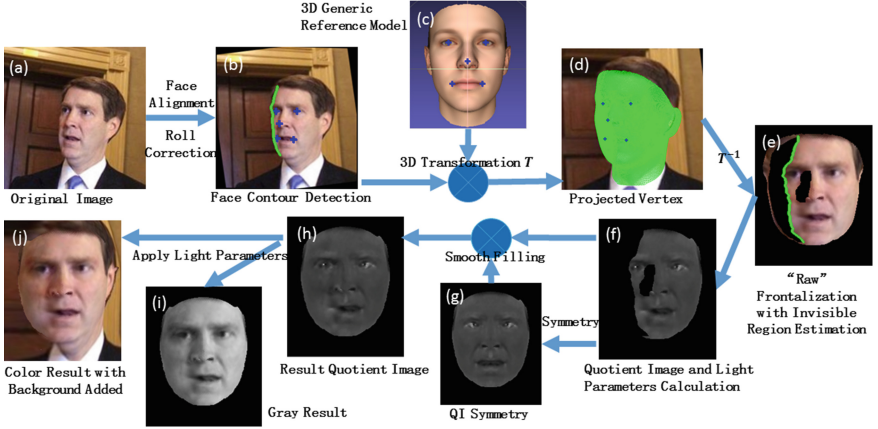


Fig. 1. Visual illustration of proposed pose normalization method. (Color figure online)

part. We can get a “raw” frontalization result by rendering the appearance-assigned 3D mesh at frontal pose with self-occlusion part unfilled (Sect. 2). In order to fill the invisible part naturally, we apply Quotient Image [17] as a face symmetrical feature which is robust to illumination. After estimating lighting parameters and making use of Quotient Image, natural normalization result is obtained where the self-occlusion part is filled with keeping the illumination conditions of input face (Sect. 3). Large scale face recognition experiments on LFW [3] and MultiPIE [2] achieve comparative results with state-of-the-art methods, verifying the effectiveness of proposed method (Sect. 4). The overall procedure of proposed method is shown in Fig. 1.

The advantage of proposed method is that the whole procedure does not depend on any specific training data and can be generalized well in unconstrained setting. Based on only five fiducial landmarks, proposed method is very suitable for practical applications.

2 “Raw” Frontalization

In this part, we will describe the “raw” frontalization process in detail. Inspired by the previous work [14, 16] of using single 3D reference model to make pose normalization, we emphasize on keeping the appearance of input face rather than keeping its shape because the shapes produced from different pose images of the same identity are not guaranteed to be similar. Our target is to obtain highly aligned normalization results for better comparison between different face images.

Given a query image, five stable facial landmarks are located automatically or manually (see the blue ‘+’ in Fig. 1(b)). The five fiducial landmarks in the 3D generic reference model (see Fig. 1(c)) have full correspondence with the

landmarks of the query image. A 3D-to-2D projection matrix T is fitted using generalized least squares solution to the linear system for least square residual:

$$V_{Q-2d} \sim V_{R-3d}T \quad (1)$$

where V_{Q-2d} is a 5×2 matrix with each row representing the (x, y) coordinates of Query-2d landmarks. V_{R-3d} is a 5×4 matrix with each row representing the $(x, y, z, 1)$ coordinates of Reference-3d landmarks where the fourth component 1 is for translation.

The underlying assumption is that sparse correspondence (five points correspondence) is able to represent dense correspondence of face vertices for the reason that human face can be roughly considered as a rigid structure. Although this assumption can not be strictly satisfied, highly aligned results can be obtained by this way. With projection matrix T , all vertices of reference model are projected onto the query image (see Fig. 1(d)) and the intensities of projected positions are assigned to the corresponding vertices by bi-linear interpolation. By rendering the appearance-assigned reference model at frontal pose, we can obtain an initial frontalization result.

When the landmarks do not include face contour, *e.g.*, the five landmarks we used, the problem that the semantic positions of face contour landmarks changes from pose to pose can be avoided. In addition, some previous works, *e.g.*, [14, 15], are based on dozens of landmarks and detecting them accurately for profile faces will be difficult because of severe self-occlusion. With using that five stable landmarks, the ranges that face recognition system can handle with will be extended largely.

2.1 Face Borderline Detection for Self-occlusion Region Estimation

As face deviates from frontal to profile, some regions become invisible due to self-occlusion and it is considered as kind of information loss. Inaccurately estimation of invisible region position will lead to unnatural results because unwanted texture, such as background texture may be introduced to face region. In [14, 18], Z-Buffer [19] is applied to estimate the visibility of each vertex. The idea is that the visibility condition of aligned 3D model approximates the visibility condition of the query face. But when the facial shape of query face differs from the generic face shape largely, the estimation will be inaccurate and introduce unwanted texture.

