Facial performance illumination transfer from a single video using interpolation in non-skin region
Hongyu Wu, Xiaowu Chen*, Mengxia Yang and Zhihong Fang
State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China

ABSTRACT
This paper proposes a novel video-based method to transfer the illumination from a single reference facial performance video to a target one taken under nearly uniform illumination. We first filter the key frames of the reference and the target face videos with an edge-preserving filter. Then, the illumination component of reference key frame is extracted through dividing the filtered reference key frames by the corresponding filtered target key frames in skin region. The differences in non-skin region caused by different expressions between the reference and target face may bring about artifacts. Therefore, we interpolate the illumination component of the non-skin region by that of the surrounded skin region to ensure the spatial smoothness and consistency. After that, the illumination components of key frames are propagated to non-key frames to ensure the temporal consistency between the two adjacent frames. We obtain convincing results by transferring the illumination effects of a single reference facial performance video to a target one with the spatial and temporal consistencies preserved. Copyright © 2013 John Wiley & Sons, Ltd.

KEYWORDS
illumination transfer; facial performance; single video; propagation; interpolation

*Correspondence
Xiaowu Chen, State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China.
E-mail: chen@buaa.edu.cn

1. INTRODUCTION
Image/video-based face illumination transfer has been extensively studied in the computer community and has found wide application in film production, personal video editing, and so on. In film production industry, Peers et al. [1] used a special designed capture system to transfer illumination of a facial performance video to another one. The system contains 16 synchronized cameras and 150 lights. To separate the illumination components of the reference video, several frames of the same person with uniform illumination are captured. In daily life, the hand-held camera is available for common users and the illumination condition is not easily controlled. In this case, only a single reference video can be obtained and the frame with nearly uniform illumination may be unavailable. Transferring the illumination from only a single reference video to a facial performance video taken under nearly uniform illumination is still a challenging task.

Recently, several image-based illumination transfer methods are proposed. The quotient image based methods [2,3] use two reference images. The artist-drawn illumination templates are used for artistic illumination transfer [4]. Logarithmic total variation-based method [5,6] and edge-preserving filter based method [7] need only one reference image. However, repeating these single image-based methods frame by frame in facial performance video with various varying expressions may causes artifacts in non-skin region and cannot retain consistency between adjacent frames.

As Figure 1 shows, our purpose is to transfer the illumination from a single reference video (henceforth called reference video) to a facial performance video (henceforth called target video). The challenges of our task include the following: (i) The temporal consistency of illumination effect between the two frames must be retained. (ii) A frame with uniform illumination is probably unavailable in the reference frame. (iii) The facial expression in the target video is probably different from that of the reference video, leading to significant geometry differences in non-skin region (eyebrows, eyes and mouth).

Chen et al. [3] use local linear adjustments to remove the small-scale details in quotient image and global optimization to ensure the global consistency of quotient image. Chen et al. [7] use edge-preserving filter to extract the large-scale layer that can be considered as the
In this paper, we present a face video illumination transfer method. Firstly, the reference key frame is aligned to the target key frame and the two frames are filtered with the edge-preserving filter. Secondly, the illumination component of the key frame is extracted by the ratio between the two filtered key frames. The illumination transfer result is obtained by the product of illumination component and target frame. Thirdly, the illumination components of the key frames are propagated to the non-key frames.

Figure 1. Our illumination transfer results.

In this paper, we present a face video illumination transfer method. Firstly, the reference key frame is aligned to the target key frame and the two frames are filtered with the edge-preserving filter. Secondly, the illumination component of the key frame is extracted by the ratio between the two filtered key frames. The illumination transfer result is obtained by the product of illumination component and target frame. Thirdly, the illumination components of the key frames are propagated to the non-key frames.

Figure 2. The overview of our method.
Our main contributions include the following: (i) a framework of facial performance video illumination transfer from a single reference video and (ii) a face illumination transfer method for key frames.

