



Lighting-aware face frontalization for unconstrained face recognition



Weihong Deng*, Jiani Hu, Zhongjun Wu, Jun Guo

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China

ARTICLE INFO

Article history:

Received 4 December 2016

Revised 25 February 2017

Accepted 20 March 2017

Available online 22 March 2017

Keywords:

Face frontalization

Pose normalization

Illumination normalization

Unconstrained face recognition

Labeled face in the wild learning

ABSTRACT

Face recognition under variable pose and lighting is still one of the most challenging problems, despite the great progress achieved in unconstrained face recognition in recent years. Pose variation is essentially a misalignment problem together with invisible region caused by self-occlusion. In this paper, we propose a lighting-aware face frontalization method that aims to generate both lighting-recovered and lighting-normalized frontalized images, based on only five fiducial landmarks. Basic frontalization is first performed by aligning a generic 3D face model into the input face and rendering it at frontal pose, with an accurate visible region estimation based on face borderline detection. Then we apply the illumination-invariant quotient image, estimated from the visible region, as a face symmetrical feature to fill the invisible region. Lighting-recovered face frontalization (LRFF) is conducted by rendering the estimated lighting on the invisible region. By adjusting the combination parameters, lighting-normalized face frontalization (LNFF) is performed by rendering the canonical lighting on the face. Although its simplicity, our LRFF method competes well with more sophisticated frontalization techniques, on the experiments of LFW database. Moreover, combined with our recently proposed LRA-based classifier, the LNFF based method outperforms the deep learning based methods by about 6% on the challenging experiment on Multiple PIE database under variable pose and lighting.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Face recognition is an active research area for many potential 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 affect the performance of face recognition system. Among these factors, pose variation is considered the most challenging one. An appropriate reduction of pose variation is largely beneficial to the following tasks such as feature extraction or facial attributes analysis. The major challenge of pose variation 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].

Face frontalization is a process of synthesizing frontal facing views of faces appearing in single unconstrained photo. It has been suggested that this process substantially boosts the performance of face recognition systems [5,6] by comparing the face in the

canonical (frontalized) view. Previous works mostly learn the (non-linear) frontalized transformation by applying regression techniques [7] or more complex deep neural network [8–10] on a large external multi-poses data set. Recent work [6] have demonstrated that a single 3D reference model is sufficient for face frontalization, which emphasizes on preserving the appearance of input face rather than accurately recovering its shape. 3D face shape estimation from unconstrained photos may be a harder problem than frontalization and can potentially introduce facial misalignments [6]. Moreover, face frontalization by a single 3D reference model avoids the dependence on any specific training data and can be generalized well in unconstrained setting by preserving the detailed texture of the specific face.

In this paper, we propose a practical face frontalization method which requires only five landmarks for alignment and is able to recover or normalize the lighting condition on the frontalized face. To the best of our knowledge, this lighting-aware face frontalization (LAFF) method is the first attempt to jointly handle the pose and lighting normalization using a generic 3D face model. The basic idea of LAFF procedure is illustrated in Fig. 1. There are three contributions of our work compared with the current techniques.

First, we propose a basic face formalization procedure using only five stable fiducial landmarks, which can easily integrated in the automatic recognition system. Specifically, a generic 3D face model is aligned to the input face image based on five landmarks,

* Corresponding author.

E-mail address: whdeng@bupt.edu.cn (W. Deng).

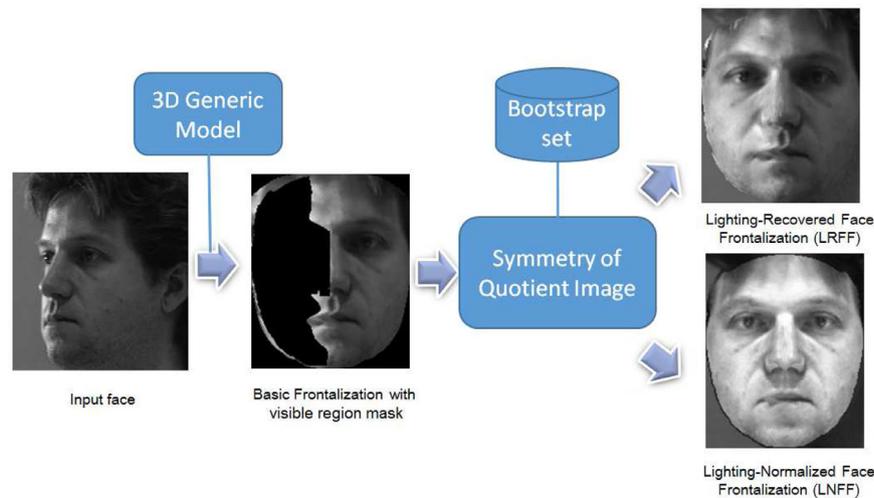


Fig. 1. Basic idea of Lighting-Aware Face Frontalization procedure. Leveraging the prior knowledge provided by a 3D generic model and bootstrap set of lighting, our method can perform both the lighting-recovered face frontalization (LRFF) and the lighting-normalized face frontalization (LNFF).

and a robust borderline detection algorithm is applied to accurately estimate the visible region. Basic frontalization is obtained by rendering the appearance-assigned 3D mesh at frontal pose with the visible region mask (Section 3.1). Empirical results show that our method produces similar visual effect and recognition accuracy as the previous works, e.g., [5,11], that are based on dozens of landmarks.

Second, we apply the symmetry of the quotient image to explicitly estimate the lighting intensity of the invisible (self-occluded) facial part, rather than simply consider the facial appearance symmetry. To achieve this goal, we apply facial symmetry to the estimated Quotient Image [12], which is lighting invariant, from the visible part of frontalized face. After estimating parameters accompanying with the quotient image, lighting-recovered face frontalization (LRFF) is performed by rendering the estimated lighting on the self-occluded part. By adjusting the combination parameters, lighting-normalized face frontalization (LNFF) is performed by rendering the canonical lighting on the face.

Third, we demonstrate the effectiveness of our method by both the visual effects and the verification/recognition performance on the large-scale face verification and recognition experiments on LFW [3] and MultiPIE [2]. Although its simplicity, our LRFF method achieves comparable, even better, performance than the state-of-the-art methods that rely on dozens of landmarks and sophisticated 3D model fitting on the LFW benchmark. Moreover, combined with our recently proposed LRA-based classifier [13], the LNFF method outperforms the deep learning based face normalization methods by about 6% on the challenging experiment on Multiple PIE database under variable poses and lightings.

2. Related works

In general, there are two families of the pose-invariant face recognition methods: 2D-based and 3D-based. For the comprehensive survey, one can refer to [4,14]. 2D-based methods handle pose variations by the 2D image (patch) mapping across different poses, which can be roughly divided into two categories as followed [15]: shallow 2D mapping and deep 2D mapping.