Face borderline is the boundary that separates visible texture and invisible texture. So face borderline detection facilitate accurate estimation of invisible region. Example comparison of results from Z-Buffer and borderline detection is shown in Fig. 2.

In [14], face borderline detection is formulated as finding a curve running from the top row to the bottom row of a certain rectangle with target of maximize edge strength and smoothness. The smoothness is constrained that the difference between adjacent row is within one pixel. But the objection function with only edge strength is too simple and it would fail with complex background texture in unconstrained environment.

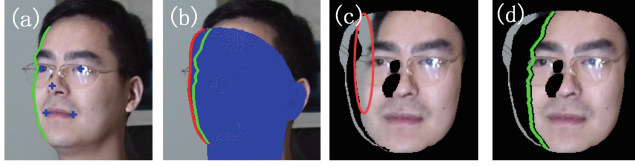


Fig. 2. Comparison of visibility detection from Z-Buffer and borderline detection. (a) Example input image from MultiPIE. (b) Aligned 3D model on input image. The texture on the left side of true borderline (green line) in Z-Buffer method is considered to be visible while actually not. (c) Results of visibility estimation of Z-Buffer. Black pixels indicates invisible. Red ellipse marks the unwanted background texture. (d) Results of visibility estimation using borderline detection which is more accurate. (Color figure online)

We use the information of borderline of projected 3D model and extend the object function in [14] to conduct a robust detection. After aligning the generic reference model into the query image, we can easily detect the borderline of aligned 3D model, which can constrain a certain borderline search region (see the red box in Fig. 3). The gradient magnitude is defined as

$$g(\mathbf{I}) = \left| \frac{\partial}{\partial x} \mathbf{I} \right| + \left| \frac{\partial}{\partial y} \mathbf{I} \right| \quad (2)$$

\mathbf{I} is the search region of the query image. This magnitude is subtracted and divided by the mean and variance of itself for normalization. Since the direction of borderline are close to vertical, in order to reduce imposters, those pixels with large ratio of vertical gradient to horizontal gradient will not be saved.

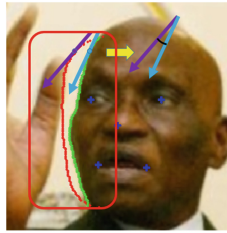


Fig. 3. Similarity of found curve and projected borderline. (Color figure online)

With the borderline of projected 3D model, we introduce the term of similarity between found curve and projected borderline. For each pixel in search region, we calculate its tangential direction through its vertical and horizontal gradient (the blue arrow in Fig. 3), represented as $\mathbf{T}_i(x, y)$. For the projected 3D model borderline, the tangential direction of row y can also be calculated (the purple arrow in Fig. 3), represented as $\mathbf{T}_r(y)$. The similarity of direction

$\mathbf{T}_i(x, y)$ at pixel (x, y) to projected 3D model borderline is calculated as cosine similarity,

$$s(x, y) = \frac{\mathbf{T}_i(x, y) \cdot \mathbf{T}_r(y)}{\|\mathbf{T}_i(x, y)\| \|\mathbf{T}_r(y)\|} \quad (3)$$

The basic idea is that the found curve should share a similar curve shape to the projected 3D borderline. The total optimization problem can be defined as

$$\max_{\{x_i\}} \sum_i g(x_i, y_i) + \lambda \sum_i s(x_i, y_i) \quad (4)$$

with constraint that x_{i-1} and x_i has to be within one pixel. λ is a parameter that balances the importance of gradient magnitude and the importance of curve shape similarity, which is set to 5 in our implementation. This optimization can be solved by dynamic programming and examples of found curves are shown in Fig. 4. The found face contour is back-transformed to the frontal 3D reference model through matrix \mathbf{T}^{-1} and we can get a rather accurate visible region mask as our “raw” frontalization result (see Figs. 1(e) and 2(d)). It is noted that the visibility of nose region is estimated using Z-Buffer method [19].



Fig. 4. Examples of face borderline detection from LFW database.