2. RELATED WORK

**Illumination Transfer Using Multi-images.** Quotient image was introduced by Riklin-Raviv and Shashua [2]. The illumination component of a face image taken under specific illumination conditions is extracted by dividing it by the face image of the same person taken under fairly uniform illumination conditions. The extracted illumination component can be used to relight a face image of another person by image warping technologies. Chen et al. [3] use local linear adjustments and global optimization to ensure the smoothness and global consistency of the quotient image. The quotient image-based methods require two reference face images of the same person with similar expressions. In contrast, we only need a single reference key frame and extract the the reference key frame illumination component though the ratio between the filtered reference key frame and the corresponding target key frame with potentially different expressions.

**Illumination Transfer Using Single Image.** Li et al. [6] proposed an image-based method that employed the logarithmic total variation model [5,8] to factorize the target and reference images into illumination-dependent component and illumination-invariant component. The illumination transfer result is obtained by combining the illumination-dependent component of the reference image with the illumination-invariant component of the target image. Chen et al. [7] proposed an illumination transfer method based on edge-preserving filters. The WLS (Weighted Least Square) filter [9] is used to divide the target face image and the reference face image into the large-scale layer and the detail layer. Then, the guided filter [10] is used to filter the large-scale layer of the reference image while preserving the edges of the large-scale layer of the target image. Finally, the illumination transfer result is obtained by substituting the large-scale layer of the target image for the filtered large-scale layer of the reference image. In our method, we use the edge-preserving filter to smooth the key frames to remove the small-scale details.

**Illumination Transfer Using Templates.** Chen et al. [4] use hand-drawn templates for artistic illumination transfer. The templates are visually smooth because the facial details are removed by artists. Their method has good performance on artistic illumination transfer.

**Facial Performance Illumination Transfer.** Peers et al. [1] used reflectance acquisition system to capture the face reflectance database. Firstly, the reference reflectance filed is transferred to the target key frames. Then, the reflectance is propagated to the non-key frame. In this paper, we also propagate the illumination components of key-frames to non-key frames to maintain the temporal consistency of illumination effects.
3. OVERVIEW

Figure 2(a) shows the framework of our method. Repeating the illumination transfer algorithm on every frame will lead to temporal discontinuous of illumination effects [1,11], so we extract the illumination components of the key frames and the illumination components are propagated to non-key frames. The number of the reference frames is assumed to be equal to the number of the target frames. The reference and target key frames are selected at 10 frame intervals. Our transfer procedure is divided into two steps: (i) illumination transfer for key frames and (ii) illumination propagation from key frames to non-key frames.

The illumination transfer for key frames is illustrated in Figure 2(b). The reference key frame is aligned to the corresponding target key frame at first. Subsequently, with different strategies, we compute the illumination components in skin and non-skin regions. In skin region, the warped reference key frame and target key frame are smoothed by...
Figure 6. (a) The reference key frames, (b) target key frames, (c) illumination transfer results of Ref. [7], (d) illumination transfer results of Ref. [7] with non-skin region interpolation, (e) our results without non-skin region interpolation and (f) results with non-skin region interpolation.

the edge-preserving filter to ensure the spatial smoothness of illumination component. The illumination component is extracted by dividing each pixel in the filtered reference frame by the corresponding pixel in the filtered target frame. In non-skin region (mouth, eyes, and eyebrows), the illumination component is interpolated by that of the surrounded skin region to ensure the spatial consistency. Finally, the illumination transfer result of the key frame is generated as the product between the illumination component and the target key frame.

The optical flow is used for illumination propagation [1] to ensure the temporal consistency between the two adjacent frames. The forward and backward optical flows from current non-key frame to its two bounding key frames are computed. The illumination component of two bounding key frames are aligned to current non-key frame by the two optical flows. Then, the illumination component of current non-key frame is computed as the linear combination of the two aligned illumination component. Finally, the illumination transfer results of the non-key frame is generated as the product between the combined illumination component and current non-key frame.