Shallow 2D mapping: Local linear regression (LLR) [16] learns appearance transformation between different poses by the key assumption that the manifold structure of a local patch stays the same across poses. Heo and Savvides [17] use 2D affine warps of the view-based AAM to approximately map non-frontal face from to frontal face, while Gao et al. [18] use a single AAM to fit non-

frontal faces. Li et al. in [7] represents a test image using some bases or exemplars and the coefficients can be regarded as one kind of pose-invariant features. Du and Ward [19] propose to use a set of prototype non-frontal face images that are in the same pose as the input non-frontal face, which is recently extended by the sparse representation [20,21]. The performance depends heavily on correlation between test subject and the external training data. Ho and Chellappa [22] proposed to learn a globally optimal set of local warps for frontal face synthesis by considering the consistency at the overlapped pixels between two nearby patches. However, these methods are limited for the incapability of capturing 3D rotations as well as solving self-occlusion problem with using 2D warping.

Deep 2D mapping: Deep learning approaches become popular due to their premier accuracy on recognition with massive external training data. Deep Auto-Encoder (DAE) method [23] learns pose-robust features by modeling the complex non-linear transformation from the nonfrontal face images to frontal ones through deep auto-encoder, which directly converts the non-frontal face images to the frontal ones. Stacked progressive auto-encoders (SPAEE) method [8] learns pose-robust features by modeling the complex non-linear transformation from the non-frontal face images to frontal ones through a deep network in a progressive way. The face identity-preserving (FIP) features [9] are learned by a deep network that combines the feature extraction layers and the reconstruction layers. The former layers encode a face image into the pose-invariant FIP features, while the latter transform them to an image in the canonical view. RL+LDA method [9] further improves the performance by applying local descriptors and LDA on the frontal reconstructed images. Recently, a Multi-View Perceptron (MVP) [24] is proposed to untangle the identity and pose by using random hidden neurons. Controlled Pose Face (CPF) method [10] is a recent work which can rotate an arbitrary pose and illumination image to a target-pose face image by multi-task deep neural network.

3D-based methods handle pose variations based on the prior knowledge provided by a reference 3D face model or a deformable model with shape and illumination parameters. 3D methods are divided into three categories as followed [15]: modeling fitting, pose synthesis, and pose normalization.

Modeling fitting: 3D Morphable Model (3DMM) [25] is a powerful 3D representation for human face which fits parameters of 3D shape, pose and illumination and use them for recognition. Breuer et al. [26] present a method for automatically fitting the 3D Morphable Model, but it has a high failure rate and high computational

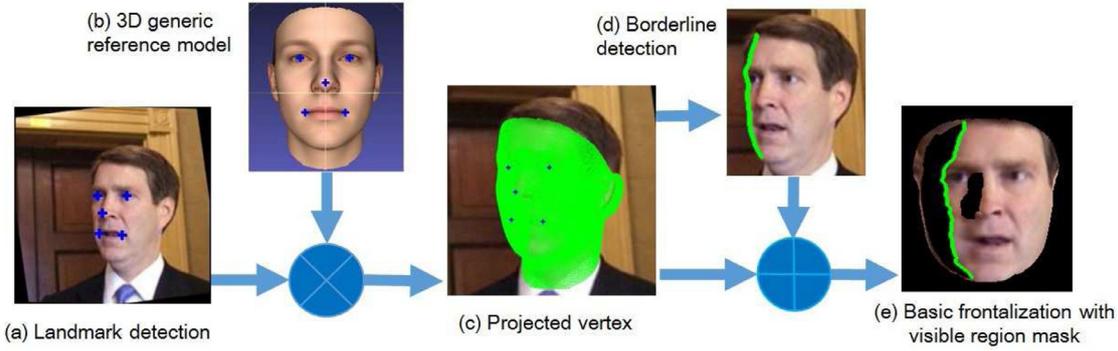


Fig. 2. Procedures of the proposed basic face frontalization with visibility detection. Only 5 landmarks are required for face frontalization with the 3D generic reference model. Borderline detection is conducted to achieve accurate visibility detection result. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

cost. Aldrian et al. [27] present an efficient framework to inverse render faces with a 3D Morphable Model (3DMM) by decomposing the image formation process into geometric and photometric parts. The PCA subspace used in 3DMM may not enough to accurately characterize the textures of test faces. Besides 3DMM, several 3D shape based methods are proposed to rotate the non-frontal face to the frontal one. Recently, Jo et al. [28] propose a person-specific 3D facial reconstruction method that is person-specific by combining the simplified 3DMM and the SfM (Structure from Motion) methods to improve reconstruction quality.

Pose synthesis: Generic Elastic Model (GEM) [29] is an efficient 3D face modeling method, which estimates 3D shape by assigning generic face depth information directly to probe 2D images. Virtual face images under arbitrary poses can be generated using 3D models generated from gallery images. Probe face is matched to the virtual images with similar pose to the probe. 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 by eliminating the pose variations. Asthana et al. [11] synthesize 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. Abiantun et al. [30] propose to recover the un-occluded pixels to a PCA model with sparse coefficient that is trained by the frontal faces. Li et al. [31] proposed the generation of the template displacement fields using images synthesized by a set of 3D face models. The pixel-wise correspondence between the synthesized images can be easily inferred via the 3D model vertices, therefore this approach implicitly utilizes 3D facial shape priors for pose normalization. High-fidelity Pose and Expression Normalization (HPEN) [5] fits 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 [6] employs a generic 3D face model to “frontalize” non-frontal images and synthesizes the occlusion part based on face symmetry with occlusion degree estimation.

Previous pose normalization methods have achieved promising results on the unconstrained images by preserving the texture information, but they are limited in the filling of self-occluded parts. Asthana et al. [11] leave the invisible region unfilled and can not produce a consistent result. Ding et al. [32] use mirrored pixels which would produce incoherent face texture especially when the illumination conditions on both sides of face are largely different. LFW3D [6] designs a “soft symmetry” that combines mirrored pixels with occlusion degree estimation. None of these methods attempts to recover the lighting conditions of the occluded face. In

contrast, our LAFF method aims to recover the lighting condition of the invisible (self-occluded) parts of the face, which is demonstrated to improve unconstrained face recognition performance.

3. Lighting-aware face frontalization (LAFF)

Leveraging the prior knowledge provided by a 3D generic model and bootstrap set of lighting, our method can perform lighting-recovered face frontalization (LRFF) and lighting-normalized face frontalization (LNFF). This section presents the detailed procedure of our method.

3.1. Basic frontalization and visibility detection

Our basic frontalization assumes that human face is a rigid structure, and thus sparse correspondence (five points correspondence) is able to represent dense correspondence of face vertices. Given a (non-frontal) facial image, five stable facial landmarks are located automatically or manually (see the blue ‘+’ in Fig. 2(a)). The five fiducial landmarks in the 3D generic reference model (see Fig. 2(b)) 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:

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

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. With projection matrix T , all vertices of reference model are projected onto the query image (see Fig. 2(c)) 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 a basic frontalization result.