3 Self-occlusion Region Filling

If the yaw angle of face is too large, some face regions become invisible due to self-occlusion. In order to obtain consistent frontalization result for completely texture comparison, the self-occlusion region should be filled naturally. Asthana *et al.* [14] leaves the invisible region unfilled and can not produce a consistent result. Ding *et al.* [18] use mirrored pixels which would produce incoherent face texture especially when the illumination conditions on both sides of face are largely different. The recent work, LFW3D [16] combines mirrored pixels with occlusion degree estimation but still suffers the illumination inconsistency problem. The self-occlusion problem is kind of information loss and the basic idea is to use face symmetry. In order to keep the illumination condition of input image, we are driven to find a feature that is not sensitive to illumination and satisfy face symmetry condition.

Quotient Image [17] is essentially the ratio of surface reflectance (gray-level) of an object against another object. For example, Caucasian face commonly has higher surface reflectance than Black people face and so has higher value of Quotient Image. Quotient Image feature is only relative to surface reflectance and is insensitive to illumination. It also satisfy face symmetry condition which is suitable for filling the self-occlusion region. We first briefly review the Quotient Image.

3.1 Quotient Image

Face, as a class of object, can be considered as Lambertian Surface with a reflection function: $\rho(u, v)n(u, v)^T s$, where $0 \leq \rho(u, v) \leq 1$ is the surface reflectance (gray-level) associated with point u, v in the image, $n(u, v)$ is the surface normal direction associated with point u, v in the image, and s is the (white) light source direction (point light source) and whose magnitude is the light source intensity.

In [17], the concept *Ideal Class of Object*, i.e., objects that have same shape but differ in surface albedo is defined. Under this assumption, the *Quotient Image* $Q_y(u, v)$ of face y against face a is defined:

$$Q_y(u, v) = \frac{\rho_y(u, v)}{\rho_a(u, v)} \quad (5)$$

where u, v range over the image. Thus, Q_y depends only on the relative surface texture information and is independent of illumination.

A bootstrap set containing N (N is small) identities under M unknown independent illumination (totally $M \times N$ images) is adopted. Q_y of a input image $Y(u, v)$ can be calculated as

$$Q_y(u, v) = \frac{Y(u, v)}{\sum_{j=1}^M \bar{A}_j(u, v)x_j} \quad (6)$$

where $\bar{A}_j(u, v)$ is the average of images under illumination j in the bootstrap set and x_j is linear combination coefficient which can be determined by the bootstrap set images and the input image $Y(u, v)$.



Fig. 5. Example bootstrap images from one identity. The illumination ids are marked as 00–09 in the first row, 10–19 in the second row.

3.2 Illumination Consistence Filling for Self-occlusion Region

We use YUV color space instead of RGB because Y is gray-level and it is independent of the other two channels. We fill the invisible region in Y channels using Quotient Image and combine directly symmetrical UV channels texture to get the final RGB result.

Our bootstrap set is formed by the frontal images from 12 identities under 20 lighting conditions from session *one* in MultiPIE [2] database. The selection of identities hardly affects final result [17]. Example bootstrap set images (gray level) from one identity is shown in Fig. 5.

For the “raw” frontalization result, we have the visible region mask, representing the valid texture, which can be used to estimate Quotient Image and lighting condition. We mask all the images in the bootstrap set using the visible mask of the query image and estimate Quotient Image on valid texture as well as lighting coefficient x_j , which can be represented as:

$$Q_{y-mask}(u, v) = \frac{Y_{mask}(u, v)}{\sum_{j=1}^M \bar{A}_{j-mask}(u, v)x_j} \quad (7)$$

Q_{y-mask} denotes Quotient Image of incomplete frontalization result. Y_{mask} denotes “raw” frontalization result. We make symmetry of the visible side and get Q_{y-sym} , which is blended with Q_{y-mask} smoothly using poisson editing [20] mentioned in [15] and finally we get Q_{y-full} . Since we have estimating lighting coefficient x_j to represent lighting conditions, we combine \bar{A}_{j-full} and x_j to get Y_{full} , represented as:

$$Y_{full}(u, v) = Q_{y-full}(u, v) \cdot \sum_{j=1}^M \bar{A}_{j-full}(u, v)x_j \quad (8)$$

The basic idea of our filling is that we estimate lighting conditions from incomplete valid texture and use it as global representation. As we mentioned before, UV channels of invisible region is filled by directly mirrored pixels and we can get colored frontalization result by back transforming YUV space into RGB space. The ability of keeping illumination consistence is better viewed in

