Face Illumination Manipulation using a Single Reference Image by Adaptive Layer Decomposition

Xiaowu Chen*, Hongyu Wu, Xin Jin* and Qinping Zhao

***Abstract*—This article proposes a novel image-based framework to manipulate the illumination of human face through adaptive layer decomposition. According to our framework, only a single reference image, without any knowledge of the 3D geometry or material information of the input face, is needed. To transfer illumination effects of a reference face image to a normal lighting face, we first decompose the lightness layers of the reference and the input images into large-scale and detail layers through Weighted Least Squares (WLS) filter with adaptive smoothing parameters according to the gradient values of the face images. The large-scale layer of the reference image is filtered with the guidance of the input image by Guided Filter with adaptive smoothing parameters according to the face structures. The relit result is obtained by replacing the large-scale layer of the input image with that of the reference image. To normalize the illumination effects of a non-normal lighting face (i.e. face delighting), we introduce Similar Reflectance Prior (SRP) to the layer decomposition stage by WLS filter, which make the normalized result less affected by the high contrast light and shadow effects of the input face. Through the above two procedures, we can change the illumination effects of a non-normal lighting face by first normalizing the illumination and then transferring the illumination of another reference face to it. We acquire convincing relit results of both face relighting and delighting on numerous input and reference face images with various illumination effects and genders. Comparisons with previous works show that our framework is less affected by geometry differences and can preserve better the identification structure and skin color of the input face.**

***Index Terms*—Face Illumination Manipulation, Face Illumination Transfer, Face Illumination Normalization, Adaptive Layer Decomposition, Edge-preserving Filter**

I. Introduction

Image-based photo-realistic face illumination manipulation has been extensively studied in the computer community and has found wide application in film production, portrait photo editing, etc. The manipulation include transferring illumination of a reference face image to a normal lighting face, normalizing the illumination of a non-normal lighting face, and change illumination of a non-normal lighting face by combining the above two procedures. However, it is still a challenging problem when only a single reference image is available, such as a portrait image with professional lighting (large lighting contrast in common as shown in Fig.1(b)) shot by a professional photographer, which is excellent in lighting usage but is not easy to be restricted in laboratory for generating reference images so as to meet the requirements of current face relighting methods. Recently, many face relighting

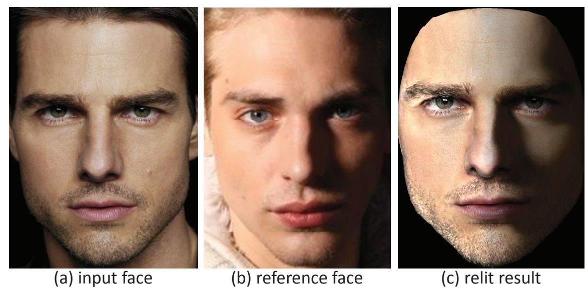

Fig. 1. (a) is the input face image, and (b) is the single reference image. (c) is our relit result which has similar illumination effects to that of (b). The input picture (a) and the reference picture (b) are from Chen et al [1].

methods have been proposed such as morphable model based methods (Wang et al. [2]) and quotient image based methods (Peers et al. [3]; Chen et al. [4]). However, according to morphable model based methods, a collection of scanned textured 3D faces is often needed; while in quotient image based methods are required two reference face images: one has similar lighting effects to the input image, and the other has the desired novel lighting effects.

For more convenient use and wider application, our objective is to generate photo-realistic relit result of an input face image, so as to make the result as similar as possible in illumination effects to those of a single reference face under different illumination (such as Fig.1). This objective is related to 2 basic face illumination manipulation tasks: (1) to transfer illumination effects to normal lighting face (i.e. a face image taken under nearly uniform illumination), and (2) to normalize the illumination effects of a non-normal lighting face (i.e. face illumination normalization or face delighting). Through the above 2 procedures, we can achieve arbitrary face relighting task: to change the illumination effects of a non-normal lighting face by first normalizing the illumination and then transferring the illumination of another reference face to it.

In the community of computational photography, edge-preserving filters (Li [5]; Farbman et al. [6]; Kornprobst et al. [7]; He et al. [8]; Gu et al. [9]) are applied in a variety of applications. In practice, an image is decomposed into a piecewise smooth layer called as a large-scale layer and a detail layer. Eisemann and Durand [10] use the method of decomposition for flash/non-flash image pair fusion. According

Xiaowu Chen, Hongyu Wu, Xin Jin and Qinping Zhao are with State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, 100191, P.R. China.

Xiaowu Chen and Xin Jin are the corresponding author. E-mail: {chen, jinxin}@vrlab.buaa.edu.cn

to [10], flash image is relit by non-flash image, but the flash and non-flash images have the same content. Retinex theory (Land and McCann [11]) tells us that large-scale variance in image is caused by illumination variance, and small-scale variance is caused by reflectance variance. Edge-preserving smoothing can separate the image into large-scale layer and detail layer, detail-scale majorly retaining small variance and large-scale large variance of the original image.

The point of the achievement of this objective is that with the interference from the material and geometry information, how to extract the illumination component from a single reference image. The illumination component is to be used to relight the input image. We are thus inspired to apply edge-preserving smoothing to approximate the process of decomposing lightness into illumination independent layer (reflectance) and illumination dependent layer (shading). Bilateral filter (Kornprobst et al. [7]) and Guided Filter (He et al. [8]) are used to smooth an image and preserve the edges of another image at the same time. Peers et al. [3] use bilateral filter to refine the quotient image used for face relighting. We are also inspired to use similar tools for our face illumination task.

To transfer illumination effects to normal lighting face, we present a face illumination transfer framework based on a single reference face image. We first decompose images into three layers: color, large-scale, and detail layers by adaptive WLS filter according the gradient values; second, as it is assumed that lighting variance retains on the large-scale layer, we operate only on the large-scale layer. We next apply adaptive Guided Filter to smooth the large-scale layer of the reference image and preserve the edges of the large-scale layer of the input image; we get convincing relit result while preserving good identification characteristics of input face.

To normalize the illumination effects of a non-normal lighting face, we tried to transfer the normal lighting effects to a non-normal lighting face using the above face illumination transfer framework (i.e. the reference is a normal lighting face and the input face is a non-normal lighting face.). However, the approximation error can be large.

We then consider the face illumination normalization task as a face intrinsic image decomposition problem. The reflectance image of the input face image is considered as the illumination normalized result. Edge-preserving filters with adaptive smoothing parameters have shown their great ability in decomposing a single face image into the large-scale layer (similar to illumination) and the detail layer (similar to reflectance) (Chen et al. [1]). Thus, we first use edge-preserving filter to remove the detail layers of the input and the reference face images. The quotient between the large-scale layers of the input and the reference face images are then considered as the illumination component (shading image) of the input face image. The the reflectance image (the normalized result) can be considered as the quotient between the input image and the shading image.