The Z-buffer method [33] is commonly used to detect the visible region of the 3D model. Unfortunately, since our 3D face model is generic, its borderline may not be fully consistent with specific face surface, as shown in Fig. 3(b), resulting in an inaccurate visible region in Fig 3(c). To address this limitation, we design a joint optimization function to detect the borderline by considering both the gradient magnitude and the similarity to the (Z-buffer) borderline of 3D model. The curve is defined by pixel coordinates (x_i, y_i) , where y_i is the row index. The total optimization problem is defined as

$$\max_{\{x_i, y_i\}} \sum_i g(x_i, y_i) - \sum_i d(x_i, x_{i-1}) + \lambda \sum_i s(x_i, y_i) \quad (2)$$

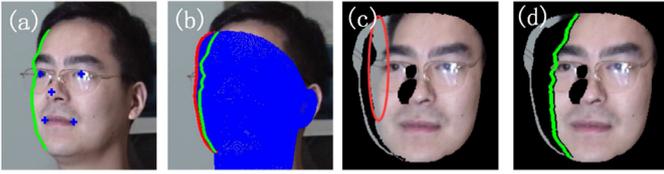


Fig. 3. Comparison of visibility detection from Z-Buffer and our method. (a) Example input image from MultiPIE. (b) Aligned 3D model on input image. The texture on the left side of true borderline (green line) in Z-Buffer method is considered to be visible while actually not. (c) Results of visibility detection of Z-Buffer. Black pixels indicates invisible. Red ellipse marks the unwanted background texture. (d) Result of our visibility detection method is more accurate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where g is the image gradient magnitude and d constrains x_{i-1} and x_i to be within one pixel, and s is the similarity between the projected borderline of 3D model and the found curve. The term g is used to catch the strong edge (large gradient magnitude) along the borderline of the face and background, where the gradient magnitude is defined as

$$g(\mathbf{I}) = \left| \frac{\partial}{\partial x} \mathbf{I} \right| + \left| \frac{\partial}{\partial y} \mathbf{I} \right| \quad (3)$$

\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. For each pixel in search region, we calculate its tangential direction through its vertical and horizontal gradient (the blue arrow in Fig. 4), represented as $\vec{T}_i(x, y)$. For the projected 3D model borderline, the tangential direction of row y can also be calculated (the purple arrow in Fig. 4), represented as $\vec{T}_r(y)$. The similarity of direction $\vec{T}_i(x, y)$ at pixel (x, y) to projected 3D model borderline is calculated as cosine similarity,

$$s(x, y) = \frac{\vec{T}_i(x, y) \cdot \vec{T}_r(y)}{\|\vec{T}_i(x, y)\| \|\vec{T}_r(y)\|} \quad (4)$$

The parameter 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.

The joint optimization (2) is solved by dynamic programming in a search region around the Z-buffer borderline of the aligned 3D model. Examples of the found curves are shown in Fig. 4(b). The found face contour is back-transformed to the frontal 3D reference model through matrix \vec{T}^{-1} and we can get a rather accurate visible region mask as our “raw” frontalization result (see Figs. 2(e) and 3(d)). Note that the visibility of nose region is estimated using Z-Buffer method [33] in our implementation.

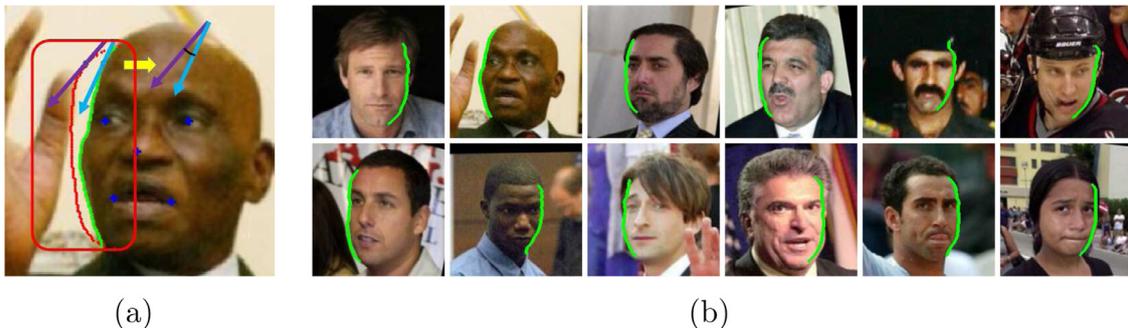


Fig. 4. Similarity of found curve and projected borderline. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

3.2. Quotient image symmetry based recovery

The face symmetry is the common prior to recover the invisible region for face frontalization [6]. Unfortunately, in the cases under uniform lighting condition, the face appearance is asymmetric and the recovered invisible regions may contain unrealistic lighting effects. To avoid the lighting artifacts, we decide to apply the symmetry prior to the quotient image (QI) that is invariant to lighting changes, rather than the appearance image. To achieve this goal, we derives quotient image [12] by the surface reflectance ratio of an (input) face against another (reconstructed) face, from the partial frontalized face and recovers the lighting condition for the full face.

In light of the quotient image technique [12], human face is assumed to be a 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. The classical quotient image technique [12] introduced the concept *Ideal Class of Object*, i.e., objects that have same shape but differ in surface albedo are 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 small bootstrap set containing N ($N=12$ in our experiments) identities under M ($M=20$ in our experiments) 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 $Y(u, v)$. In our experiments, the 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 [12]. Example bootstrap set images (gray level) from one identity is shown in Fig. 5.

The basic frontalized face with visible region mask is used to estimate Quotient Image and lighting condition. We mask all the images in the bootstrap set using the visible mask of the input image and estimate Quotient Image on valid texture as well as



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.