Algorithm 1. Invisible Region Filling

Input: “Raw” frontalization result, bootstrap set images

Output: Full frontalization result

- 1: Mask bootstrap set images with same mask of “Raw” frontalization result.
 - 2: Solve Q_{y-mask} and light coefficient $x_j (1 \leq j \leq 20)$ according to Eq. 7.
 - 3: Mirror Q_{y-mask} and get Q_{y-sym} . Blend Q_{y-sym} into Q_{y-mask} smoothly using poisson editing and get Q_{y-full} .
 - 4: Compute Y_{full} by Q_{y-full} , x_j and full bootstrap set images according to Eq. 8.
 - 5: Mirror UV channels and back transform to RGB space. Adding background texture using affine transformation and get full frontalization result.
-

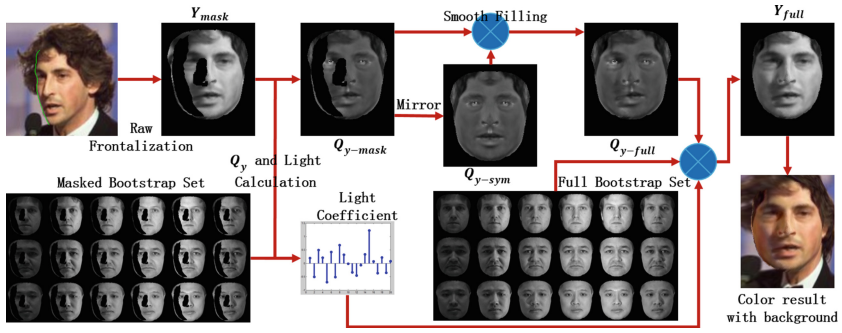
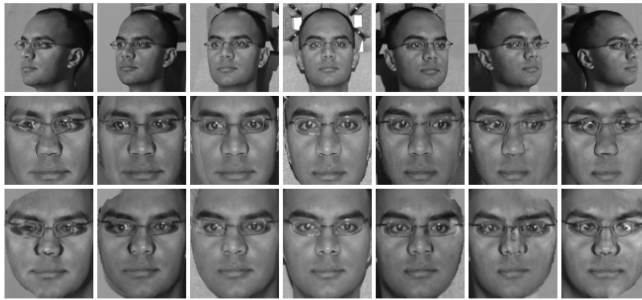


Fig. 6. Process of self-occlusion part filling. The images from 3 identities of 6 lightings in the bootstrap set are shown for convenience. There are actually 12 identities and 20 lightings.



(a)



(b)

Fig. 7. Example frontalization results from (a) LFW and (b) MultiPIE (pose variation from -45° to $+45^\circ$ in step of 15°). First Row: Input images. Second Row: Results of LFW3D [16]. Third Row: Results of Proposed Method. Our results keep illumination consistence and produce less artifacts for accurate borderline detection and smooth filling.

gray result. With adding background texture using affine transformation in [15], a complete frontalization is generated. Figure 6 demonstrates process of self-occlusion part filling, which is also summarized in the following algorithm block. Example frontalization results from LFW and MultiPIE are shown in Fig. 7.

4 Experiments and Results

In this section, we evaluate the performance of proposed method on LFW and MultiPIE databases for face verification and face identification settings respectively.

4.1 Face Verification on LFW

Labeled Faces in the Wild (LFW) [3] is the most commonly used database for unconstrained face recognition this years. LFW contains 13233 face images of 5749 persons collected from Internet with large variations including pose, age, illumination, expression, resolution, etc. We report our results following the “View 2” setting which defines 10 disjoint subsets of image pairs for cross validation. Each subset contains 300 matched pairs and 300 mismatched pairs. We follow the “Image-Restricted, Label-Free Outside Data” protocol and outside data includes BFM [21] as 3D reference model and frontal, multiple illumination facial images from MultiPIE [2] as bootstrap set images in Quotient Image.

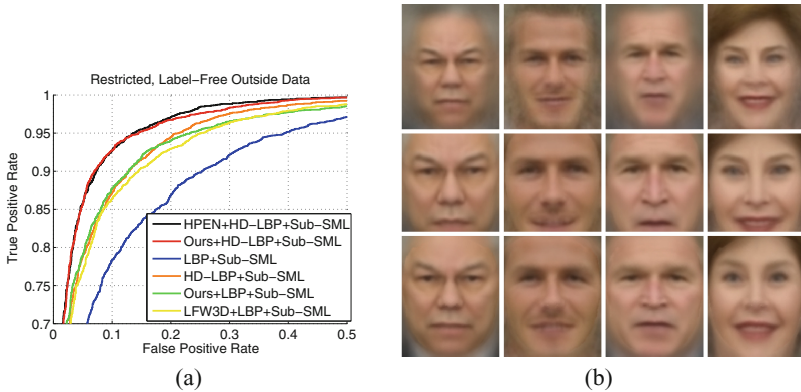
For an input image, we frontalized it with invisible region filling when the estimated yaw angle is larger than 13° . We compare our method with two state-of-the-art method in terms of 3D face frontalization, HPEN [15] and LFW3D [16]. High dimensional LBP (HD-LBP) [22] is extracted on HPEN and proposed method for comparison. The images released by LFW3D are $90 * 90$ pixels only containing face region, which are not suitable for HD-LBP extraction. So we just extract LBP features on LFW3D and proposed method for comparison. Similarity metric learning (Sub-SML) [23] is adopted to boost face verification performance.

