Method Article

A method of single reference image based scene relighting

Xin Jin, Yannan Li, Ri Xu, Xiaokun Zhang, Xiaodong Li*

Department of Computer Science and Technology, Beijing Electronic Science and Technology Institute, Beijing, 100070, PR China

A B S T R A C T

Image relighting is to change the illumination of an image to a target illumination effect without known the original scene geometry, material information and illumination condition. We propose a novel outdoor scene relighting method, which needs only a single reference image and is based on material constrained layer decomposition. Firstly, the material map is extracted from the input image. Then, the reference image is warped to the input image through patch match based image warping. Lastly, the input image is relit using material constrained layer decomposition. The experimental results reveal that our method can produce similar illumination effect as that of the reference image on the input image using only a single reference image.

© 2018 Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

A R T I C L E  I N F O

Method name: Single reference image based scene relighting

Keywords: Image relighting, Single reference image, Material map, Layer decomposition

Article history: Received 14 December 2017; Accepted 6 August 2018; Available online 13 August 2018

Specifications Table

| Subject area | Computer Science |
|--------------|------------------|
| More specific subject area | Computer Vision, Image Relighting |
| Method name | Single Reference Image based Scene Relighting |
| Name and reference of original method | None |
| Resource availability | coming soon |

* Corresponding author.

E-mail address: zuigueili@gmail.com (X. Li).

https://doi.org/10.1016/j.mex.2018.08.004

2215-0161/© 2018 Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Method details

Our proposed method can be divided into 4 steps, as shown in Fig. 1 (1) the input image is segmented to the material map using the method of Bell et al. [1]. Every pixel of the material map is assigned by a material label; (2) the reference image is warped to the structure of the input image by the patch match warping; (3) each channel of the input image and the reference is decomposed to large-scale layer and detail layer under material constrain; (4) the final relit results are obtained by composing the details of the input image and the large-scale layer of the warped reference image. Different from some methods that rely on inferred geometry [2,3], the proposed method does not rely on inference geometry.

The input image is segmented according to the material of each pixel. We use the method of Bell et al. [1] to obtain material label of each pixel. We make the material segmentation because that in different material region, different relighting operations should be conducted. We select 9 sorts of materials, which often appear in outdoor scene images, as shown in Fig. 2. We recolor each pixel according to the material label to get the material map. The first and the third lines are the input images. The second and the forth lines are the corresponding material maps.

In face image relighting, the reference face image can be warped by face landmark detection/face alignment. However, in outdoor scene, we cannot find such similar structure easily. The outdoor scene contains multiple objects. Thus, we use the patch match method to warp the reference image to the input image, i.e. to align the reference and the input image. The patch match algorithm is similar as the method of Barnes et al. [5]. We use the neighbor patches whose best matched patches have already been found to improve matching result of current patch. The difference from Barnes et al. [5] is that we use 4 neighbor patches instead of 3 ones.

The basic idea is to find the most similar patch in the reference image to substitute the original patch in the input image. Two basic assumptions are made: (1) the matched patches of the neighbor patches in the input image are mostly neighbor; (2) large scale random search region may also contain matched patch.

We denote the input image as $A$ and the reference image as $B$. The coordinate of a patch is represented as coordinate of the left up corner of the patch. The Nearest Neighbor Field (NNF) is defined as $f$, whose definition domain is the coordinates of all the patches in $A$. The value of the NNF is the offset of the coordinate of matched patch in $B$. We denote the coordinate of the original patch in $A$ as $a$ and the coordinate of the matched patch in $B$ as $b$, then:

$$f(a) = b - a$$  \hspace{1cm} (1)
The distance between the original patch and the matched patch is defined as \( D(v) \), which describes the distance between the patch \( a \) in \( A \) and patch \( a + v \) in \( B \). The distance is computed by the Euclidean distance \([6]\). The warping method contains three steps: initialization, propagation and random search.

- **Initialization.** The initial offset of each patch in \( A \) is randomized around the patch.
- **Propagation.** As assumed above, the matched patches of the neighbor patches in the input image are mostly neighbor. We use the neighbor patches whose best matched patches have already been found to improve matching result of current patch. The \( f(x-1, y), f(x, y - 1) \) and \( f(x-1, y - 1) \) are used:

  \[
  f(x, y) = \min \{ D(f(x, y)), D(f(x-1, y)), D(f(x, y - 1)), D(f(x-1, y - 1)) \} \tag{2}
  \]

- **Random Search.** As assumed above, large scale random search region may also contain matched patch. We use a search window whose size is declined exponentially.

  \[
  u_i = v_0 + \omega \alpha^i R_i \tag{3}
  \]

where \( v_0 = f(x, y) \), \( R_i \) is a random point in \([1,1][-1,1] \), \( \omega \) is the max search radius. \( \alpha \) is the declining rate of the radius. The warped results of some reference images are shown in Fig. 3.