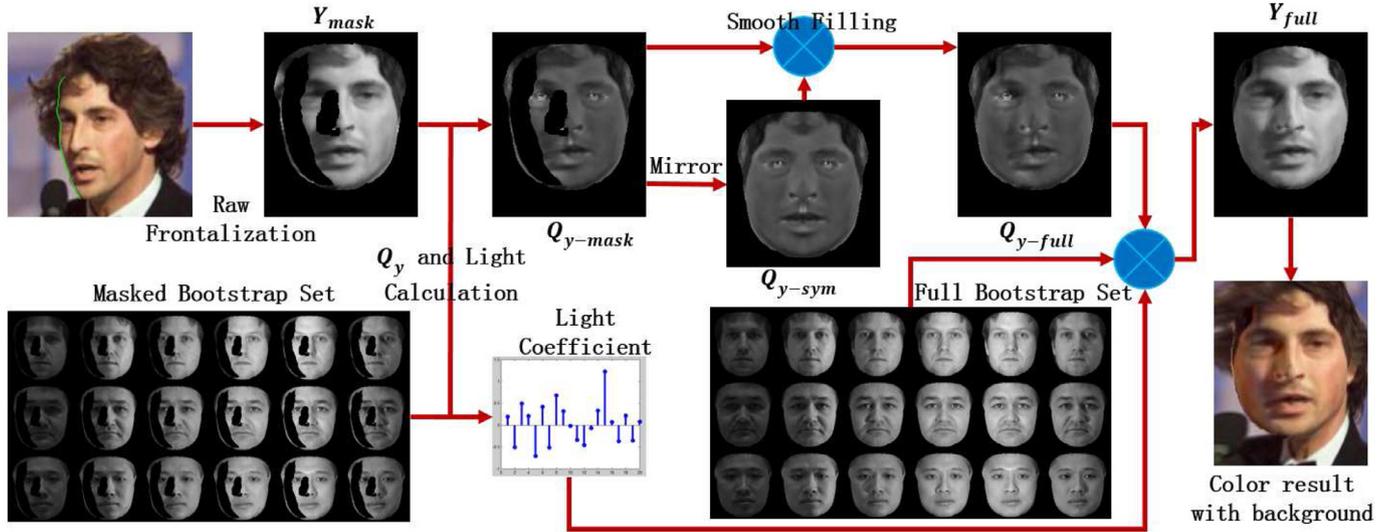


Fig. 6. Process of self-occlusion part filling based on the symmetry of the estimated quotient image. The images from 3 identities of 6 lightings in the bootstrap set are shown for convenience. There are actually 12 identities and 20 lightings.

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 [34] mentioned in [5] 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 to *estimate lighting conditions from incomplete valid texture and use it as global representation*. With adding background texture using affine transformation in [5], a complete frontalization is generated. Fig. 6 demonstrates process of self-occlusion part filling, which is also summarized in the following algorithm block. For the color image, we first fill the invisible region in Y channels using Quotient Image and combine directly symmetrical UV channels texture to get the final RGB result.

3.3. Lighting recovered/normalized face frontalization

When a facial image is presented, LAFF first aligns the 3D generic model to the input face according to the 5 feature points, and refines the borderline to obtain accurate visible face parts. Then, the facial quotient image and corresponding lighting coefficients are estimated from the visible face region, based on which

the lighting condition of occluded parts are recovered by the symmetry of the facial quotient image. The result frontalized image, which is named Lighting Recovered Frontalized face, is natural for visual inspection and facilitates further image processing and recognition.

However, lighting recovered frontalized images still display diverse illumination variation, which is not optimal for the recognition purpose. Fortunately, the quotient image technique, which has been integrated in LAFF, provides a natural manner to address the illumination variation 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 [35] and DCT [36] mainly focus on frontal face while our idea provide a simple, unify framework for illumination normalization after pose normalization (Algorithm 1).

4. Experiments and results

In this section, we first visually inspect the results of the proposed LRFF/LNFF methods and then evaluate their performance of on LFW and MultiPIE databases for face verification and face identification task respectively.

4.1. Qualitative visual results

Fig. 7(a) shows the front-facing new views of Labeled faces in the Wild images, where we compare our LAFF results with the most relevant method in [6] that has been release as the LFW3D

Algorithm 1 Invisible region filling.**Require:** “Raw” frontalization result, bootstrap set images**Ensure:** Full frontalization result

- 1: Mask bootstrap set images with same mask of basic 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.



(a)



(b)

Fig. 7. (a) Example frontalization results from LFW database. First Row: Input images. Second Row: Results of LFW3D [6]. Third Row: Results of Proposed LRFF Method. Our results keep illumination consistence and produce less artifacts by accurate borderline detection and smooth filling. (b) Mean faces by averaging corresponding multiple images of four subjects from LFW. First Row: Deep-Funneled [37], Second Row: LFW3D [6], Third Row: Our proposed LRFF Method.

dataset. One can see from the figures that both our LRFF method and LFW3D can preserve the texture of input, but LRFF show additional advantage of lighting consistence on the frontalized face. In general, the frontalized images by LRFF seem more similar to real frontal faces, especially when there is uneven lighting on the face.

To illustrate a more general result, Fig. 7(b) further shows the mean faces with different alignment methods. Average faces from the 31 David Beckham, 41 Laura Bush, 236 Colin Powell, and 530 George W. Bush images in the LFW set. On the average faces of LFW3D and LRFF, Wrinkles on the forehead of George W Bush in our result are faintly visible. These were preserved despite having been averaged from many images captured under varying condi-

tions. 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. In addition, one should note that the LFW3D method uses 48 facial landmarks for alignment, but it is surprising that LRFF method uses only 5 landmarks to achieve similar accuracy.

Fig. 8 shows the example images of MPIE database across different poses and lightings, and one can see from the figures that the original images have extremely large intra-class variations, which are much larger than those in other unconstrained data sets like LFW. The lighting recovered face frontalization (LRFF) images display much smaller variations, and all the images are transferred to the frontal face with common lighting variations. After lighting normalized face frontalization (LNFF), the intra-class variations on both poses and lightings are almost eliminated, and at the same time, the inter-class difference between faces are largely preserved. One can expect that the difficulty on recognizing LNFF images would be reduced significantly (Algorithm 2).

Algorithm 2 Lighting-aware face frontalization.**Require:** A non-frontal input face and the bootstrap set of images**Ensure:** Lighting-recovered/normalized frontalized face

- 1: Locate the 5 feature points, i.e. two eye centers, the nose tip, and two mouth corners by off-the-shelf face alignment method or manual labeling.
- 2: Align a 3D generic model to the input face according to the 5 feature points according to Eq. (1).
- 3: Optimize the criterion (2) to seek the face borderline and then back-transform it to the frontal 3D reference model; Apply Z-buffer method to find the visible nose region; Combine two results to obtain the visible region mask.
- 4: Generate lighting-recovered frontalized face (LRFF) by the invisible region filling according to Algorithm 1; Generate Lighting-normalized frontalized face (LNFF) by the invisible region filling according to Algorithm 1 with $x_j = 1 (j = 8)$ and $x_j = 0 (j = 1 : 7, 9 : 19)$ in Eq. (8);

4.2. 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 13,233 face images of 5749 persons collected from Internet with large variations including pose, age, illumination, expression, resolution, etc. We report our results following the “View2” 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 [38] as 3D reference model and frontal, multiple illumination facial images from MultiPIE [2] as bootstrap set images in Quotient Image.

We aim to evaluate the improvement of face recognition performance by using our frontalized faces. Thus, rather than using up-to-date learning methods that may overlay the contribution of frontalization, we first use the L2 distance between basic local descriptors, such as LPB, TPLBP, and FPLBP. For an input image, we first predict the yaw angle based on the 5 feature points and frontalized the face with invisible region filling when the estimated yaw angle is larger than 13° . Results are compared to those reported on the Deep-funneled images, LFW-a and LFW-3D collections. Table 1 enumerates the comparative results. Evidently, both LFW3d and LRFF, which use a generic 3D reference model, outperform the conventional affine-transformation base face alignment by a large margin. This suggest that 3D prior knowledge can play important role in unconstrained face alignment. Although applying



Fig. 8. Example frontalization results from MPIE database. First row: original images of four persons. Second row: resulting images of lighting recovered face frontalization (LRFF). Third row: resulting images of lighting normalized face frontalization (LNFF).