In order to discover how much the face verification performance could be improved by using proposed face frontalization method, we extract LBP on LFW-a [24] and HD-LBP on original LFW images and apply Sub-SML for comparison.

Results and Analysis. Table 1 shows the verification performance on LFW of different methods and Fig. 8(a) shows corresponding ROC curves. We first show the result of directly extracting LBP or HD-LBP with using Sub-SML, which achieves 83.92% and 88.78% respectively. With adding proposed method, we boost the performance to 88.82% and 91.50%, an improvement of 4.90% and 2.72% respectively. The improvement on feature HD-LBP is smaller because HD-LBP is already an excellent and expressive feature. These improvements come from our explicitly frontalization with natural, consistent results generated for directly texture comparison.

Table 1. Verification performance on LFW give by mean accuracy and standard error under image restricted, label-free outside data protocol.

Methods	Accuracy ($\bar{\mu} \pm S_E$)
LBP+Sub-SML [23]	0.8392 \pm 0.0065
HD-LBP+Sub-SML	0.8878 \pm 0.0046
LFW3D [16]+LBP+Sub-SML	0.8818 \pm 0.0047
OURS+LBP+Sub-SML	0.8882 \pm 0.0041
HPEN [15]+HD-LBP+Sub-SML	0.9152 \pm 0.0037
OURS+HD-LBP+Sub-SML	0.9150 \pm 0.0058

**Fig. 8.** (a) ROC curves on LFW under image restricted, label-free outside data protocol. (b) Mean faces by averaging corresponding multiple images of four subjects from LFW. First Row: Deep-Funneled [25], Second Row: LFW3D [16], Third Row: Proposed Method.

With same condition of using LBP with Sub-SML, proposed method outperforms LFW3D by 0.64% since we accurately estimate the self-occlusion region and fill it smoothly with keeping illumination consistence, resulting in less artifacts than LFW3D. Under the setting of using HD-LBP with Sub-SML, we achieve 91.50%, nearly the same performance as HPEN. It is noticed that HPEN utilizes 68 facial landmarks for shape fitting along with expression normalization. Bad face normalization result may occur due to un-precise 68 landmarks localization under large pose. Our method adopts five stable landmarks which are easier to detect even under large pose, indicating the simpleness and superiority of proposed method.

Qualitative Result. LFW3D [16] shows how well the frontalization method have preserved the texture of input identity by showing mean faces of several subjects. We follow this qualitative experiment and results are shown in Fig. 8(b).

It can be seen that the details around the eyes and mouth are better preserved and more consistent in our method, compared with the other two methods.

4.2 Face Identification Across Pose on MultiPIE

MultiPIE contains 754,204 images of 337 identities, where each identity has images captured under controlled environment with 15 poses and 20 illumination in four sessions during different periods, supporting development of algorithms for face recognition across pose, illumination and expression. A common setting for face recognition across pose, proposed in [5, 14], is used for evaluation. This setting adopts images with different poses with neutral illumination marked as ID 07. The first 200 identities in all the *four* sessions are used for training and remaining 137 identities for test. During test, one frontal image of each identity from the earliest session in the test set is selected as gallery. The remaining images from -45° to $+45^\circ$ except 0° are selected as probes. This setting evaluates recognition robustness affected by pose, as well as other real-world factors, such as appearance changes by glasses or mustache.

For all gallery images and each probe image, we make frontalization and invisible region filling is used when estimated yaw angle is larger than 13° , the same as operation on LFW. HD-LBP is adopted as feature extractor. For better comparison, we apply several classifiers, including PCA, LDA and LRA [26]. PCA and LDA are trained on frontalized images from the first 200 identities. LRA is directly trained on frontalized gallery images by mapping gallery faces to equidistant space targets which could enhance the discrimination between similar faces.

