Face Illumination Transfer through Edge-preserving Filters

Xiaowu Chen, Mengmeng Chen, Xin Jin* and Qinping Zhao
State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China

Abstract

This article proposes a novel image-based method to transfer illumination from a reference face image to a target face image through edge-preserving filters. According to our method, only a single reference image, without any knowledge of the 3D geometry or material information of the target face, is needed. We first decompose the lightness layers of the reference and the target images into large-scale and detail layers through weighted least square (WLS) filter after face alignment. The large-scale layer of the reference image is filtered with the guidance of the target image. Adaptive parameter selection schemes for the edge-preserving filters is proposed in the above two filtering steps. The final relit result is obtained by replacing the large-scale layer of the target image with that of the reference image. We acquire convincing relit result on numerous target and reference face images with different lighting 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 target face.

1. Introduction

Image-based photo-realistic face relighting without 3D model has been extensively studied in the computer community and widely used in film production and game entertainment. However, it is still a challenging problem when only a single reference image is available.

Recently, many face relighting methods have been proposed such as morphable model based methods [17] and quotient image based methods [14, 2]. 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 target image, and the other has the desired novel lighting effects.

For more convenient use and wider application, our objective is to generate photo-realistic relighting result of a frontal face image taken under nearly uniform illumination, so as to make the result as similar as possible in lighting effects to that of only a single reference face image under another illumination (see Figure 1).

(a) target face (b) reference face (c) relit result

Figure 1. The objective. (a) is the target face image, and (b) is the single reference image. (c) is our relit result which has similar lighting effects to that of (b).

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 target image. Due to the ill-posedness of the problem, current automatic methods of extracting illumination component from a single image such as [15] will often fail to handle complex natural images, especially when large lighting contrast exists. Even by using an user-assisted as approach proposed by Bousseau *et al.* [1], the separated illumination components still contain not just illumination but also some geometry and material information. This is far from true illumination component where object reflectance is all the same. The face relighting results would thus be influenced.

Most of current single face relighting methods such as [7] are related to face recognition. 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 [16]. Jin *et al.* [9] have used local lighting contrast features to learn artistic lighting template from portrait pho-

*Corresponding author (email: jinxin@vrlab.buaa.edu.cn)

tos. However the template is designed for classification and numerical aesthetic quality assessment of portrait photos. Thus, the template is not suitable to transfer illumination. In the application of face makeup transfer by a single reference image, Guo and Sim [6] adopt a gradient-based editing method, which add only large changes in the gradient domain of the reference image to the target image so as to transfer highlight and shading. However, their assumption that large changes are caused by makeup holds if the illumination of the reference image is approximately uniform.

To the best of our knowledge, the work most similar to ours is [12]. They relight a single image using a single reference image. They decompose images into large-scale layer (considered as illumination dependent) and small-scale layer (considered as illumination independent) in RGB channels respectively, and then get the result by re-combination of the target small scale layer and the reference large-scale layer in each channel. However, the color of the target face is not preserved well.

Recently, edge-preserving filters have been an active research topic in computational photography [10, 3, 8]. Edge-preserving filters can 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. Retinex theory [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 retaining small variance and large-scale large variance of the original image. For our single face relighting task, 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).

We analyze the performance of edge-preserving filters in our scenario (see Figure 1). As Weight least square(WLS) filter [3] shows best performance in image detail decomposition, we choose WLS filter for our detail decomposition process. The joint bilateral filter [10] could make the filtered result preserve well the structure of the target 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 target object. Guided filter [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. For our task we observe guided filter [8] could perform a better trade-off between shading distribution preservation and edge structure preservation.

In this paper, we present a method of frontal face image relighting based on a single reference face image. The light color is considered as nearly white. We first decompose images into three layers: color, large-scale, and detail layers; 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 edge-preserving filters to smooth the large-scale layer of the reference image and preserve the edges of the large-scale layer of the target image; Finally, we get convincing relit result while preserving good identification characteristics of target face.

Our main contributions include: (1) A framework of face illumination transfer based on edge-preserving filters , and (2) Adaptive parameter selection schemes in WLS filter and guided filter for face relighting.