Table 1

Local descriptors verification results on the LFW benchmark. The best accuracy of each feature is highlighted in bold.

Methods	Funneled [37]	LFW-a [39]	LFW3D [6]	LRFF (Appearance)	LRFF (QI)
LPB	0.6767	0.6824	0.7465 ± 0.0053	0.7395 ± 0.0065	0.7565 ± 0.0055
TPLBP	0.6875	0.6926	0.7502 ± 0.0055	0.7433 ± 0.0057	0.7622 ± 0.0053
FPLBP	0.6865	0.6818	0.7265 ± 0.0143	0.7231 ± 0.0126	0.7335 ± 0.0153

Table 2

Verification performance on LFW given by mean accuracy and standard error under image restricted, label-free outside data protocol. The best accuracy of each feature is highlighted in bold.

Methods	Accuracy ($\bar{\mu} \pm S_{\epsilon}$)
LFW-a + LBP + Sub-SML [40]	0.8392 ± 0.0065
LFW3D [6] + LBP + Sub-SML	0.8818 ± 0.0047
LRFF + LBP + Sub-SML	0.8882 ± 0.0041
HD-LBP + Sub-SML	0.8878 ± 0.0046
HPEN [5] + HD-LBP + Sub-SML	0.9152 ± 0.0037
LRFF + HD-LBP + Sub-SML	0.9150 ± 0.0058

much simpler alignment and fitting procedure, our LRFF-based face is slightly better than the LFW3D-based face by about 1–2%. More importantly, To validate the superiority of the quotient image symmetry based recovery, we also implement LRFF with appearance symmetry. One can see from the Table 1 that QI-symmetry based recovery improves the accuracy by 1–3% compared to the commonly used appearance symmetry. The lighting-recovered faces are more suitable for recognition than the appearance symmetry faces.

Then, we evaluate the effectiveness of our LRFF-face by learning metric on the local face descriptors. Specially, We extract LBP on the LFW-a, LFW3D, and LRFF faces, and apply similarity metric learning (Sub-SML) [40] to boost the face verification performance. Table 2 shows the verification performance on LFW of different methods and Fig. 9(a) shows corresponding ROC curves. With same Sub-SML learned metric on the LBP descriptor with , our LRFF outperforms LFW3D by 0.64% because we accurately estimate the self-occlusion region and fill it smoothly with keeping illumination consistence, resulting in less artifacts than LFW3D.

Finally, we extract high-dimensional LBP (HD-LBP) [41] face descriptor on LFW-a, HPEN and LRFF-face for comparison. The images released by LFW3D are 90°90 pixels only containing face region,

which are not suitable for HD-LBP extraction. Table 2 shows the verification performance on LFW of different methods and Fig. 9(b) shows corresponding ROC curves. Compared with the basic LBP descriptor, the improvement on HD-LBP feature is smaller because HD-LBP is already an excellent and expressive feature that overlay the contribution of frontalization. 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 only five stable landmarks which are easier to detect even under large pose, indicating the simpleness and superiority of proposed method.

4.3. Face recognition across poses and lightings on MPIE

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. To fully evaluate our lighting-aware face frontalization method, our experiment is conducted under the **Setting-III** which has been widely used [7,9,10] to evaluate the robustness of recognition algorithms under variable poses and lightings. This setting is more realistic than the first two settings with only pose variation. Specifically, Setting-III adopts images in session one for training and test, which has 249 identities. Images from -45° to $+45^{\circ}$ (seven poses) under 20 illuminations (marked as ID 00–19) are used. As listed in Table 4, previous studies used all the images of first 100 identities for model training (LRFF and LNFF do not use these external training data), and the images of the remaining 149 identities for test. During the test, the frontal image under illumination marked as ID 07 of each

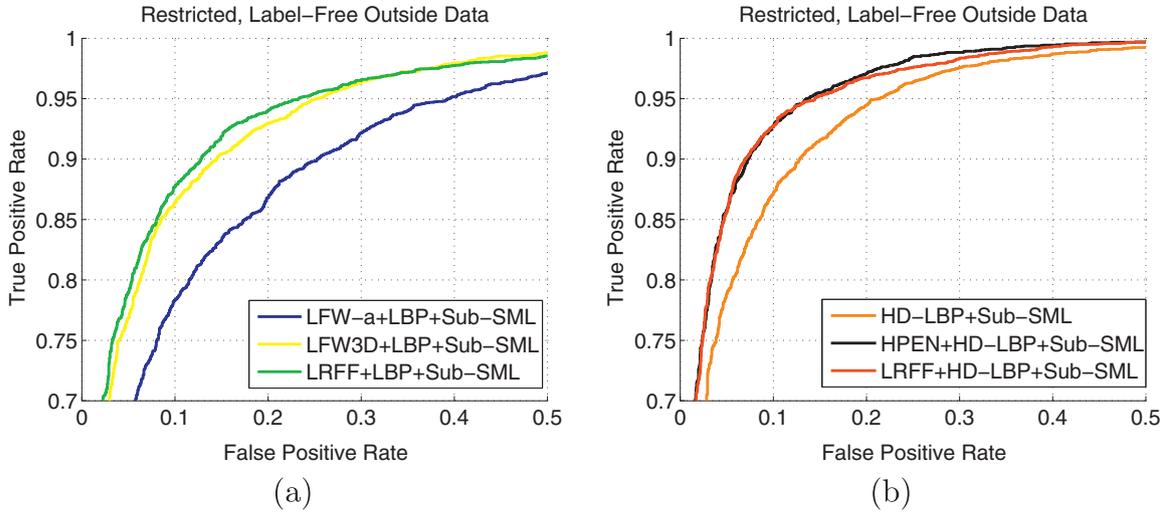


Fig. 9. ROC curves on LFW under image restricted, label-free outside data protocol.

Table 3

Comparative recognition rates (percentage) on MultiPIE under variable poses and lightings. The recognition rate under a specific illumination condition is the averaged result of 6 tested poses. Lighting recovered face frontalization is briefly denoted as **LRFF**. Lighting normalized face frontalization is briefly denoted as **LNFF**.