We compare our method with several pose normalization methods, including two 3D methods, Asthana11 [14] and HPEN [15], and three 2D methods, MDF [27], FIP [7] and MVP [9], the latter two are representative deep learning methods. Rank-1 identification rates are reported as results.

Table 2. Rank-1 identification rates (Percentage) on **MultiPIE** across pose. The first and the second highest performance are in **Bold**.

Methods	-45°	-30°	-15°	$+15^\circ$	$+30^\circ$	$+45^\circ$	Avg.
Asthana11 [14]	74.1	91	95.7	95.7	89.5	74.8	86.9
HPEN+PCA [15]	88.5	95.4	97.2	98.0	95.7	89.0	94.0
HPEN+LDA [15]	97.4	99.5	99.5	99.7	99.0	96.7	98.6
MDF [27]	93	98.7	99.7	99.7	98.3	93.6	97.2
FIP [7]+LDA	95.6	98.5	100	99.3	98.5	97.8	98.3
MVP [9]	93.4	100	100	100	99.3	95.6	98.1
OURS+PCA	89.3	97.3	97.3	97.7	94.7	86.0	93.7
OURS+LDA	98.0	99.3	99.3	99.0	99.3	96.3	98.5
OURS+LRA [26]	98.7	99.7	100	99.7	100	98.7	99.5

Results and Analysis. Table 2 presents recognition results on MultiPIE across pose. Asthana11 exploited 3D information learnt from 200 subjects of training set and achieved mean accuracy of about 87%. The shortage is that self-occlusion part and background were not filled. MDF [27] transformed a non-frontal face to a frontal one by face pixels rearrangement, using morphable displacement field learnt from 3D face models and achieved competitive result. Proposed method with LDA outperform above two methods possibly for the natural filling of self-occlusion region and background. Similar to the experiment on LFW, proposed method with PCA as well as LDA achieve very close results to HPEN, indicating the effectiveness of our accurate invisible region estimation and natural filling, by using just five stable landmarks. With applying LRA classifier, we further boost our performance to 99.5% and outperform other methods, especially in large angles ($\pm 45^\circ$). As we can see, FIP and MVP are two representative deep learning methods. They achieve competitive results with taking advantage of same pose distribution of training data and test data. In contrast, proposed method does not utilize any database-dependent information and would generalize well across continuous pose.

4.3 Further Discussion on Illumination Normalization

We apply light coefficients estimated from incomplete texture as global light representation to preserve lighting condition of input face, which inspires an idea of illumination normalization, by applying light coefficients of canonical lighting condition. Concretely, 20 lighting conditions exists in the bootstrap set marked as id 00–19 (shown in Fig. 5), among which id 07 represents canonical lighting condition. We set $x_j = 1$ ($j = 8$) and $x_j = 0$ ($j = 1 : 7, 9 : 19$) in Eq. 8 and get illumination normalization result. Previous illumination normalization methods, such as WA [28] and DCT [29] mainly focus on frontal face while our idea provide a simple, unify framework for illumination normalization after pose normalization.

We perform face identification experiments across pose and illumination variations on MultiPIE. Images of 249 ids from session *one*, covering 7 poses (-45° to $+45^\circ$) and 20 illumination are used. The first 100 ids are for training, and the remaining 149 ids for test. The frontal image under illumination marked as ID 07 of each identity in the test set is chosen as the gallery. The remaining images from -45° to $+45^\circ$ except 0° (illumination ID 07 excluded) are selected as probes. We examine the performance of just Pose Normalization (briefly denoted as **PN**) and both Pose and Illumination Normalization (briefly denoted as **PIN**). Feature extractor is LBP and classifiers PCA, LDA and LRA [26] are tested as previous experiments in Sect. 4.2.

Results and Analysis. From Table 3 we can see that, explicitly illumination normalization largely improve total performance towards all classifiers, from 45.4% to 67.8% in PCA, from 63.1% to 75.5% in LDA, from 77.0% to 86.5% in LRA, verifying the effectiveness of proposed idea in solving both pose and

Table 3. Rank-1 Identification Rates (Percentage) on **MultiPIE** Across Illumination. Recognition rate under one illumination condition is the averaged result of 6 possible poses. Pose Normalization is briefly denoted as **PN**. Pose and Illumination Normalization is briefly denoted as **PIN**.