For general images, it is not easy to distinguish whether the variance of the images is caused by illumination variance or by reflectance variance in nature images. Recently, in the community of intrinsic image decomposition, user-assisted methods (Bousseau et al. [12]; Shen et al. [13]) are proposed by introducing simple indications by users such as regions of similar illumination or reflectance. With face landmark detection, it is easy to locate regions of similar reflectance in human face. Thus, to obtain more convincing results, we introduce such similar reflectance prior (SRP) in human face into the adaptive parameter selection procedure of the WLS filter. This can enable us to recovery the illumination and reflectance components from a single input face image without computing its 3D geometry using a generic geometric model such as Wen et al. [14] and make the normalized result less affected by the high contrast light and shadow effects of the input face.

For the task of face illumination normalization, a normal lighting reference image is first aligned to the input face according to their facial landmarks. Second, edge-preserving filter with the SRP based adaptive parameter selection scheme is used to extract the large-scale layer of the input and reference face images. Finally, the shading image is obtained by dividing the large-scale layer of the input face image with that of the reference face image. The reflectance image is the quotient of the input face image and the shading image and is considered as the illumination normalized result of the input non-normal lighting face image. For the face illumination normalization task, the large-scale layer of the reference image is NOT filtered with the guidance of the input image by Guided Filter because the incorrect structures caused by shadow or highlight can be also preserved in the normalized result.

Our main contributions include: (1) A framework of face illumination manipulation based on adaptive edge-preserving filters , and (2) A method of face illumination normalization based on similar reflectance prior.

The remainder of this paper is organized as follows: We describe the related work in Section II. Transferring illumination effects to normal lighting face is demonstrated in Section III, parts of this paper which originally appeared as our previous work (Chen et al. [1]) are mainly described in this section. The proposed method of face illumination normalization is described in Section IV. The experiments of face illumination manipulation and comparisons with previous works are demonstrated in Section V. Finally, we conclude our paper and discuss future work in Section VI.

II. Related Work

We review related work in three aspects: face illumination manipulation, intrinsic image decomposition and edge- preserving filters.

Face Illumination Manipulation. Many literatures have addressed image based face illumination manipulation in computer vision and computer graphics (Leung et al. [15], Shim et al. [16], Wu et al. [17]). Wang et al. [2] integrate the spherical harmonics representation (Ramamoorthi and Hanrahan [18]) into the morphable model framework (Blanz and Vetter [19]) by a 3D spherical harmonic basis morphable model (SHBMM). The re-constructed textured 3D faces can be re-rendered under novel illumination conditions. 3D information of a face can also be recovered from depth and intensity Gabor features (Xu et al. [20]). The fitted 3D models are used to generate the relit results, which seem less realistic if insufficient faces are scanned.

Quotient image has been widely used in face illumination transfer such as Riklin-Raviv and Shashua [21]; Peers et al. [3] and Chen et al. [4]. In all these methods, two reference images are needed, which limit their applications when only one reference image is available. To adapt to the formation of quotient image, Peers et al. [3] use a sophisticated Light Stage system to capture linearly radiometric and high dynamic range images.

Most of current single face illumination manipulation methods such as Li et al. [22], Han et al. [23], and Xie et al. [24] are related to face recognition (Zhang et al. [25]). They are often used to deal with low- resolution face images, aiming to remove the light and the shadows in the face for recognition rather than generate convincing relighting results as surveyed in Struc and Pavesic [26]. In the excellent work of face makeup transfer by a single reference image, Guo and Sim [27] adopt a gradient-based editing method, which add only large changes in the gradient domain of the reference image to the input image so as to transfer highlight and shading. Their assumption that large changes are caused by makeup holds if the illumination of the reference image is approximately uniform. This method inspire us to design an adaptive edge-preserving filter for our task. Jin et al. [28] have used local lighting contrast features to learn the artistic lighting template from portrait photos. However, the learned coarse templates can be used for classification rather than illumination transfer. Chen et al. [29] has proposed a method to transfer artistic illumination from artworks. However, they need a database containing artistic illumination templates.

Edge-Preserving Filters. Recently, edge- preserving filters have been an active research topic in computational photography. Edge-preserving filters could smooth an input image and preserve the edges of another image at the same image, and the two input images can be the same or not.The output of bilateral filter (Kornprobst et al. [7]) at a pixel is a weighted average of its neighbouring pixels. The weights simultaneously consider the spatial distance and the range distance of the neighbouring pixels. It is generalized to joint bilateral filter when the edge image is different from the input image.

It is reported in Farbman et al. [6] that bilateral filter may have halo artifacts near some edges in detail decomposition. When the kernel radius is large, it will be quite slow. Weight least square (WLS) filter (Farbman et al. [6]) is an optimization-based method. It adjusts the matrix affinities to non-linearly scale the gradients. It can often get high-quality detail decomposition results. Guided Filter (He et al. [8]) presents a straightforward method to smooth a single image with the guidance of another image by taking the assumption that the output image is a linear transform of the guidance image in local windows. When the input image and the edge image are different, it can get good results.

As WLS filter shows best performance in image detail decomposition, we choose WLS filter for our detail decomposition process. The edge-preserving filters can be used for filtering one image while preserving the edges of another image. We analyze the performance of edge-preserving filters in our scenario (the input object is always under normal illumination condition, and the reference object is under another illumination condition such as a single light source from one side). The joint bilateral filter could make the filtered result preserve well the structure of the input object, but may lose much shading distribution. WLS filter could make the filtered result preserve shading distribution of reference object well but may lose the edge structure of the input object. For our task we observe Guided Filter could perform a better trade-off between shading distribution preservation and edge structure preservation.

Intrinsic Image Decomposition. We consider the face illumination normalization task as a face intrinsic image decomposition problem. Intrinsic image decomposition was first introduced by Barrow and Tenenbaum [30]. An input image can be a product of illumination component and reflectance component. The reflectance describes how an object reflects light. The illumination corresponds to the amount of light incident at a point (essentially irradiance). Although it is often referred to as shading, it includes effects such as shadows and indirect lighting (Bousseau et al. [12]).

Intrinsic image decomposition from a single image is an ill-posed problem. Multiple images are used to overcome the problem. In Weiss [31] and Matsushita et al. [32] registered images captured under different illumination conditions are used to recovery the intrinsic image. While in our scenario, only a single input face image is provided. In Bell and Freeman [33] and Tappen et al. [34], [35], learning-based approaches are proposed to distinguish the variance caused by illumination from reflectance. Other works related to highlight removal such as Tan et al. [36] and Yang et al. [37] are not designed for complex face images.