We use the WLS filter \([7]\) to decompose image into large-scale layer and detail layer, which can be considered as the illumination component and non-illumination component. Using the large-scale layer of the warped reference to substitute the large scale layer of the input can produce the final relit result. The outdoor scene contains various objects with various materials. Thus for different material, different decomposition parameters should be used. Each channel \( l \) of the input image and the

\[
\text{cloud} \quad \text{sky} \quad \text{architecture} \quad \text{tree} \quad \text{grass} \quad \text{water} \quad \text{ground} \quad \text{other} \quad \text{light source}
\]

Fig. 2. The material maps of the input images.
reference image is filtered to a large-scale layer $s$. The detail layer $d$ is obtained by:

$$d = l/s$$  \hspace{1cm} (4)

The original WLS filter uses the same smoothness level over the whole image. When using the WLS filter for our scene relighting task, we need make regions with different materials with different smooth levels. Thus, we set different smoothness levels in regions with different materials. We modified the original WLS [7] as:

$$E = ||l - s||^2 + H(\nabla s, \nabla l)$$  \hspace{1cm} (5)

$$H(\nabla s, \nabla l) = \sum_p \left( \lambda(p) \left( \frac{(\partial s/\partial x)^2}{(\partial l/\partial x)^2 + \varepsilon} + \frac{(\partial s/\partial y)^2}{(\partial l/\partial y)^2 + \varepsilon} \right) \right)$$  \hspace{1cm} (6)

where, $||l - s||^2$ is the data term, which is to let $l$ and $s$ as similar as possible, i.e., to minimize the distance between $l$ and $s$. $H(\nabla s, \nabla l)$ is the regularization (smoothness) term, which makes $s$ as smooth as possible, i.e. to minimize the partial derivative of $s$. $p$ is the pixel of the image. $\alpha$ controls over the affinities by non-linearly scaling the gradients. Increasing $\alpha$ will result in sharper preserved edges. $\lambda$ is the balance factor between the data term and the smoothness term. Increasing $\lambda$ will produce

---

**Fig. 3.** Patch match based scene image warping.

**Fig. 4.** The relit results of our method.
smoother images. \( \varepsilon \) is a very small number, so as to avoid the division by 0. Our \( \lambda \) is the smoothness level constrained by different materials, using the material map (Figs. 4 and 5):

\[
\lambda = \nabla l + \text{gray}(l_m)/255
\]  

(7)
where, $\nabla l$ is the gradient of $l$. $l_m$ is the material map of $l$, and the gray is the gray value of $l_m$:

$$\text{gray} = (R \times 0.2989 + G \times 0.587 + B \times 0.114)$$

(8)

The minimization of Eqs. (1) and (2) can be solved by the off-the-shell methods such as Lischinski [6]. At last, using the large-scale layer of the warped reference to substitute the large-scale layer of the input can produce the final relit result.

**Comparison with other methods**

We compare our method with the state of the art method [4], which needs a time-lapse video captured by a fixed camera working for 24 h. While our method needs only a single reference image. We randomly select 5 input images for comparison. As shown in Fig. 6, (a): multiple input images, (b) the reference image, (c): warped reference images to corresponding input images, (d): warped reference images using method of [4], note that they need a time-lapse video for warping, (e): the relit results using our proposed method, (f): the relit results using the method of [4]. The results reveal that our method can produce similar relit results as those of [4], with only a single reference image.

**References**

[1] S. Bell, P. Upchurch, N. Snavely, K. Bala, Material recognition in the wild with the materials in context database, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015 (2015) 3479–3487.
[2] J.T. Barron, J. Malik, Shape, illumination, and reflectance from shading, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1670–1687.
[3] B. Henz, M.M. Oliveira, Artistic relighting of paintings and drawings, Vis. Comput. 33 (1) (2015) 1–14.
[4] Yichang Shih, Sylvain Paris, Frédéric Durand, William T. Freeman, Data-driven hallucination of different times of day from a single outdoor photo, ACM Trans. Graph. 32 (6) (2013) 11 Article 200.
[5] C. Barnes, E. Shechtman, D.B. Goldman, A. Finkelstein, The generalized patchmatch correspondence algorithm, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Proceedings of the 11th European conference on computer vision conference on Computer vision: Part III (ECCV’10), Springer-Verlag, Berlin, Heidelberg, 2018, pp. 29–43.
[6] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[7] Z. Farbman, R. Fattal, D. Lischinski, R. Szeliski, Edge-preserving decompositions for multi-scale tone and detail manipulation, ACM Trans. Graph. (Proceedings of ACM SIGGRAPH) 27 (3) (2008).