Methods	00	01	02	03	04	05	06	08	09	10
PN+PCA	30.2	20.1	24.6	32.3	48.8	67.8	79.4	79.2	71.0	50.8
PIN+PCA	48.1	35.1	45.3	55.8	68.7	81.3	89.0	88.5	84.3	71.3
PN+LDA	58.7	37.9	46.9	57.3	71.0	84.9	91.7	93.5	87.4	78.3
PIN+LDA	65.1	47.4	58.7	71.4	82.0	92.4	96.9	96.4	92.6	82.0
PN+LRA	64.9	44.1	51.9	67.3	87.3	96.3	99.5	99.7	97.1	88.8
PIN+LRA	79.6	59.3	73.2	85.0	94.9	97.7	99.3	99.0	98.3	94.5
	11	12	13	14	15	16	17	18	19	Avg.
PN+PCA	34.1	26.5	20.3	43.4	51.1	57.5	51.2	43.4	31.4	45.4
PIN+PCA	57.7	44.0	32.0	64.3	70.9	80.8	71.0	63.1	46.9	67.8
PN+LDA	63.5	47.6	39.0	70.0	74.6	81.3	74.7	71.6	58.4	63.1
PIN+LDA	69.0	56.0	47.5	80.3	84.9	89.5	82.1	76.0	64.4	75.5
PN+LRA	67.8	51.7	46.4	81.1	88.3	94.2	89.8	82.9	64.4	77.0
PIN+LRA	83.3	70.4	59.5	91.2	95.1	97.8	95.0	90.8	80.5	86.5

illumination problem in such a simple way. We can observe that performance of some lighting condition are relatively low, *e.g.*, 01, 02, 12, 13 and also the improvement from **PN** to **PIN** under these conditions are relatively large. We have selected -45° for demonstration. The 5 lowest performance light conditions and 5 highest performance ones in **PN+LRA** are shown in Fig. 9. In pose normalization results (third row), large illumination variance (strong specular light or dark ambient light) exists in the former group and leads to uneven, unsmooth face texture, resulting in low performance. From another perspective, under former group, our illumination normalization can largely reduce the lighting difference between probes and galleries and thus boost the performance with large proportion. The illumination conditions of latter group are close to gallery and thus achieve higher performance.

4.4 Discussion and Limitations

The normalization process takes about 1.5s, running on a 2.8Ghz CPU with matlab code. The bottleneck part is face and background rendering, which takes about 0.8s and can be accelerated by C++.

In the process of invisible region filling, we use Quotient Image as a feature insensitive to illumination, satisfying face symmetry. The assumption of Quotient Image is Lambertian reflectance surface. When strong specular light occur, it can not model very well and would generate unnatural results. Also, it is hard to

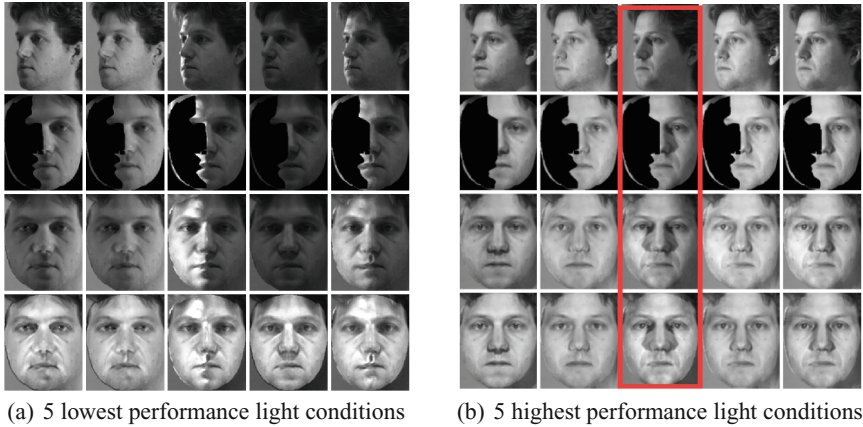


Fig. 9. Example results from various lighting of -45° from MultiPIE. First Row: Input images. Second Row: “Raw” frontalization results. Third Row: Pose normalization result. Fourth Row: Pose and Illumination Normalization results. (Color figure online)

eliminate the presence of cast shadow which leads to obvious artifacts when applying face symmetry (see the group in red box in Fig. 9(b)).

5 Conclusion

In this paper, considering the pose factor in unconstrained face recognition, we propose a continuous identity-preserving face normalization method which produces natural results in terms of illumination condition. With face borderline detection, the self-occlusion part is accurately detected and natural result is obtained by applying Quotient Image as a face symmetrical feature which is robust to illumination. We also provide a simple idea for illumination normalization in our framework. Our method achieve very competitive performance on LFW and MultiPIE datasets. With using only five stable landmarks and advantage of being database independent, our work is suitable for practical applications. In the future, we will focus on more sophisticated illumination modeling method to handle with strong specular light and cast shadow problem.