Recently, various assumptions have been made. Shen et al. [38] assume that similar textures should have similar reflectance. Bousseau et al. [12] and Shen et al. [13] introduce simple indications by users such as regions of similar illumination or reflectance. This can be done easily in human face image by facial landmark detection technologies. Bousseau et al. [12] assume that reflectance variations lie locally in a plane in color space that does not contain the origin. However, this assumption breaks in the presence of very dark reflectance in face images (center of the eyes, eyebrow, etc.). In Shen and Yeo [39], it is assumed that neighboring pixels with similar chromaticities usually have the same reflectance and this sparsity of reflectance is used as a constraint on local reflectance. Shen et al. [13] use the premise that neighbouring pixels in a local window having similar values should have similar reflectance.

Inspired by these works, we also make the assumption (*similar reflectance prior*) that in human face, similar reflectance is shared inside each separated local regions (skin, eyebrows, eyeballs, the white of the eyes, nostrils and mouth).

III. Transfer Illumination Effects to Normal Lighting Face

Our face illumination transfer method is illustrated in Fig.2. The face alignment, layer decomposition and adaptive parameter selection schemes for edge-preserving filtering are described in this section.

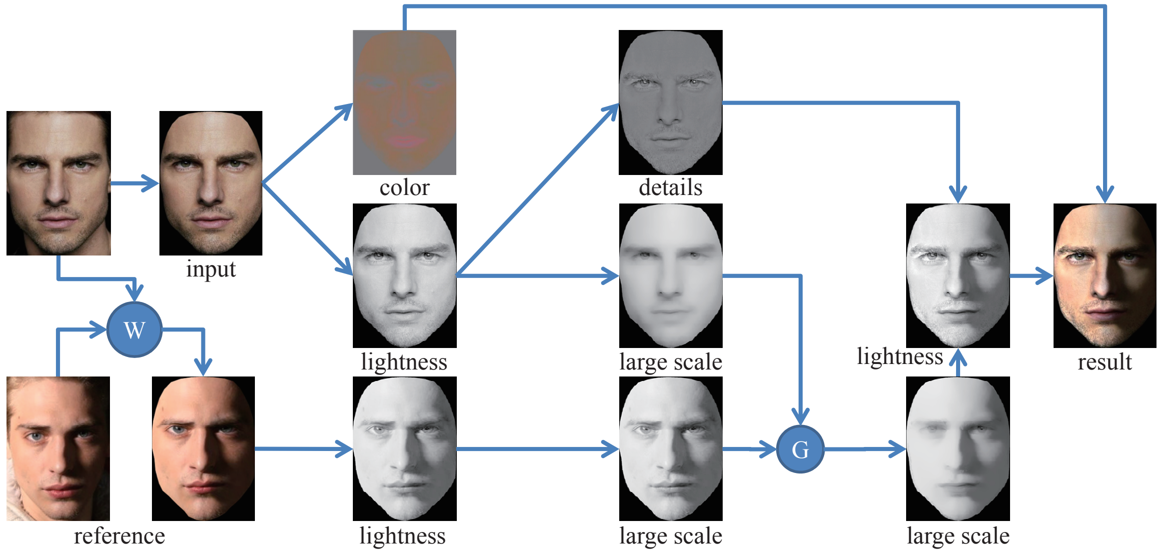

Fig. 2. The proposed face illumination transfer method. The reference face image is warped according to the shape of the input face. Both the input image and the warped reference image are cropped to the contour of the mark points. For the face illumination transfer task, we consider that the reference face image is taken under nearly white light sources. Then the two cropped images are decomposed into lightness layer and color layer, and only lightness layer is operated on. The two lightness layers are decomposed into the large-scale layer and the detail layer by using WLS filter. The reference large-scale layer is filtered with the guidance of the input large-scale layer using Guided Filter to form the large-scale layer of the relit result. By compositing the filtered large-scale layer and the input detail layer, the lightness layer of relit result is obtained. Finally, the relit result is calculated by compositing the lightness layer and the color layer of the input face image.

A. Face Alignment

To transfer the illumination effects from a reference face to an input face, we need to warp the reference face image according to the input face image. We employ Active Shape Model (ASM) (Milborrow and Nicolls [40]) with 104 mark points to identify the mark points on both images. Due to various changes of different faces under various illuminations, current ASM methods tend to fail to locate the accurate mark points, therefore we get a rough initial mark points by using ASM, and then refine their accurate position in an interactive way. In our experiments, one minute is enough to fix the mark points accurately. We then take the mark points as control points to warp the reference image according to the input image by using the affine transform.

B. Layer Decomposition by Adaptive WLS Filter with Gradient Value

First, we decouple the image into lightness and color, and the illumination effects are considered mainly retained on lightness. The Lab, HSV and YCrCb color spaces can be used in our method, as they can separate color image to lightness and color well, the brightness channel (L in Lab space, V in HSV space, Y in YCrCb space) contains lightness information (similar to human perception lightness). We employ edge-preserving filters to smooth the lightness layer so as to obtain the large-scale layer and then use division to obtain the detail layer: $d = l/s$.

Lightness layer, large-scale layer and detail layer are denoted as l, s, and d. The detail layer d can be considered as illumination independent, and large-scale layer as illumination dependent. We choose to apply WLS filter to decompose the lightness layer into the large-scale layer and the detail layer. It is observed that WLS filter can perform well in decomposing lightness into the large-scale layer and the detail layer.

Here, we simply describe WLS filter (Farbman et al. [6]), WLS filter is a solver by minimization of the energy function.

$$E = |l - s|^2 + \lambda H(\nabla s, \nabla l) \quad (1)$$

$$H(\nabla s, \nabla l) = \sum_p \left(\frac{(\partial s/\partial x)_p^2}{(\partial l/\partial x)_p^\alpha + \epsilon} + \frac{(\partial s/\partial y)_p^2}{(\partial l/\partial y)_p^\alpha + \epsilon}\right), \quad (2)$$

where, $|l - s|^2$ is the data term to keep s as similar as to l, and $H(\nabla s, \nabla l)$ is the regularization term to make s as smooth as possible. The subscript p denotes the spatial location of a pixel. α controls over the affinities by non-linearly scaling the gradients. Increasing α will result in sharper preserved edges. λ is the balance factor between the data term and the smoothness term. Increasing λ will produce smoother images.

WLS filter in Farbman et al. [6] performs the same level of smoothing all over the image. But when WLS filter is used in our task, it is expected to perform different levels of smoothing on different regions of the image. Similar to Guo and Sim [27], we set different λ in different regions of the image. Then Eq.1 and Eq.2 can be modified into Eq.3 and Eq.4 ,

$$E = |l - s|^2 + H(\nabla s, \nabla l) \quad (3)$$

$$H(\nabla s, \nabla l) = \sum_p \left(\lambda(p)\left(\frac{(\partial s/\partial x)_p^2}{(\partial l/\partial x)_p^\alpha + \epsilon} + \frac{(\partial s/\partial y)_p^2}{(\partial l/\partial y)_p^\alpha + \epsilon}\right)\right). \quad (4)$$

It is observed that the less flat the region is, the larger λ is required. In the flat region, a small λ is enough to produce a good separation of the large-scale and the detail layers. Most reflectance information can then be retained in the detail layer. However, in the regions such as facial hair and eyebrows, a larger λ is required to perform higher level of smoothing, so that reflectance can be better maintained on the detail layer.