2. Face Illumination Transfer

The workflow of our method is illustrated in Figure 2. The face alignment, layer decomposition and adaptive parameter selection schemes for edge-preserving filtering will be described in this section.

2.1. Face Alignment

To transfer the illumination effects from a reference face to a target face, we need to warp the reference face image according to target face image. We employ Active Shape Model (ASM[13]) 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 target image by using the affine transform.

2.2. Layer Decomposition

First, we decouple the image into lightness and color, and the lighting effects are considered mainly retained on lightness. We choose CIE 1976 (L^*,a^*,b^*) color space, as it could separate color image to lightness and color well, L^* channel contains lightness information (similar to human perception lightness), and a^* channel and b^* channel contain color information. 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.

$$d = l/s \tag{1}$$

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

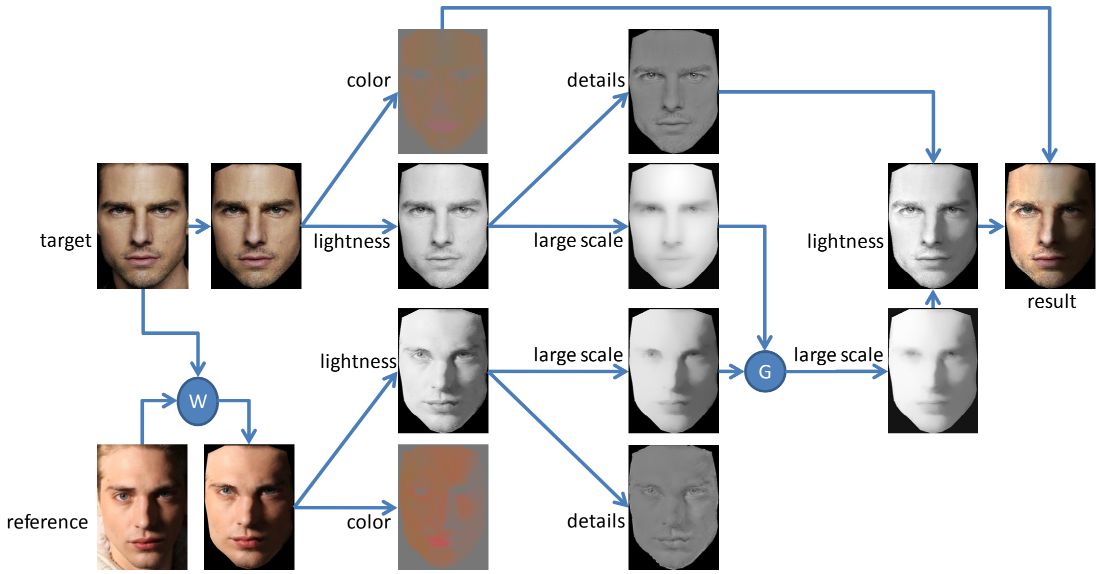

Figure 2. The workflow of the proposed method. The reference face image is warped according to the shape of the target face. Both the target image and the warped reference image are cropped to the contour of the mark points. Then the two cropped images are decomposed into lightness layer and color layer, 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 target 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 target 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 target face image.

detail layer. It is observed that WLS filter can perform well in decomposing lightness into the large-scale layer (similar to shading) and the detail layer (containing reflectance).

2.3. WLS filter with Adaptive Parameter

WLS filter in [3] 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 [6], we set different different smooth levels in different regions of the image. Thus, the modified version of the energy function of WLS[3] filter is

$$E = |l - s|^2 + H(\nabla s, \nabla l) \tag{2}$$

$$H(\nabla s, \nabla l) = \sum_p (\lambda(p)(\frac{(\partial s/\partial x)_p^2}{(\partial l/\partial x)_p^\alpha + \epsilon} + \frac{(\partial s/\partial y)_p^2}{(\partial l/\partial y)_p^\alpha + \epsilon})). \tag{3}$$

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.