Lighting ID	Eigenfaces		Fisherfaces		LRA	
	LRFF	LNFF	LRFF	LNFF	LRFF	LNFF
00	30.2	48.1 (↑ 59%)	58.7	65.1 (↑ 11%)	64.9	79.6 (↑ 23%)
01	20.1	35.1 (↑ 75%)	37.9	47.4 (↑ 25%)	44.1	59.3 (↑ 34%)
02	24.6	45.3 (↑ 84%)	46.9	58.7 (↑ 25%)	51.9	73.2 (↑ 41%)
03	32.3	55.8 (↑ 73%)	57.3	71.4 (↑ 25%)	67.3	85.0 (↑ 26%)
04	48.8	68.7 (↑ 41%)	71.0	82.0 (↑ 15%)	87.3	94.9 (↑ 9%)
05	67.8	81.3 (↑ 20%)	84.9	92.4 (↑ 9%)	96.3	97.7 (↑ 1%)
06	79.4	89.0 (↑ 12%)	91.7	96.9 (↑ 6%)	99.5	99.3 (↑ 0%)
08	79.2	88.5 (↑ 12%)	93.5	96.4 (↑ 3%)	99.7	99.0 (↑ -1%)
09	71.0	84.3 (↑ 19%)	87.4	92.6 (↑ 6%)	97.1	98.3 (↑ 1%)
10	50.8	71.3 (↑ 40%)	78.3	82.0 (↑ 5%)	88.8	94.5 (↑ 6%)
11	34.1	57.7 (↑ 69%)	63.5	69.0 (↑ 9%)	67.8	83.3 (↑ 23%)
12	26.5	44.0 (↑ 66%)	47.6	56.0 (↑ 18%)	51.7	70.4 (↑ 36%)
13	20.3	32.0 (↑ 58%)	39.0	47.5 (↑ 22%)	46.4	59.5 (↑ 28%)
14	43.4	64.3 (↑ 48%)	70.0	80.3 (↑ 15%)	81.1	91.2 (↑ 12%)
15	51.1	70.9 (↑ 39%)	74.6	84.9 (↑ 14%)	88.3	95.1 (↑ 8%)
16	57.5	80.8 (↑ 41%)	81.3	89.5 (↑ 10%)	94.2	97.8 (↑ 4%)
17	51.2	71.0 (↑ 39%)	74.7	82.1 (↑ 10%)	89.8	95.0 (↑ 6%)
18	43.4	63.1 (↑ 45%)	71.6	76.0 (↑ 6%)	82.9	90.8 (↑ 10%)
19	31.4	46.9 (↑ 49%)	58.4	64.4 (↑ 10%)	64.4	80.5 (↑ 25%)
Avg.	45.4	67.8 (↑ 49%)	63.1	75.5 (↑ 20%)	77.0	86.5 (↑ 12%)

Table 4

Comparison of the number of training images used in our experiment. Compared with previous methods, LAFF uses a much smaller bootstrap set for reference.

Methods	Training images
Li [7]	100 identities × 7 poses × 20 illuminations, totally 14,000 images
RL [9]+LDA CPF [10]	
LAFF	12 identities × 1 frontal pose × 20 illuminations, Totally 240 images
Bootstrap set	

identity in the test set is chosen as the gallery. The remaining images from -45° to $+45^\circ$ except 0° , excluded the illumination ID 07, are selected as probes. All images that we selected were converted to gray scale. In the bootstrap set of quotient image, 20 lighting conditions exists in the bootstrap set marked as id 00–19 (shown in Fig. 5), among which id 07 represents canonical light-

ing condition. We set $x_j = 1$ ($j = 8$) and $x_j = 0$ ($j = 1 : 7, 9 : 19$) in Eq. (8) and get LNFF result.

The frontalized facial images are first represented by LBP descriptor followed by linear regression analysis (LRA) [13] for recognition. For the comparison purpose, we also implemented the Eigenfaces (PCA) and Fisherfaces (LDA). LRA is a single sample based face recognition method trained with the gallery set with a single sample per person. For LDA, we use the external training set of 100 persons with 140 images per person for training. Table 3 compares the performance of LRFF and LNFF under each lighting condition using the three well-established recognition methods. The recognition rate under one lighting condition is the averaged result of 6 tested poses. In general, our LRA method performs better than the Fisherfaces method followed by the Eigenfaces method. The LRA method performs better than other methods because the LRA take full advantage of the discriminative information contained in the gallery images.

Table 3 also suggests that the LRFF method is not sufficient to address the variable lighting condition in MPIE database. No mat-

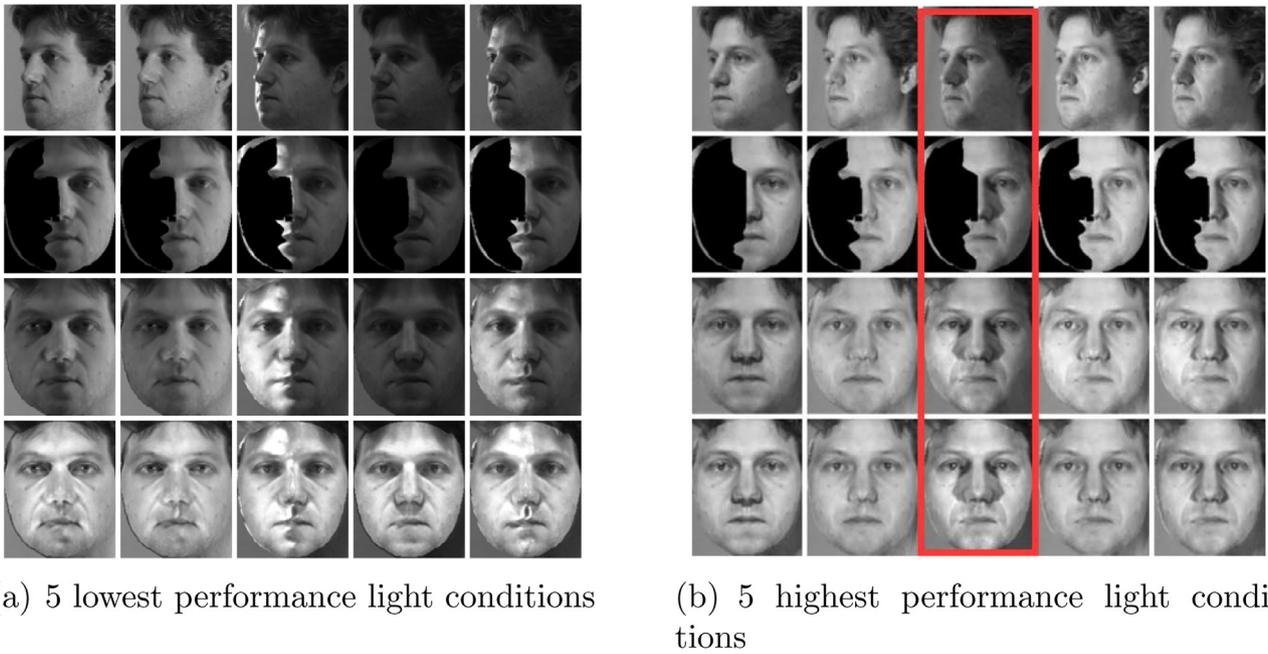


Fig. 10. Example results from various lighting of -45° from MultiPIE. First Row: Input images. Second Row: “Raw” frontalization results. Third Row: Lighting-recovered frontalized faces. Fourth Row: Lighting-normalized frontalized faces. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Table 5

Average recognition rate (in percentage) under different lighting conditions on the MPIE **Setting-III**. The best performance are in Bold.

