DEPTH OF FIELD GUIDED REFLECTION REMOVAL

Renjie Wan, Boxin Shi, Tan Ah Hwee and Alex C. Kot*

Nanyang Technological University, Singapore, 639798
rwan001@e.ntu.edu.sg, {asahtan,eackot}@ntu.edu.sg

ABSTRACT

Reflection removal aims at separating the mixture of the desired scene and the undesired reflections. Locating reflection and background edges is a key step for reflection removal. In this paper, we present a visual depth guided method to remove reflections. Our idea is to use Depth of Field (DoF) to label the background and reflection edges. We propose a DoF confidence map where pixels with higher DoF values are assumed to belong to the desired background components. Moreover, we observe that images with different resolutions show different properties in the DoF map. Thus, we introduce a multi-scale DoF computing strategy to classify edge pixels more efficiently. Based on the results of edge classification, the background and reflection layers can be separated. Experimental results validate the effectiveness of our method using real-world photos.

***Index Terms*—** Depth of field, reflection removal, layer separation

1. INTRODUCTION

When taking photos through transparent surfaces, the captured pictures are usually a mixture of both the desired scene and undesired reflections. Reflection removal aims at solving this problem. As a subfield of the blind source separations, it can be expressed by the following linear model:

$$I = I_B + I_R, \tag{1}$$

where I_B and I_R are background and reflection images, respectively.

Our goal is to remove the reflections I_R and to recover the background I_B. Obviously, additional information or constrains are required to solve this problem. Different priors were introduced in previous research works. Existing methods on reflection removal can be classified into three categories based on the number of images used.

The first category aims at removing reflections from a single image. Levin *et al.* [1] proposed an algorithm using image priors including statistics of derivative filters and corner detectors in natural images to decompose the image into two different layers. Later, they modified their algorithm by imposing a gradient sparsity prior on the recovered layers. Recently, Li *et al.* [2] introduced a method under the assumption that the background is in focus and the reflection is blurrier. However, even with these priors, single image reflection removal is still a very challenging problem.

The second category focuses on removing reflections from two images. For example, Agrawal *et al.* [3] introduced a method to reduce the reflection effect by taking a flash and no-flash image pair. Farid *et al.* [4] and Schechner *et al.* [5] tried to separate the reflections with two images taken by rotating a polarizer with different angles and then linearly combine the two images. Although this kind of approaches can efficiently suppress the reflections, special skills and devices are required when capturing the input data [4, 5, 3].

The third category uses an image sequence as the input. Gai *et al.* [6] assumed that the motion of each layer follows an affine transformation and a new image prior based on joint relationship of both background and reflection gradient is adopted to separate reflections. Li *et al.* [7] used SIFT flow to calculate the motion field and register images. Guo *et al.* [8] assumed the motion of each layers is a homography, and a rank minimization method is applied to solve this problem robustly. However, the requirement for the motion of camera limits the practicability of these approaches for dynamic scene.

Our method belongs to the first category and the basic pipeline is illustrated in Fig. 1. The key assumption in our paper is that the photographers always focus on the background in a particular depth when they take photos. Reflections in different depth layers would be blurred in the images. Thus, DoF, the distance between the nearest and farthest objects in a scene that appears reasonably sharp [9], can be used as an important feature to distinguish background and reflection edges. The existing method [9] can compute a DoF value for a whole image to evaluate its blurred degrees. Inspired by [9], we propose a DoF confidence map computing strategy to evaluate the blurred degrees for all pixels. We also observe that images with different resolution can exhibit different levels of details in the DoF map. Combining the assumption and observation, we develop a multi-scale inference scheme to select background and reflection edges to guide the reflec-

*Boxin Shi is now affiliated with Artificial Intelligence Research Center, National Institute of AIST, Tokyo, Japan (boxin.shi@aist.go.jp).

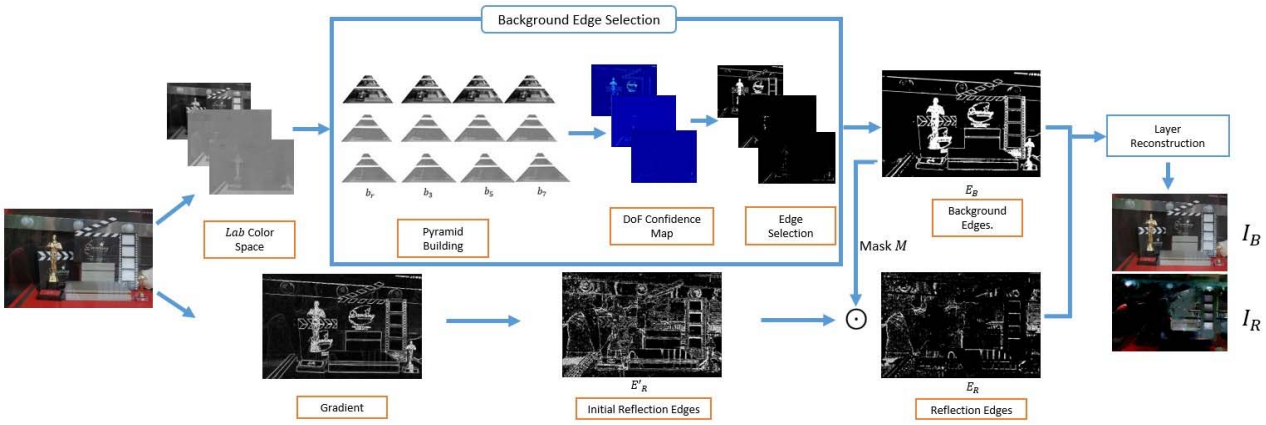

Fig. 1. The pipeline of our method. For the background edge selection, the input image is first converted to the Lab color space. For each channel, we build one reference pyramids and three blurred pyramids. Then a DoF confidence map for each channel is computed. Finally, E_B are selected based on the confidence map. For the reflection edge selection, we compute the gradient of the input image to get the initial reflection edges. Based on the initial reflection edges and background edges obtained before we can get E_R. With the two sets of edges, I_B and I_R can be separated. We multiple I_R by 10 for better visualization

tion removal process. With the selected edges, the classical approach in [10] can be directly applied for layer separation. Compared with the state-of-the-art methods (*e.g.* [2, 7]), our method shows better separation results.