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 in 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,

$$\gamma(p) = \sum_{i \in w_p} (\sqrt{(\partial l/\partial x)_i^2 + (\partial l/\partial y)_i^2} \geq t_1). \tag{4}$$

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

$$\lambda(p) = \lambda_s + (\lambda_l - \lambda_s) * \gamma(p), \tag{5}$$

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

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 Figure 3 , by setting different λ spatially, WLS filter performs different levels of smoothing and

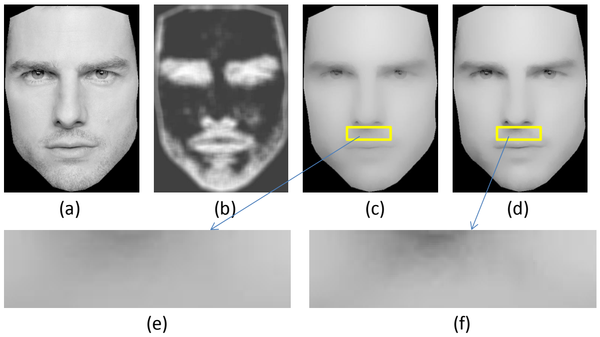

Figure 3. (a) Target lightness layer, decomposed to large-scale layer (d) by using a same λ over all the image; (b) is normalized γ, which is used to calculate spatial λ. (c) is the large-scale layer calculated by using spatial λ determined by (b). It could be observed that (c) can obtain less detail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

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.

2.4. Guided Filter with Adaptive Parameter

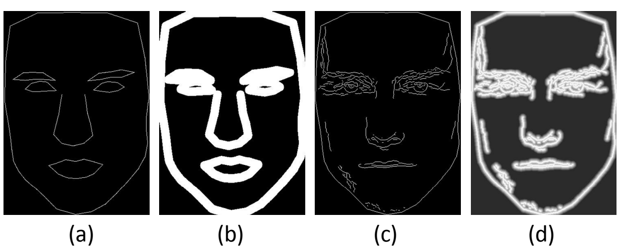

Figure 4. (a) A rough contour line is decided by face mark points; (b) the face structure region is then determined by the contour line; (c) the Canny edge detector is applied to the face structure region; (d) distance transform is employed to the detected edges, for the pixels far from edges, smaller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

The reference large-scale layer is filtered by guided filter with the guidance of the target large-scale layer. Guided filter [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 centered at the pixel k:

$$q_i = a_k I_i + b_k, \forall i \in w_k, \tag{6}$$

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.7.

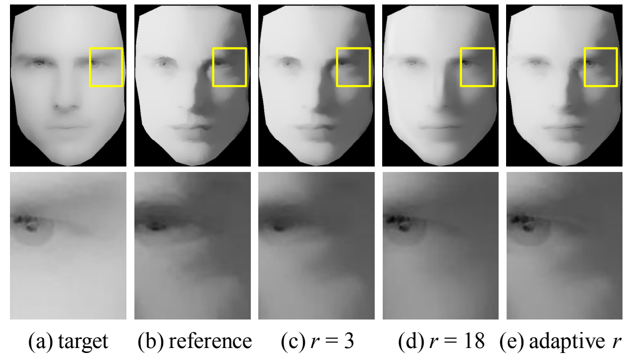

Figure 5. Large-scale layers with different guided filter parameters. (a) target large-scale layer; (b) reference large-scale layer; (c) a single kernel size $r = 3$ is used over all the image; (d) a single kernel size $r = 18$ is used over all the image; and it can be observed that (c) preserves much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

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - P_i)^2 + \epsilon a_k^2). \tag{7}$$

[8] gives the output of a_k and b_k:

$$a_k = \frac{\frac{1}{|w|} \sum_{i \in w_k} I_i P_i - \mu_k \overline{P}_k}{\sigma_k^2 + \epsilon} \tag{8}$$

$$b_k = \overline{P}_k - a_k \mu_k, \tag{9}$$

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. Then the linear model is applied to all local windows in the image, for a pixel i involved in all the windows w_k that contain i. In [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 centered. 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 centered window of pixel i as the representation of all a_k and b_k from the windows involved in pixel i.