Methods	00	01	02	03	04	05	06	08	09	10
Li [7]	51.5	49.2	55.7	62.7	79.5	88.3	97.5	97.7	91.0	79.0
RL+LDA [9]	72.8	75.8	75.8	75.7	75.7	75.7	75.7	75.7	75.7	75.7
CPF [10]	59.7	70.6	76.3	79.1	85.1	89.4	91.3	92.3	90.6	86.5
LNFF+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.
Li [7]	64.8	54.3	47.7	67.3	67.7	75.5	69.5	67.3	50.8	69.3
RL+LDA [9]	75.7	75.7	75.7	73.4	73.4	73.4	73.4	72.9	72.9	74.7
CPF [10]	81.2	77.5	72.8	82.3	84.2	86.5	85.9	82.9	59.2	80.7
LNFF+LRA	83.3	70.4	59.5	91.2	95.1	97.8	95.0	90.8	80.5	86.5

ter which classifier is applied, the illumination normalization of LNFF largely improves average performance, from 45.4% to 67.8% in Eigenfaces, from 63.1% to 75.5% in Fisherfaces, from 77.0% to 86.5% in LRA, clearly suggesting the effectiveness of proposed idea in solving both pose and illumination problem in such an unified manner. For the Eigenfaces method that simply matches two images, the relative improvement from LRFF to LNFF is as large as 49%, i.e. $[(67.8 - 45.4)/45.4] \times 100\%$. Moreover, it is surprising that, with the LNFF images, even the conventional Eigenfaces approach achieves a 67.8% recognition rate on this challenging experiments. This clearly indicates that the LNFF method is effective to reduce unconstrained pose and lighting variations simultaneously.

We can observe that performance of some lighting conditions are relatively low, e.g., 01, 02, 12, 13 and also the improvement from LRFF to LNFF under these conditions are relatively large. We have selected -45° for demonstration. The 5 lowest performance light conditions and 5 highest performance ones in LRFF+LRA are shown in Fig. 10. 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 boosts the performance with large proportion. The illumination

Table 6

Average recognition rate (in percentage) under different poses on the MPIE **Setting-III**. The best performance are in Bold.

Methods	-45°	-30°	-15°	$+15^\circ$	$+30^\circ$	$+45^\circ$	Avg.
Li [7]	63.5	69.3	79.7	75.6	71.6	54.6	69.3
RL [9]+LDA	67.1	74.6	86.1	83.3	75.3	61.8	74.7
CPF [10]	73.0	81.7	89.4	89.5	80.4	70.3	80.7
LNFF+LRA	77.2	87.7	94.9	94.8	88.1	76.4	86.5

conditions of latter group are close to gallery and thus achieve higher performance.

Then, we compare our best LNFF+LRA method to three existing state-of-the-art pose and illumination invariant methods. (1) Li [7] represents a test face as a linear combination of training images, and utilizes the regularized linear regression coefficients as features for face recognition. (2) RL+LDA [9] first reconstructs the frontal-view face images using FIP features extracted from an image under any pose and illumination, and then applies LDA to further enhance class separation. (3) CPF [10] is a recent work which can rotate an arbitrary pose and illumination image to a target-pose face image by multi-task deep neural network.

The comparative accuracy under different lighting and pose conditions is enumerated in Tables 5 and 6 respectively. The over-

all performance of LNFF+LRA is about 6% better than that of the state-of-the-art methods [10]. The significant higher accuracy of LNFF clearly shows the superiority of the 3D generic model based method for face frontalization, by which the identity related fine-texture can be preserved. Combination of LNFF images and the discriminative power of LRA achieves state-of-the-art performance on MPIE database. CPF [10] and RL+LDA [9] attempt to learn an unified deep neural network for pose-invariant feature extraction, but our results show they may not be the premier solution even given a large representative training set to the test condition.

Finally, it is worth mentioning that 1) our LNFF+LRA method doesn't use any non-frontal images from MPIE database to learn the pose-invariant features. LNFF just needs a few frontal images under different illumination conditions as bootstrap set to learn illumination information (see Table 4 for comparison). 2) Given the LNFF images, the performance can be further improved by feature/metric learning methods, as our results are reported by the basic LBP face descriptor. We would like to distribute the LNFF images for the research purpose after the publication of this paper.

4.4. Discussion and limitations

The normalization process takes about 1.5 s, running on a 2.8 Ghz CPU with matlab code. The bottleneck part is face and background rendering, which takes about 0.8 s 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 eliminate the presence of cast shadow which leads to obvious artifacts when applying face symmetry (see the group in red box in Fig. 10(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 supported by the National Natural Science Foundation of China (No. 61375031, No. 61573068, and No. 61471048), the Fundamental Research Funds for the Central Universities under Grant No. 2014ZD03-01, This work was also supported by the Program for New Century Excellent Talents in University, Beijing Nova Program under Grant No. Z161100004916088.