4. ILLUMINATION TRANSFER FOR KEY FRAMES

4.1. Skin Region Transfer

The faces in the target video and the reference video may have significantly different facial geometrical features, expressions, and poses. Thus, we use multilevel free-form deformation [12] to align the reference key frame to the target key frame. We use active shape model [13] to detect the feature points at the first key frame and refine their accurate position manually. After that, the feature points are propagated to the following key frame by the optical flow [14] from the current frame to the next. Whereas the propagation may be inaccurate due to the image noise, so we manually refine the points after each propagation procedure.

Although the reference key frame is aligned to the target key frame, the small-scale details of the target key frame and that of the warped reference key frame are not identical. The illumination component directly extracted through the ratio between the two key frames contains noticeable noise (Figure 3(c)). We use edge-preserving filter [15] to remove the small-scale details of the target key frame and warped reference key frame. Figure 3 shows that the illumination component is smoother after edge-preserving filtering.

4.2. Non-Skin Region Interpolation

The significant geometrical differences in non-skin region (eyes, eyebrows, and mouth) between the warped reference key frame and the target key frame may result in artifacts in the illumination components (Figure 4(c)), leading to the spatial inconsistency in the illumination component. The ‘ideal’ quotient image obtained in Ref. [1] is visually consistent all over the image. Therefore, we interpolate the illumination component in non-skin region to ensure the visual consistency. We triangulate those regions according to the feature points. The triangular mesh is illustrated in Figure 4(a). The pixels on the mesh lines are denoted as $C_I(\cdot)$ and the pixels in the triangles are denoted as $C_I(\cdot)$. Figure 4(b) illustrates the interpolation procedure; $a$, $b$, and $c$ are the vertexes of a triangle, $\Omega$ denotes the boundary of a region (the green line of Figure 4(a)), $d(\cdot, \cdot)$ denotes the Euclidean distance between two points, the value $C_I(e)$ of a pixel $e$ on the mesh line is computed using Equation (1):

$$C_I(e) = \begin{cases} 
C_I(e) & \text{if } e \in \Omega \\
\frac{C_I(c)d(b,e)+C_I(b)d(e,c)}{d(b,c)} & \text{else}
\end{cases}$$

(1)
Figure 7. Illumination transfer results on key frames. (a) The reference face video is captured in a room with little ambient light and the reference face is lit by a spot light source. (b) The reference face video is captured in a room lit by a fluorescent lamp as ambient light and the reference face is lit by a surface light source.

The value $C_t(f)$ of a pixel $f$ in the triangle is computed using Equation (2):

$$C_t(f) = \frac{C_t(d)(f,e) + C_t(e)d(d,f)}{d(d,e)} \tag{2}$$

Figure 4(c) and (d) shows the comparison of illumination component before and after interpolation. As is illustrated in Figure 4, our non-skin region interpolation procedure maintains the spatial consistency of the illumination in the eyebrows, eyes, and mouth regions.

5. ILLUMINATION PROPAGATION FOR NON-KEY FRAMES

In the previous section, the illumination components of all reference key frames are computed. To maintain the temporal consistency of the illumination effects, the illumination components of the key frames are propagated to the non-key frames. The optical flow estimation method [14] is used to build the correspondences between two frames [1].
Figure 8. Illumination propagation from key frames to non-key frames. The left-most and right-most columns are the illumination transfer results of key frames. The middle columns are the illumination propagation results from key frames to non-key frames.

Denote the current non-key frame index by \( t \). The illumination component of the two bounding key frames are warped to current non-key frame by forward and backward optical flow. The illumination component \( Q_{n,k}(t) \) of current non-key frame is computed using Equation (3):

\[
Q_{n,k}(t) = w_p Q_{w,k}(t_p) + w_f Q_{w,k}(t_f) \\
w_p = (t_f - t)/(t_f - t_p) \\
w_f = 1 - w_p
\]

where \( t_p \) and \( t_f \) denote the previous key frame index and the following key frame index. \( Q_{w,k}(t_p) \) and \( Q_{w,k}(t_f) \) denote the warped illumination component of the previous and the following frames. Figure 5 shows that the illumination propagation can maintain the temporal consistency of the illumination effects.