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 target 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 target 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, and 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 Figure 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.10 and 11.

$$dist(p) = |p - q(\min_q(|q-p|))| \tag{10}$$

$$r(p) = \begin{cases} r_0 + (r_1 - r_0) * \frac{(T_d - dist(p))}{d_t} & \text{if } dist(p) \le T_d \\ r_0 & \text{others} \end{cases} \tag{11}$$

$|p_1 - p_2|$ 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 Figure 5, our modified guided filter with different kernel sizes over the entire image performs well the trade-off between preserving the edges of the target image and preserving the shading distribution of the reference image.

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

3. Experiments

Relit results with adaptive parameters. Performance of our relighting method majorly relies on decomposition of lightness layer into large-scale and detail layer using WLS filter, and guided filtering process of the reference large-scale layer with the guidance of the target large-scale layer.

To validate the efficiency of the proposed method, we first check the function of our modification on WLS filter parameter choice. As shown in Figure 6, by setting larger λ in the facial hair and eyebrow regions, less reflectance information would retain in the large-scale layer and more reflectance information retain in detail layer, and the relit result can then retain more detail identification characteristics of the target face.

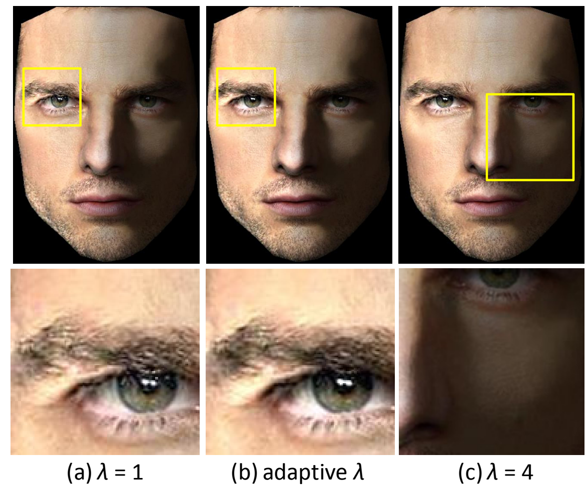

Figure 6. Our relit results with different WLS filter parameters. (a) is relit result when over all the image, the WLS filter parameter $\lambda = 1$; (b) is the relit result when the WLS filter parameter is set to spatially different by using our method; (c) is the relit result when over all the image, WLS filter parameter $\lambda = 4$. It's observed that the larger the value of λ is, the more details are retained on the final result, but the lighting effect is also weakened. Our strategy can ensure that the final relit result preserve well the details and generate well the shading effect at the same time.

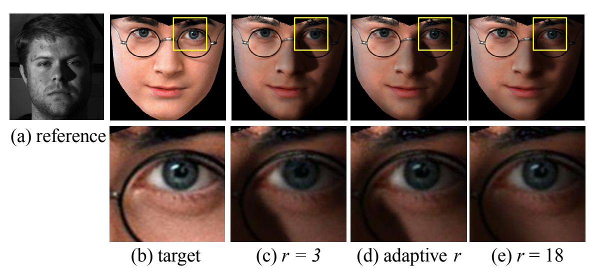

Figure 7. Our relit results with different guided filter parameters. (a) reference image, (b) target image, (c) the relit result for setting guided filter parameter $r = 3$ over all the image, (d) the relit result when the guided filter parameter r is set to be different spatially over all the image, (f) the relit result for the guided filter parameter $r = 18$ over all the image. It can be found that blurring can be seen around the eye structure in (c), shading contrast is weakened in (e), and our method in (d) can preserve the identification structure and retain the shading contrast at the same time.

We then check of the effect of our modification on guided filtering process. As shown in Figure 7, our method performs well in preserving the identification structure and retaining the shading of the reference object at the same time.

In Figure 8, we present some experiment results. The results show that our method can perform well in illumination transfer between genders, and also between color and grey images. More results are in the supplemental material.