A simple way to set λ over the image is as follows: First, vertical and horizon gradients g_x and g_y of lightness l are calculated, and a threshold t_1 is given, second, for each pixel p, compute the number of the pixels with the gradient scale larger than threshold t_1 in the local window of p,

$$\gamma(p) = \sum_{i \in w_p} (\sqrt{(\partial l/\partial x)_i^2 + (\partial l/\partial y)_i^2} \geq t_1). \qquad (5)$$

After γ is normalized to 0-1, we set λ as follows,

$$\lambda(p) = \lambda_s + (\lambda_l - \lambda_s) * \gamma(p), \qquad (6)$$

where λ_s and λ_l refer to the smaller and larger λ to control the lowest and highest levels of smoothing.

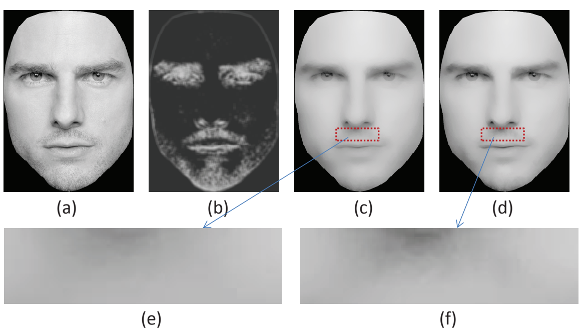

Fig. 3. (a) Input lightness layer, decomposed to large-scale layer (d) by using the same λ over all the image; (b) is normalized γ, which is used to calculate spatial λ. (c) is the large-scale layer calculated by using spatial λ determined by (b). It could be observed that (c) can obtain less detail information than (d) in the regions of facial hair and eyebrows.

In our experiments, $\alpha = 1.2$, the local window radius is 8, $\lambda_s = 1, \lambda_l = 4 * \lambda_s$ and $t_1 = 0.02$.

As shown in Fig.3, by setting different λ spatially, WLS filter performs different levels of smoothing and higher level of smoothing on the regions such as facial hair, eyebrows, etc.. Large-scale layer can be obtained to retain less detail information on the corresponding regions than using a same λ over all the image.

C. Structure Preservation by Adaptive Guided Filter with Edge Detection

To preserve the structure of the input face, the reference large-scale layer is filtered by Guided Filter with the guidance of the input large-scale layer. Guided Filter (He et al. [8]) is briefly described here. Guided Filter has a key assumption that it is a local linear model between guidance I and filtered output q, as q is the linear transform of I in the window w_k centred at the pixel k:

$$q_i = a_k I_i + b_k, \forall i \in w_k, \qquad (7)$$

where, a_k and b_k are assumed to be constant in w_k. Then, the linear coefficients a_k and b_k are determined by minimizing the difference between q and filter input P. The cost function is defined in Eq.8.

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - P_i)^2 + \epsilon a_k^2). \qquad (8)$$

He et al. [8] give the output of a_k and b_k:

$$a_k = \frac{\frac{1}{|w|}\sum_{i \in w_k} I_i P_i - \mu_k \overline{P}_k}{\sigma_k^2 + \epsilon} \qquad (9)$$

$$b_k = \overline{P}_k - a_k \mu_k, \qquad (10)$$

where μ_k and σ_k are the mean and the variance respectively in w_k, $|w|$ is the pixel number in w_k, and $\overline{P}_k$ is the mean of P in w_k. For a pixel i involved in all the windows w_k that contain i, the linear model is applied to all local windows in the image. In the implementation of He et al. [8], they average all possible values of a_k and b_k. We observe that the averaged a_k and b_k is similar to the smoothed version of a_k and b_k which are directly computed by the windows that they are centred. In fact, when the guidance I and the filter input P are different images, the filtered result with the averaged a_k and b_k and the filtered result with a_k and b_k from the centred window have little difference. Thus, we omit the average process, and consider a_k and b_k from the centred window of pixel i as the the representation of all a_k and b_k from the windows involved in pixel i.

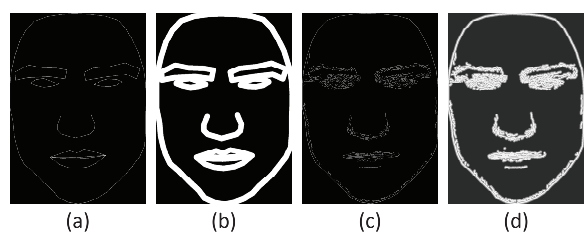

Fig. 4. (a) A rough contour line is decided by face mark points; (b) the face structure region is then determined by the contour line; (c) the Canny edge detector is applied to the face structure region; (d) distance transform is employed to the detected edges, for the pixels far from edges, smaller kernel sizes are used, and for the pixels near the edges, larger kernel sizes are used.

Since it is hard to fix a kernel size (window radius above) for our task, a large kernel size can make the filtered result preserve the edge structure of the input object, but blur the shading information; a small kernel size can get the opposite result. As Guided Filter is a totally local method, it can be extended to different kernel sizes in different regions. Edges in the face structure region (such as eyes, eyebrows, nose and mouth) are important and edges in other regions are less important. We thus set the kernel size a small value in the non-face-structure region and treat the face structure region carefully, which can preserve better the structure of the input face by sacrificing part of the shading distribution in the face structure region.

We extend Guided Filter to different kernel size spatially as follows: we first define a mask containing face structure region, and then treat pixels in the mask region carefully. Our basic idea is to set the kernel size near the edges in the face structure region to be of larger value. Distance transform is applied to set the kernel size that gradual changes of the gradual change of the distance away from the edges in face structure region. As shown in Fig.4, Kernel size r in the face structure region is decided as follows: First, the mark points can construct a rough contour of the face structure, and then the face structure region is determined; second, the Canny edge detector is applied to detect the edges in the face structure region of the large-scale layer of the reference face; third, for the edges in the face structure region, compute the distance of all pixels from these edges. Finally, spatially kernel sizes are defined by Eq.11 and 12.