Acknowledgments. This work was partially sponsored by supported by the NSFC (National Natural Science Foundation of China) under Grant No. 61375031, No. 61573068, No. 61471048, and No.61273217, the Fundamental Research Funds for the Central Universities under Grant No. 2014ZD03-01, This work was also supported by Beijing Nova Program, CCF-Tencent Open Research Fund, and the Program for New Century Excellent Talents in University.

References

1. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **22**, 1090–1104 (2000)
2. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. *Image Vis. Comput.* **28**, 807–813 (2010)
3. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst (2007)
4. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. *arXiv preprint [arXiv:1502.04383](https://arxiv.org/abs/1502.04383)* (2015)
5. Li, A., Shan, S., Gao, W.: Coupled bias-variance tradeoff for cross-pose face recognition. *IEEE Trans. Image Process.* **21**, 305–315 (2012)
6. Chai, X., Shan, S., Chen, X., Gao, W.: Locally linear regression for pose-invariant face recognition. *IEEE Trans. Image Process.* **16**, 1716–1725 (2007)
7. Zhu, Z., Luo, P., Wang, X., Tang, X.: Deep learning identity-preserving face space. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 113–120. IEEE (2013)
8. Kan, M., Shan, S., Chang, H., Chen, X.: Stacked progressive auto-encoders (spae) for face recognition across poses. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1883–1890. IEEE (2014)
9. Zhu, Z., Luo, P., Wang, X., Tang, X.: Multi-view perceptron: a deep model for learning face identity and view representations. In: *Advances in Neural Information Processing Systems*, pp. 217–225 (2014)
10. Yim, J., Jung, H., Yoo, B., Choi, C., Park, D., Kim, J.: Rotating your face using multi-task deep neural network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 676–684 (2015)
11. Yi, D., Lei, Z., Li, S.: Towards pose robust face recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3539–3545 (2013)
12. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1063–1074 (2003)
13. Prabhu, U., Heo, J., Savvides, M.: Unconstrained pose-invariant face recognition using 3D generic elastic models. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 1952–1961 (2011)
14. Asthana, A., Marks, T.K., Jones, M.J., Tieu, K.H., Rohith, M.: Fully automatic pose-invariant face recognition via 3D pose normalization. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 937–944. IEEE (2011)
15. Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z.: High-fidelity pose and expression normalization for face recognition in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 787–796 (2015)
16. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4295–4304 (2015)
17. Shashua, A., Riklin-Raviv, T.: The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 129–139 (2001)
18. Ding, L., Ding, X., Fang, C.: Continuous pose normalization for pose-robust face recognition. *Sig. Process. Lett.* **19**, 721–724 (2012). IEEE

19. Foley, J.D.: *Computer Graphics: Principles and Practice*, vol. 12110. Addison-Wesley Professional, Reading (1996)
20. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. *ACM Trans. Graph. (TOG)* **22**, 313–318. ACM (2003)
21. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009*, pp. 296–301. IEEE (2009)
22. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3025–3032 (2013)
23. Cao, Q., Ying, Y., Li, P.: Similarity metric learning for face recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2408–2415 (2013)
24. Wolf, L., Hassner, T., Taigman, Y.: Similarity scores based on background samples. In: Zha, H., Taniguchi, R., Maybank, S. (eds.) *ACCV 2009*. LNCS, vol. 5995, pp. 88–97. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-12304-7_9](https://doi.org/10.1007/978-3-642-12304-7_9)
25. Huang, G., Mattar, M., Lee, H., Learned-Miller, E.G.: Learning to align from scratch. In: *Advances in Neural Information Processing Systems*, pp. 764–772 (2012)
26. Deng, W., Hu, J., Zhou, X., Guo, J.: Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning. *Pattern Recogn.* **47**, 3738–3749 (2014)
27. Li, S., Liu, X., Chai, X., Zhang, H., Lao, S., Shan, S.: Morphable displacement field based image matching for face recognition across pose. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012*. LNCS, vol. 7572, pp. 102–115. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-33718-5_8](https://doi.org/10.1007/978-3-642-33718-5_8)
28. Du, S., Ward, R.: Wavelet-based illumination normalization for face recognition. In: *IEEE International Conference on Image Processing, ICIP 2005*, vol. 2, p. II-954. IEEE (2005)
29. Chen, W., Er, M.J., Wu, S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **36**, 458–466 (2006)