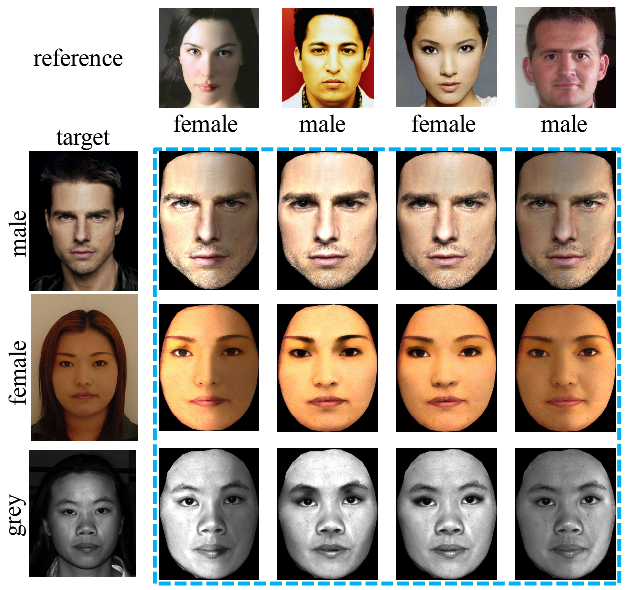

Figure 8. Our illumination transfer results between genders, and also between color and grey images.

Figure 9. Comparison with Li *et al.*[12]. Our method better retains the color information of the target face while the method of Li *et al.* transfers the reference color to the target face. Our method also maintains better the identification structure of the target face.

Comparison with the previous methods. We compare our method with the previous work. In [2], they relight a single target image under the frontal lighting condition by using two reference images (one is the reference image under the frontal lighting condition and the other is the reference image under the desired lighting condition). We use our method to relight a single image by using a single reference image. For a sake of a fair comparison, we also relight a single image by using a pair of reference images: first, a pair of reference images are divided pixel-to-pixel to get a quotient image, then the image warping technique is used to warp the reference quotient image according to the target image, and a rough relit result of the target image is acquired by multiplying the target image and the warped quotient image pixel-to-pixel, then the rough relit result is used as the input of our 2D-2D method.

As shown in Figure 10, the method in [2] over-stresses that the quotient image of the relit result and the original target image is locally scaled to warped reference quotient image; therefore, it introduces more reference shading details to the relit result. But the assumption that the target face and the reference face have similar geometry and reflectance usually could not be fulfilled in practice. Preserving more reference shading details of the target face may not ensure more naturalness and realness of the relit result. In contrast, according to our method, the image is decomposed into three layers, and only the large-scale layer of lightness is operated on so as to it avoid the introduction of much shading variance caused by the geometry difference of the reference and the target faces. In addition, our method with guided filter stresses that the relit result with the target image in local window can preserve the liner model and maintain better identification characteristics of the target face.

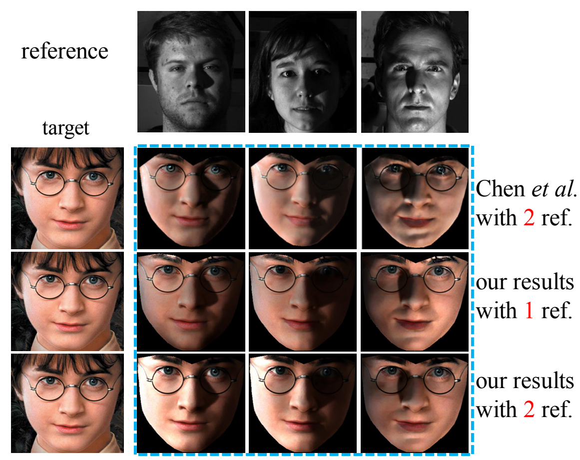

Figure 10. Comparisons with Chen *et al.*[2]. The first row is 3 reference images from YaleB [4] (it should be noted that according to the method in [2], one reference image under the frontal lighting condition and one reference image under the desired lighting condition are needed, and the reference images under the frontal lighting condition are not shown here), and the images in the first column are target images. The second row contains results of [2]. The third row contains our results using a single image as the reference image. The fourth row contains results of our method by using the rough results of quotient image based method as the reference images.