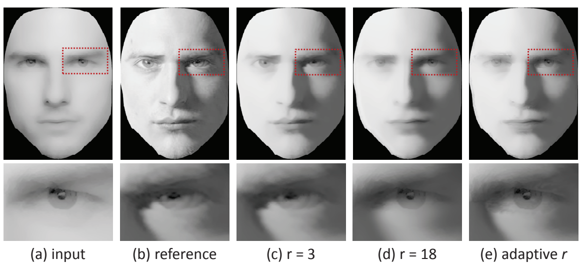

Fig. 5. Large-scale layers with different Guided Filter parameters. (a) input large-scale layer; (b) reference large-scale layer; (c) a single kernel size $r = 3$ is used over all the image; (d) a single kernel size $r = 18$ is used over all the image; and it can be observed that (c) preserves much more structures of the reference image, it retains much shading information as well as more identification characteristics of the reference face; (d) maintains the structure of the input face well, but blurs the shading information of the reference face; (e) performs a trade-off between preserving the input structure in the face structure region and retaining the shading information of the reference face by using adaptive kernel sizes.

$$dist(p) = |p - q(\min_q(|q - p|))| \tag{11}$$

$$r(p) = \begin{cases} \frac{dist(p)}{T_d} r_0 + (1 - \frac{dist(p)}{T_d}) r_1 & \text{if } dist(p) \le T_d \\ r_0 & \text{else} \end{cases} \tag{12}$$

where $|p_1 - p_2|$ means the Euclidean distance from pixel p_1 to pixel p_2. We set $r_1 = 18, r_0 = 3$. T_d is a threshold of $dist(p)$ and we set $T_d = 10$ in our experiments.

As shown in Fig.5, our modified Guided Filter with different kernel sizes over the entire image performs well the trade-off between preserving the edges of the input image and preserving the shading distribution of the reference image.

After we filter the reference large-scale layer with the guidance of the input large-scale layer by using our modified Guided Filter, the desired input lightness is got by compositing the filtered reference large-scale layer and the input detail layer. Finally, we get the relit result by incorporating the color information of the input face.

IV. Face Illumination Normalization

We consider the face illumination normalization task as a face intrinsic image decomposition problem. The reflectance image of the input face image is considered as the illumination normalized result. The workflow of our face illumination normalization method is illustrated in Fig. 6. The new adaptive parameter selection schemes for WLS filter based on similar reflectance prior will be described in this section. The large-scale layer of the reference image is NOT filtered with the guidance of the input image by Guided Filter because the incorrect structures caused by shadow or highlight can be also preserved in the normalized result. A comparison of using Guide Filter or not is illustrated in Section V.

A face image $\boldsymbol{I}$ can be denoted as the per-color-channel product of a reflectance component $\boldsymbol{R}$ and an illumination component $\boldsymbol{S}$: $\boldsymbol{R} * \boldsymbol{S}$. To follow this assumption, we operate on RGB color space for illumination normalization to consider color light sources.

$\boldsymbol{R}_{in}$ and $\boldsymbol{S}_{in}$ refer to the reflectance and illumination components of the input face image. $\boldsymbol{R}_{ref}$ and $\boldsymbol{S}_{ref}$ refer to the reflectance and illumination components of the reference face image. We denote the WLS filter with adaptive parameter in our method as $f_{wls}(\cdot)$. The illumination component can be obtained as: $\boldsymbol{S}_{in} = f_{wls}(\boldsymbol{I}_{in})/f_{wls}(\boldsymbol{I}_{ref})$. The reflectance component is obtained as: $\boldsymbol{R}_{in} = \boldsymbol{I}_{in}/\boldsymbol{S}_{in}$.

A. Similar Reflectance Prior

Using face landmark detection with manual adjustment, the face image is divided into these regions: skin, eyebrows, eyeballs, the white of the eyes, nostrils and mouth (see Fig. 6). We assume that inside each region the reflectance is similar. Thus, the large variance inside each region is assumed to be caused by illumination. And the intensity changes on the boundary between adjacent regions are assumed to be caused by reflectance. For example, illumination changes smoothly on most of the skin and mouth regions except on the boundary of cast shadow and highlight area, which is shown in Fig. 7 (a) and (b).

B. WLS Filter by Similar Reflectance Prior

We add similar reflectance prior to the parameter selection of WLS filter. We consider that gradient values correspond to the variance of image. If the gradient values is larger than a threshold, the variance is considered to caused by illumination. This type of variance should be extracted as the illumination component and the smoothing parameter λ on these pixels should be a small value. If the gradient values are less than a threshold the variance is considered to be caused by reflectance. The illumination should be smooth on these pixels, so that the λ should be selected according to their gradient values. The parameter $\lambda(p)$ inside each region can be obtained as follows:

$$\lambda(p) = \begin{cases} C_s & \text{if } ga(p) \ge T_g \\ ga(p) & \text{others} \end{cases} \tag{13}$$

$$ga(p) = \sqrt{(\partial l/\partial x)_i^2 + (\partial l/\partial y)_i^2} \tag{14}$$

where, C_s is a constant small value, which helps the WLS filter to extract the variance caused by illumination. T_g is the

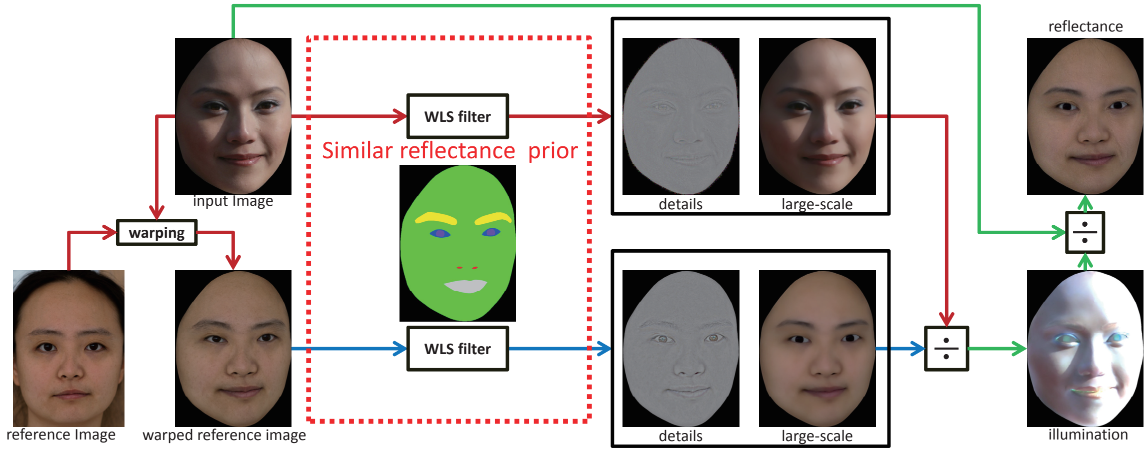

Fig. 6. The workflow of face illumination normalization method. The reference face image (a normal lighting face) is warped according to the landmarks of the input face. The two images are filtered with WLS filter with parameter selection scheme constrained by the similar reflectance prior. The large-scale layer of input image is divided by that of reference image to obtain the illumination component. The reflectance component (the normalized result) is the quotient between the input image and the illumination component.

threshold of the square sum of gradient values, and is used to distinguish whether the variance is caused by reflectance or by illumination. It can be seen that, if $ga(p)$ is larger than the threshold T_g, we can considered that the variance is caused by illumination. Thus, a small value C_s is needed to extract the variance. In our experiments $C_s = 0.15$, $T_g = 2 * mean(ga)$ inside each region.