References

- [1] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The feret evaluation methodology for face-recognition algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (10) (2000) 1090–1104.
- [2] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, *Image Vis. Comput.* 28 (5) (2010) 807–813.
- [3] G.B. Huang, M. Ramesh, T. Berg, 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.
- [4] C. Ding, D. Tao, A comprehensive survey on pose-invariant face recognition, *ACM Trans. Intell. Syst. Technol.* 7 (3) (2016) 37.
- [5] X. Zhu, Z. Lei, J. Yan, D. Yi, S.Z. Li, High-fidelity pose and expression normalization for face recognition in the wild, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.
- [6] T. Hassner, S. Harel, E. Paz, R. Enbar, Effective face frontalization in unconstrained images, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4295–4304.
- [7] A. Li, S. Shan, W. Gao, Coupled bias-variance tradeoff for cross-pose face recognition, *Image Process. IEEE Trans.* 21 (1) (2012) 305–315.
- [8] M. Kan, S. Shan, H. Chang, X. Chen, Stacked progressive auto-encoders (spae) for face recognition across poses, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1883–1890.
- [9] Z. Zhu, P. Luo, X. Wang, X. Tang, Deep learning identity-preserving face space, in: *IEEE International Conference on Computer Vision*, 2013, pp. 113–120.
- [10] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, J. Kim, Rotating your face using multi-task deep neural network, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 676–684.
- [11] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu, M. Rohith, Fully automatic pose-invariant face recognition via 3d pose normalization, in: *IEEE International Conference on Computer Vision*, 2011, pp. 937–944.
- [12] A. Shashua, T. Riklin-Raviv, The quotient image: Class-based re-rendering and recognition with varying illuminations, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) (2001) 129–139.
- [13] W. Deng, J. Hu, X. Zhou, J. Guo, Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning, *Pattern Recognit.* 47 (12) (2014) 3738–3749.
- [14] X. Zhang, Y. Gao, Face recognition across pose: a review, *Pattern Recognit.* 42 (11) (2009) 2876–2896.
- [15] D. Yi, Z. Lei, S. Li, Towards pose robust face recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3539–3545.
- [16] X. Chai, S. Shan, X. Chen, W. Gao, Locally linear regression for pose-invariant face recognition, *Image Process. IEEE Trans.* 16 (7) (2007) 1716–1725.
- [17] J. Heo, M. Savvides, Face recognition across pose using view based active appearance models (vbaams) on cmu multi-pie dataset, in: *International Conference on Computer Vision Systems*, Springer, 2008, pp. 527–535.
- [18] H. Gao, H.K. Ekenel, R. Stiefelhagen, Pose normalization for local appearance-based face recognition, in: *International Conference on Biometrics*, Springer, 2009, pp. 32–41.
- [19] S. Du, R. Ward, Component-wise pose normalization for pose-invariant face recognition, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 873–876.
- [20] H. Zhang, Y. Zhang, T.S. Huang, Pose-robust face recognition via sparse representation, *Pattern Recognit.* 46 (5) (2013) 1511–1521.
- [21] W. Deng, J. Hu, J. Guo, Extended src: undersampled face recognition via intraclass variant dictionary, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (9) (2012) 1864–1870.
- [22] H.T. Ho, R. Chellappa, Pose-invariant face recognition using markov random fields, *IEEE Trans. Image Process.* 22 (4) (2013) 1573–1584.
- [23] Y. Bengio, Learning deep architectures for AI, *Found. Trends Mach. Learn.* 2 (1) (2009) 1–127.
- [24] Z. Zhu, P. Luo, X. Wang, X. Tang, Multi-view perceptron: a deep model for learning face identity and view representations, in: *Advances in Neural Information Processing Systems*, 2014, pp. 217–225.
- [25] V. Blanz, T. Vetter, Face recognition based on fitting a 3d morphable model, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (9) (2003) 1063–1074.
- [26] P. Breuer, K.-I. Kim, W. Kienzle, B. Scholkopf, V. Blanz, Automatic 3d face reconstruction from single images or video, in: *IEEE International Conference on Automatic Face & Gesture Recognition*, 2008, pp. 1–8.
- [27] O. Aldrian, W.A. Smith, Inverse rendering of faces with a 3d morphable model, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (5) (2013) 1080–1093.
- [28] J. Jo, H. Choi, I.-J. Kim, J. Kim, Single-view-based 3d facial reconstruction method robust against pose variations, *Pattern Recognit.* 48 (1) (2015) 73–85.
- [29] U. Prabhu, J. Heo, M. Savvides, Unconstrained pose-invariant face recognition using 3d generic elastic models, *Pattern Anal. Mach. Intell. IEEE Trans.* 33 (10) (2011) 1952–1961.
- [30] R. Abiantun, U. Prabhu, M. Savvides, Sparse feature extraction for pose-tolerant face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (10) (2014) 2061–2073.
- [31] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, S. Shan, Maximal likelihood correspondence estimation for face recognition across pose, *IEEE Trans. Image Process.* 23 (10) (2014) 4587–4600.
- [32] L. Ding, X. Ding, C. Fang, Continuous pose normalization for pose-robust face recognition, *IEEE Signal Process. Lett.* 19 (11) (2012) 721–724.
- [33] J.D. Foley, *Computer Graphics: Principles and Practice*, vol. 12110, Addison-Wesley Professional, 1996.
- [34] P. Pérez, M. Gangnet, A. Blake, Poisson image editing, in: *ACM Transactions on Graphics*, vol. 22, 2003, pp. 313–318.
- [35] S. Du, R. Ward, Wavelet-based illumination normalization for face recognition, in: *IEEE International Conference on Image Processing*, vol. 2, 2005, pp. II–954.

- [36] W. Chen, M.J. Er, S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *Syst. Man Cybern. Part B IEEE Trans.* 36 (2) (2006) 458–466.
- [37] G. Huang, M. Mattar, H. Lee, E.G. Learned-Miller, Learning to align from scratch, in: *Advances in Neural Information Processing Systems*, 2012, pp. 764–772.
- [38] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3d face model for pose and illumination invariant face recognition, in: *IEEE International Conference On Advanced Video and Signal Based Surveillance.*, 2009, pp. 296–301.
- [39] L. Wolf, T. Hassner, Y. Taigman, Similarity scores based on background samples, in: *Computer Vision-ACCV 2009*, Springer, 2009, pp. 88–97.
- [40] Q. Cao, Y. Ying, P. Li, Similarity metric learning for face recognition, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2408–2415.
- [41] D. Chen, X. Cao, F. Wen, J. Sun, 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*, 2013, pp. 3025–3032.

Weihong Deng received the B.E. degree in information engineering and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2004 and 2009, respectively. From October 2007 to December 2008, he was a postgraduate exchange student in the School of Information Technologies, University of Sydney, Australia. He is currently an associate professor in School of Information and Telecommunications Engineering, BUPT. His research interests include statistical pattern recognition and computer vision, with a particular emphasis in face recognition. He has published over 80 technical papers in international journals and conferences, such as IEEE TPAMI and CVPR. He serves as guest editor for Image and Vision Computing Journal and the reviewer for several international journals, such as IEEE TPAMI / TIP / TIFS / TNNLS / TMM / TSMC, IJCV, PR / PRL. Recently, he gives tutorials on face recognition at ICME 2014, ACCV 2014, CVPR2015 and FG2015, and organizes the workshop on feature and similarity learning in ACCV2014 with colleagues. His Dissertation titled “Highly accurate face recognition algorithms” was awarded the Outstanding Doctoral Dissertation Award by Beijing Municipal Commission of Education in 2011. He has been supported by the program for New Century Excellent Talents by the Ministry of Education of China in 2013 and Beijing Nova Program in 2016.

Jiani Hu received the B.E. degree in telecommunication engineering from China University of Geosciences in 2003, and the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008. She is currently a lecturer in School of Information and Telecommunications Engineering, BUPT. Her research interests include information retrieval, statistical pattern recognition and computer vision.

Zhongjun Wu received the B.E. degree in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. He currently is a post-graduate student major in Information and Telecommunications Engineering. His research interests include pose-invariant face recognition and deep learning.

Jun Guo received B.E. and M.E. degrees from Beijing University of Posts and Telecommunications (BUPT), China in 1982 and 1985, respectively, Ph.D. degree from the Tohoku-Gakuin University, Japan in 1993. At present he is a professor and the vice-president of BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and network management. He has published over 200 papers, some of them are on world-wide famous journals or conferences including SCIENCE, IEEE Trans. on PAMI, IEICE, ICPR, ICCV, SIGIR, etc. His book “Network management” was awarded by the government of Beijing city as a finest textbook for higher education in 2004. His team got a number of prizes in national and international academic competitions including: the first place in a national test of handwritten Chinese character recognition 1995, the first place in a national test of face detection 2004, the first place in a national test of text classification 2004, the first place of paper design competition held by IEEE Industry Application Society 2005, the second place in the competition of CSIDC held by IEEE Computer Society 2006.