Figure 9 gives a comparison of our result with [12]. Unlike [12], in our method, only the lightness layer is operated on, leading to unchanged color of the relit result. Furthermore, by using edge-preserving filters in our method is introduced less shading details caused by geometry difference

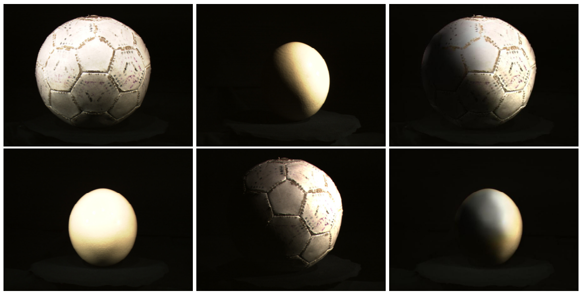

Figure 11. Illumination transfer between simple objects. ((a) and (b) are from the Amsterdam Library of Object Images [5])

between the target and the reference faces.

Relighting general objects. Our method can also be used to relight the general object when the reference object and the target object have similar geometry and reflectance. Figure 11 shows the results of illumination transfer between a ball and an egg. Mark points are marked by hand, and our method is then followed. It can be seen that the relighting result is natural and photo-realistic.

4. Conclusion and Discussions

In this paper, we have presented a novel image-based method for face illumination transfer through edge-preserving filters. We also propose adaptive parameter selection schemes in WLS filter and guided filter processes in face relighting application. The main advantage of our method lies in that only a single reference image is required with 3D geometry or material information of the target face. Convincing relit results demonstrate that our method is effective and advantageous in preserving the identification structure and skin color of the target face.

Limitation and future work. The ASM we adopt in our study could only locate frontal face well, we thus only test our method with both the target face and the reference face being frontal. And the ASM also fails to locate accurate mark points under hash lighting, and more manual manipulations are thus required. In future work, we would employ more effective methods to reduce the amount of manual correction. In our method, we make a trade-off between preserving the shading and the identification structures in face face structure regions. We would introduce learning based method to detect the shadows in human face for the parameter selection in the our relighting method. To overcome the assumption of similar shape between the reference and the target faces, we would explore a scheme for reference face selection in the future work.

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (90818003&60933006), National High-Tech (863) and Key Technologies R&D Programs of China (2009AA01Z331&2008BAH29B02), and Specialized Research Fund for the Doctoral Program of Higher Education (20091102110019). We would like to thank Qing Li for helpful suggestions, and Hongyu Wu for data processing.

References

[1] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. In *Proc. ACM SIGGRAPH*, 2009.

[2] J. Chen, G. Su, J. He, and S. Ben. Face image relighting using locally constrained global optimization. In *Proc. ECCV*, 2010.

[3] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *Proc. ACM SIGGRAPH*, 2008.

[4] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *TPAMI*, 23(6):643–660, 2001.

[5] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The amsterdam library of object images. *IJCV*, 61(1):103–112, 2005.

[6] D. Guo and T. Sim. Digital face makeup by example. In *Proc. CVPR*, 2009.

[7] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao. Lighting aware preprocessing for face recognition across varying illumination. In *Proc. ECCV*, 2010.

[8] K. He, J. Sun, and X. Tang. Guided image filtering. In *Proc. ECCV*, 2010.

[9] X. Jin, M. Zhao, X. Chen, Q. Zhao, and S.-C. Zhu. Learning artistic lighting template from portrait photographs. In *Proc. ECCV*, 2010.

[10] P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1):1–74, 2009.

[11] E. H. Land and J. J. McCann. Lightness and Retinex Theory. *Journal of the Optical Society of America (1917-1983)*, 61, Jan. 1971.

[12] Q. Li, W. Yin, and Z. Deng. Image-based face illumination transferring using logarithmic total variation models. *Vis. Comput.*, 26(1):41–49, 2009.

[13] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Proc. ECCV*, 2008.

[14] P. Peers, N. Tamura, W. Matusik, and P. E. Debevec. Post-production facial performance relighting using reflectance transfer. In *Proc. ACM SIGGRAPH*, 2007.

[15] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *Proc. CVPR*, 2008.

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

[17] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face re-lighting from a single image under harsh lighting conditions. In *Proc. CVPR*, 2007.