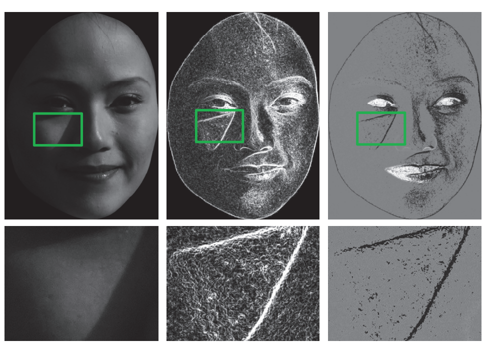

Fig. 7. Adaptive parameter selection in skin region. (a) The blue channel of the input face. (b) Visualization of the square sum of gradients on vertical and horizontal axis. (c) Visualization of the adaptive parameter constrained by SRP. From (a) and (b), Along the edge of the cast shadow, the gradients value is very large. The variance is caused by illumination, and thus the smoothing parameter λ are set small according to the similar reflectance prior.

We have the observation that, filtering using the above λ works well inside the skin and mouth regions. However, in the regions of eyebrow, the white of eyes, eyeballs and nostril, the large variance is caused not only by the illumination, but also by the uncertain details inside these regions (irregularity of the distribution of eyebrow and complex texture of the eyeballs. etc). In Farbman et al. [6], an input image is filtered by WLS filter at different resolutions. Inspired by their multi-scale method, we resize these regions to a lower resolution. Then the variance caused by detail reflectance is almost removed by down-sampling the image. Then the illumination is extract by WLS filter by similar parameter selection scheme inside the skin and mouth regions. Finally, the illumination components of those regions are resized to their original resolutions.

The eyebrow region is filtered by WLS filter directly at its original resolution. The variance cause by reflectance is retained on the large-scale layer. The eyebrow region is filtered after being resized to 1/10 of its original resolution. The variance caused by reflectance is almost removed from the large-scale layer.

V. Experiments

In this section, our face illumination manipulation results and comparisons with previous works are shown. Convincing relit results of both face relighting and delighting are output by our face illumination manipulation framework.

We test our framework on numerous input and reference face images with various illumination effects and genders. Comparisons with previous work show that our method is less affected by geometry differences and can preserve better the identification structure and skin color of the input face.

A. E1: Transfer Illumination Effects to Normal Lighting Face

In Fig.8, we present more experiment results. The results show that our method can perform well in illumination transfer between genders. We implemented our illumination transfer method by MATLAB. On a dual-core CPU (Intel E8400 3.0 GHz), the runtime of a 1000*1000 face image is about 42 seconds. The face alignment takes about 16 seconds, the layer decomposition takes about 24 seconds and the guided filtering takes about 2 seconds.

B. E2: Illumination Transfer in Multi-Color Spaces

Our illumination transfer method can be applied in various channels in different color spaces. We test the Lab, HSV and YCrCb color spaces in our mehod. Fig. 9 (d)-(f) show the

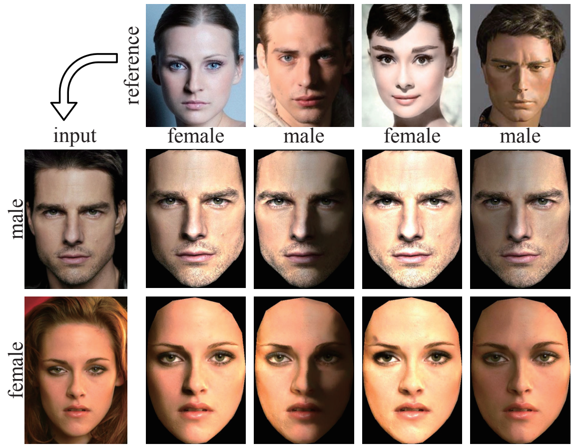

Fig. 8. Our illumination transfer results between genders.

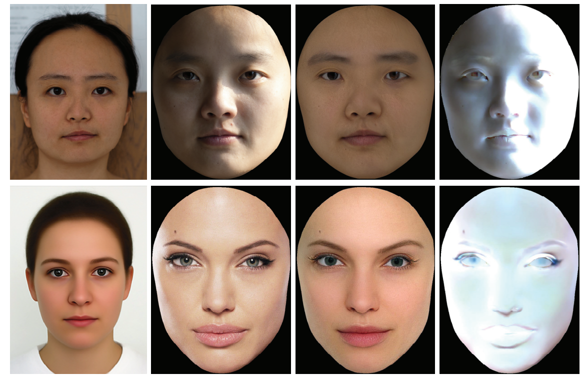

Fig. 10. The results of face illumination normalization. For the input Asia female faces we use a Asia female face image taken under normal lighting. For the input Euro female faces we use the average female face of 64 female faces photos generated by the Beauty Check project. The 64 female faces are taken under normal lighting. The illumination of the average face is more uniform. The illumination normalization results are shown in (c) with the separated illumination components shown in (d).

illumination transfer results in Lab, HSV and YCbCr color space. We also test our method on RGB channel. Fig. 9 (d) is the illumination transfer in RGB color space. The layer decomposition and guided filter procedures are performed on each color channel of the RGB color space after face alignment. The operation on L channel of Lab color space can preserve the most identification features (texture, skin color) of the input face.

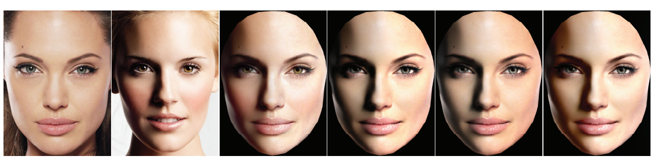

Fig. 9. Illumination transfer in multi-color spaces. (a) is the input image, (b) is the reference image, (c) is the illumination transfer result in R, G, B channel of RGB color space, (d) is the illumination transfer result in L channel of Lab color space, (e) is the illumination transfer result in H channel of HSV color space, (f) is the illumination transfer result in Y channel of YCbCr color space.

C. E3: Face Illumination Normalization

In Fig. 10, we test our method on four images of two different races. We choose two reference images from the persons of two races according to the input images. We make the assumption that the persons in the image should have little makeup on their face. The result (the fourth row in Fig. 10) shows that if the reference image is very close to the input image, more convincing result will be generated. More results are be submitted in the supplementary materials.