6. EXPERIMENTS

Illumination Transfer Results with Non-skin Region Interpolation. We interpolate the illumination component in non-skin region to ensure the spatial consistency. To validate the performance of the non-skin region interpolation, we first compare the illumination transfer results with the most recent state-of-the-art illumination transfer method from single reference image [7]. Their method cannot deal with the expression differences between the reference and the target video. Figure 6(c) illustrates the illumination transfer results using Ref. [7]. In Figure 6(d), the large-scale layer of Ref. [7] in non-skin regions is interpolated by our method. Chen et al. [7] use edge-preserving filter to extract the large-scale layer of reference image and target image. The illumination transfer result is generated by replacing the large-scale layer of the target face image with that of the reference image. Although their method does not take the expression differences into account, we generate the illumination component by the ratio of the filtered reference and target frames, and the illumination component can be interpolated in non-skin region. We first compare the illumination transfer results before and after the illumination component interpolation in the non-skin region. The comparison of Figure 6(e) and (f) shows that...
the non-skin region interpolation obviously reduces the artifacts especially in the mouth region.

**Illumination Transfer on Key Frames.** In Figure 7(a), the illumination effects contrast of reference face is relatively high, whereas illumination effects contrast in Figure 7(b) is relatively low. Experimental results show that our method performs well on face video illumination transfer. The illumination transfer result video has been included in the attachment.

**Illumination Propagation.** The illumination components are propagated from key frames to non-key frames. Figure 8 shows that the propagation procedure preserves the consistency between the two frames.

**Color Illumination Transfer.** Our method can transfer color illumination from reference video to target video. Figure 9 shows some color illumination transfer results. A green light source is moving on the left of the reference face, whereas a fuchsia light source is moving on the right of the reference face. The illumination effects on the reference face caused by the color light source is transferred to the target face.

**7. CONCLUSION AND DISCUSSION**

We propose a method for face video illumination transfer. The main advantage of our method lies in the fact that only a single reference video is required. The illumination component can be obtained more robustly when only skin region is considered. Illumination of non-skin regions are obtained using interpolation. After that, the illumination component of key frames is propagated to non-key frames. Experimental results show that our method is capable of generating convincing illumination transfer results.

**Limitation and Future Work.** In the case that the skin color and large-scale face geometry in the target and the reference video are not similar, our method does not work well. Our method assumes that the reference video and target video have the the same amount of frames that is not a practical assumption. The synchronization between the reference video and target video will be considered in the future work. Our method is purely an image-based illumination transfer method. The specular reflection effects on the face cannot be transferred to the target face video completely. In the future, we will use the facial geometrical information to enhance the specular reflection effects to generate more photo-realistic illumination effects. The landmarks of key frame are manually adjusted. This is prohibitive for a long video. In the future, we will employ some effective methods to reduce the manual interaction.

**ACKNOWLEDGEMENTS**

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), R&D Program (2012BAH07B01), and ITER (2012GB102008).

**REFERENCES**

13. Milborrow S, Nicolls F. Locating facial features with an extended active shape model, In *Proceedings of
the 10th European Conference on Computer Vision: Part IV, Marseille, France, Springer-Verlag, 2008; 504–513.


AUTHORS’ BIOGRAPHIES

**Hongyu Wu** is presently a PhD candidate in the School of Computer Science and Engineering and works in the State Key Laboratory of Virtual Reality Technology and Systems at Beihang University. His research interests are computer vision and video processing.

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

**Mengxia Yang** is presently a Master candidate in the School of Computer Science and Engineering at Beihang University. Her research interests are computer graphics and large-scale light filed.

**Zhihong Fang** is presently a Master candidate in the School of Computer Science and Engineering at Beihang University. His research interests are computer vision and image processing.