D. E4: Arbitrary Face Relighting

Through the de-light and adding light procedures, as shown in Fig. 11, we can achieve arbitrary face relighting task: to change the illumination effects of a non-normal lighting face by first normalizing the illumination and then transferring the illumination of another reference face to it. We can virtually turn the light source from left to right to human face, as illustrated in Fig. 11. This greatly extent the ability of our face illumination manipulation framework.

E. E5: Filter Performance

WLS Filter with Gradient vs. SRP. Performance of our face illumination normalization method mainly relies on the filter results of WLS filter. To validate the efficiency of the proposed method, we first check the function of our modification of the WLS filter parameter selection scheme with the similar reflectance prior in the human face. As shown in Fig. 12, by setting a small λ in the skin region, the variance caused by illumination is maintained, but the detail information of reflectance is retained on shading component. A large λ can remove the detail information of reflectance but the variance caused by illumination is also partly removed. Adaptive λ of Section III is only defined by gradient values. The larger the gradient value is the larger the smoothing parameter λ is. This scheme can be affected by large illumination changes inside similar reflectance regions. We introduce the similar reflectance prior of human face into the adaptive λ selection, which can not only remove the detail information of reflectance but also maintain the variance caused by illumination.

Guided Filter in Face Illumination Normalization. For the face illumination normalization task, the large-scale layer of the reference image is NOT filtered with the guidance of the input image by Guided Filter because the incorrect edges caused shadow or highlight can be also preserved in the input image. A comparison of using Guide Filter or not is illustrated in Fig. 12.

F. E6: Comparison with Previous Methods

Comparison with Quotient Image based Method. We compare our face illumination transfer method (Section III) with the previous work. In Chen et al. [4], they relight the input image under the frontal lighting condition by two reference images (one is the reference image under the frontal

(a) reference of illumination transfer (b) reference of illumination normalization (c) input (d) directly transfer (e) normalized (f) result

Fig. 11. Arbitrary Face Relighting. (a) is the reference of illumination transfer. (b) is the reference of illumination normalization (c) is the input face with non-normal lighting. (d) is the result of directly transferring the illumination of (a) to (c) using the method in Section III. (e) is the illumination normalized faces of (c) by using the method in Section IV. (f) is the relit result of transferring the illumination of (a) to (e).

(a) λ = 0.01 (b) λ = 4 (c) λ via only gradient value (d) reflectance prior +guided filter (e) λ via similar reflectance prior

Fig. 12. Illumination normalization in skin region with different parameters: the normalized result with a small and large constant λ over all the skin region ((a) and (b)), with adaptive λ only according to the gradient value (Chen et al. [1]) (c), with the adaptive λ using the similar reflectance prior and Guided Filter for alignment (d), with the adaptive λ using the similar reflectance prior in the WLS filter (e). The original face image of this figure is shown in Figure 6.

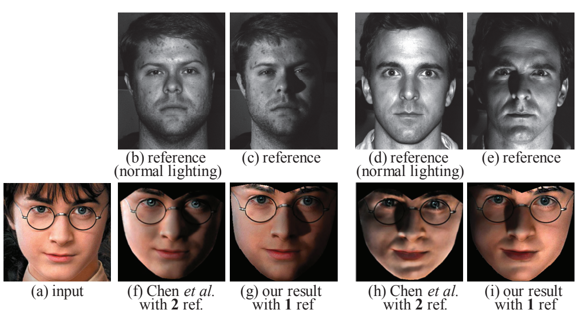

Fig. 13. Comparisons with Chen et al. [4]. (b)(d) and (c)(e) are reference faces from YaleB [41] taken under normal and the desired lighting conditions, respectively. (f) and (h) are results of Chen et al. [4] using two reference images. (g) and (i) are our results using one reference image.

lighting condition and the other is the reference image under the desired lighting condition). As shown in Fig.13, our method needs only one reference face image and the numerical comparison in Fig. 14 shows that the illumination transfer results of our method are similar to that of Chen et al. [4].

Comparison with Single Image based Method. To the best of our knowledge, the most similar to our face illumination transfer work (Section III) is Li et al. [22]. They relight a single image using a single reference image. Fig.15 gives a comparison of our result with the method in Li et al. [22]. They decompose images into large-scale layer (considered as illumination dependent) and small-scale layer (considered as illumination independent) in RGB channels respectively. Then they get the result by re-combination of the input small scale layer and the reference large-scale layer in each channel. In contrast, in our framework, only the lightness layer is operated on, leading to unchanged color of the relit result. Furthermore, using edge-preserving filters in our framework introduces less shading details caused by geometry difference between the input and the reference faces.

(a) reference images (b) input images (c) Li et al. (d) our results

Fig. 15. Comparisons with Li et al. [22]. (a) are the reference images, (b) are the input images, (c) are the illumination transfer results of Li et al. [22], (d) are the our results.

Comparison with Intrinsic Image Decomposition Method. According our face illumination normalization method (Section IV), we use similar reflectance prior and consider this task as a face intrinsic image decomposition problem. By using face landmark detection and manual adjustment of the location of face landmarks, we can locate the face regions and assume that the reflectance in the same region is similar. Decomposition results of Bousseau et al. [12] can be achieved using the global constraints of shading and reflectance provided by user scribbles. However, users may not provide correct scribbles if they are not familiar with intrinsic images. However, it is easy to locate regions of similar reflectance. As shown in Fig. 16, our method can achieve more convincing decomposition results especially on the highlight

(a) reference (normal lighting) (b) reference (c) input (d) ground truth (e) Chen et al. (f) our method accumulated percentage

Fig. 14. Quantitative comparison with quotient image based method. (a) and (b) are the reference images, (c) are the input images, (d) are the ground truth, (e) are the results of Chen et al.[4], (f) are the results of our method, (g) accumulated percentage of pixel-intensity difference between the illumination transfer results and the ground truth. We use only one reference image (b) to obtain similar relit results to those of Chen et al. [4], which needs two reference images (a) and (b).

and shadow areas.

(a) input images (b) Bousseau *et al.* (c) Shen *et al.* (d) our method

Fig. 16. The results of Bousseau et al. [12] and Shen et al. [39] are shown in (b) and (c), respectively. (d) is our result. Our method can separate the illumination from the input image more convincingly. The variance caused by illumination is not retained in our reflectance components.

VI. Conclusion and Discussion

In this paper, we have presented a novel image-based framework for face illumination manipulation through edge-preserving filters. We also propose adaptive parameter selection schemes in WLS filter and Guided Filter processes in face illumination manipulation application. For face illumination normalization, we introduce similar reflectance prior to the layer decomposition stage with adaptive WLS filter. Through removing lighting and adding lighting, we can change the illumination effects of a non-normal lighting face by first normalizing the illumination and then transferring the illumination of another reference face to it. The main advantage of our method lies in that only a single reference image is required with 3D geometry or material information of the input face. Convincing relit results demonstrate that our method is effective and advantageous in preserving the identification structure and skin color of the input face.

Limitation and future work. For face illumination transfer, we consider that the reference face image is taken under nearly white light sources. For face illumination normalization, we operate on RGB color spaces to consider color light sources. Both illumination transfer and normalization can be operated on various channels in different color spaces. In the future work, we will consider the adaptation scheme for various colors of light sources. We consider the face illumination normalization task as a face intrinsic image decomposition problem. However, human faces do not strictly respect the Lambertian model. The separating of the Lambertian shading and the specularities of the human face is to be further studied.

Acknowledgment

We would like to thank the anonymous reviewers for their help in improving the paper. This work was partially supported by NSFC (60933006), 863 Program (2012AA011504) and R&D Program (2013BAH35F01). Parts of the results and figures presented in this paper have previously appeared in our previous work.

References

[1] X. Chen, M. Chen, X. Jin, and Q. Zhao, "Face illumination transfer through edge-preserving filters," in *Proc. CVPR*, 2011, pp. 281–287.

[2] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras, "Face relighting from a single image under harsh lighting conditions," in *Proc. CVPR*, June 2007.

[3] P. Peers, N. Tamura, W. Matusik, and P. E. Debevec, "Post-production facial performance relighting using reflectance transfer," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 26, no. 3, p. 52, 2007.

[4] J. Chen, G. Su, J. He, and S. Ben, "Face image relighting using locally constrained global optimization," in *Proc. ECCV*, 2010, pp. 44–57.

[5] S. Z. Li, "Roof-edge preserving image smoothing based on mrfs," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 1134–1138, 2000.

[6] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 27, no. 3, Aug. 2008.

[7] P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–74, 2009.

[8] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. ECCV*, 2010, pp. 1–14.

[9] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping." *IEEE Transactions on Image Processing*, no. 1, pp. 70–79.

[10] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, pp. 673–678, 2004.

[11] E. H. Land and J. J. McCann, "Lightness and Retinex Theory," *Journal of the Optical Society of America*, vol. 61, p. 1, Jan. 1971.

[12] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," in *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 2009, pp. 1–10.

[13] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Proc. CVPR*, 2011, pp. 3481–3487.

[14] Z. Wen, Z. Liu, and T. S. Huang, "Face relighting with radiance environment maps," *Proc. CVPR*, vol. 2, p. 158, 2003.

[15] C.-S. Leung, T.-T. Wong, P.-M. Lam, and K.-H. Choy, "An rbf-based compression method for image-based relighting," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 1031–1041, Apr. 2006.

[16] H. Shim, J. Luo, and T. Chen, "A subspace model-based approach to face relighting under unknown lighting and poses," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1331–1341, Aug. 2008.

[17] H. Wu, X. Chen, M. Yang, and Z. Fang, "Facial performance illumination transfer from a single video using interpolation in non-skin region," *Computer Animation and Virtual Worlds*, vol. 24, no. 3-4, pp. 255–263, 2013.

[18] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2001, pp. 497–500.

[19] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1999, pp. 187–194.

[20] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3d face recognition from depth and intensity gabor features," *Pattern Recognition*, vol. 42, pp. 1895–1905, September 2009.

[21] T. Riklin-Raviv and A. Shashua, "The quotient image:class based recognition and synthesis under varying illumination conditions," *Proc. CVPR*, vol. 2, p. 2566, 1999.

[22] Q. Li, W. Yin, and Z. Deng, "Image-based face illumination transferring using logarithmic total variation models," *Visual Computer*, vol. 26, no. 1, pp. 41–49, 2009.

[23] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao, "Lighting aware preprocessing for face recognition across varying illumination," in *Proc. ECCV*, 2010, pp. 308–321.

[24] X. Xie, W.-S. Zheng, J.-H. Lai, P. C. Yuen, and C. Y. Suen, "Normalization of face illumination based on large-and small-scale features." *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1807–1821, 2011.

[25] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2599–2606, 2009.

[26] V. Štruc and N. Pavešić, *Performance Evaluation of Photometric Normalization Techniques for Illumination Invariant Face Recognition*, Y. Zhang, Ed. IGI Global, 2010.

[27] D. Guo and T. Sim, "Digital face makeup by example," in *Proc. CVPR*, 2009.

[28] X. Jin, M. Zhao, X. Chen, Q. Zhao, and S.-C. Zhu, "Learning artistic lighting template from portrait photographs," in *Proc. ECCV*, 2010, pp. 101–114.

[29] X. Chen, X. Jin, Q. Zhao, and H. Wu, "Artistic Illumination Transfer for Portraits," *Computer Graphics Forum*, vol. 31, no. 4, pp. 1425–1434, 2012.

[30] H. G. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," *Computer Vision Systems*, pp. 3–26, 1978.

[31] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. ICCV*, 2001, pp. 68–75.

[32] Y. Matsushita, S. Lin, S. B. Kang, and H.-Y. Shum, "Estimating intrinsic images from image sequences with biased illumination," in *Proc. ECCV*, 2004, pp. 274–286.

[33] M. Bell and W. T. Freeman, "Learning local evidence for shading and reflectance," in *Proc. ICCV*, 2001, p. 670.

[34] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1459–1472, 2005.

[35] M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using non-linear regression," in *Proc. CVPR*. Los Alamitos, CA, USA: IEEE Computer Society, 2006, pp. 1992–1999.

[36] P. Tan, L. Quan, and S. Lin, "Separation of highlight reflections on textured surfaces," *Proc. CVPR*, vol. 2, pp. 1855–1860, 2006.

[37] Q. ang, S. Wang, and N. Ahuja, "Real-time specular highlight removal using bilateral filtering," in *Proc. ECCV*, 2010, pp. 87–100.

[38] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *Proc. ECCV*, 2008".

[39] L. Shen and C. Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *Proc. CVPR*, 2011, pp. 697–704.

[40] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. ECCV*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 504–513.

[41] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[42] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground-truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. ICCV*, 2009, pp. 2335–2342.

Xiaowu Chen received Ph.D. degree at Beihang University in 2001. He is a professor in the State Key Laboratory of Virtual Reality Technology and Systems, also in the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, computer graphics, virtual reality and augmented reality.

Hongyu Wu is presently a Ph.D. candidate in the State Key Laboratory of Virtual Reality Technology and Systems, also in the School of Computer Science and Engineering, Beihang University. His research interests are visual computing and computer vision.

Xin Jin is presently a Ph.D. candidate in the State Key Laboratory of Virtual Reality Technology and Systems, also in the School of Computer Science and Engineering at Beihang University. His research interests are visual computing and virtual reality.

Qinping Zhao is a professor in the School of Computer Science and Engineering, Beihang University. He is the founder and director of the State Key Laboratory of Virtual Reality Technology and Systems. His research interests include virtual reality and artificial intelligence.