2. PROPOSED METHOD

2.1. Background Edge Selection

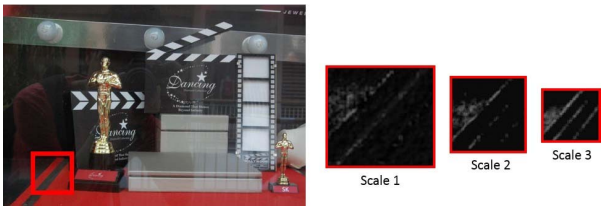

Fig. 2. DoF confidence map obtained from Scale 1 to Scale 3

In this section, we describe the multi-scale inference scheme to extract background edges. Based on our observations, it may not be accurate to know whether a region belongs to the background or not from only one resolution. From Fig. 2, it can be seen that the DoF map of Scale 3 exhibits more details of the background than in Scale 1. The scale ambiguity has been studied in many applications [11]. We resort to an image pyramid model to fuse information from different levels.

2.1.1. *Construction of Image Pyramids*

The pyramids are built as follows. First, we build a reference pyramid b_r with 3 layers where each layer is down-sampled by a factor to form the next coarser level. The resolution of the first layer is the same as that of the input image. Second, based on b_r, we build 3 pyramids which is denoted as $\{b_3, b_5, b_7\}$. The subscript here is the blurring kernel size of $k \times k$ ($k = \{3, 5, 7\}$) for that pyramid. It means that each layer of the three pyramids is a blurred version of the corresponding layer of the reference pyramid, which can be expressed as follows:

$$b_k = b_r * f_k, \tag{2}$$

where $*$ is a 2D convolution operator and f_k is a Gaussian blurring kernel. After this step, four pyramids have been created for the next step which is (b_r, b_3, b_5, b_7). The pyramids for other channels can be computed in a similar way.

2.1.2. *Pyramid Fusion*

For the four image pyramids obtained before, we use the method proposed in [9] to calculate the KL divergence between the corresponding layers of b_k and b_r after the vertical and horizontal derivatives are applied on each layer. The result is denoted as $\{D_k^n\}$ where n is the level number from 1 to 3. Unlike [9] which obtains a DoF value for a whole image, we compute a DoF confidence map for images with different resolution as

$$D^n(i,j) = \sum_{k=\{3,5,7\}} D_k^n(i,j), n = \{1,2,3\}\,. \tag{3}$$

We call the DoF maps with different scales generated by Eq. (3) as DoF pyramids, which is $\{D^1, D^2, D^3\}$, respectively.

Algorithm 1 Multi-scale Background Edge Extraction

Input:
Input image I;
Output:
Background edge map E_B;
for $m = 1$ to 3 (corresponding to $\{L, a, b\}$) **do**
  Build one reference image pyramid $\{b_r\}$with level index $\{1, 2, 3\}$;
  Build 3 image pyramids $\{b_3, b_5, b_7\}$with level index $\{1, 2, 3\}$ based on b_r (Eq. (2)) ;
  Compute KL divergence between $\{b_3, b_5, b_7\}$ and b_r using the method in [9];
  Build the DoF pyramid (Eq. (3));
  Combine multi-scale DoF maps (Eq. (4) and Eq. (5));
  if $m = 1$ **then**
    Compute τ_s using Eq. (6);
  else
    $\tau_s = \tau_s/1.5$;
  end if
  Select salient edges E_b (Eq. (5));
end for
$E_B = \vee_{m=1}^{3} E_b^m$;
return E_B;

Combining the 3 maps with different scales, we define our final DoF confidence map D as

$$D = (\lambda \cdot D^2 \uparrow + (1 - \lambda) \cdot D^3 \uparrow) \odot (D^1), \tag{4}$$

where $\odot$ is the elementary multiplication and $\uparrow$ indicates that D^2 and D^3 are upscaled to the same size of D^1. The DoF confidence maps in three channels are also shown in Fig. 1, which exhibit very large difference in reflection and background areas and can help us to distinguish the two components.

Then, we rule out the pixels belonging to reflection components in the DoF maps and the salient edges for layer reconstruction are determined by

$$E_b = H(D - \tau_s), \tag{5}$$

where H is the Heaviside step function, generating zeros for negative values and ones for positive values. τ_s is a threshold to determine the salient edges belonging to the background components.

Similar strategy can be used in computing E_b in three color channels. For different channels of Lab, to infer subtle structures during this process, we decrease the value of τ_s in each iterations to include more details of the background. Instead of using a constant threshold, we use the DoF maps in L channel to determine the initial threshold value. This is similar to the adaptive threshold proposed by Hou *et al.* [12]. In our case, the adaptive threshold is the mean confidence value shown below,

$$\tau_s = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} D(i, j), \tag{6}$$

where W and H are the width and height of the confidence map in pixels, respectively.

Finally, the edge map of the background can be generated as follows:

$$E_B = \vee_{m=1}^{3} E_b^m, \tag{7}$$

where $\vee$ denotes logical *or* and m is the channel number corresponding to $\{L, a, b\}$, respectively. At last, the background edge maps obtained are illustrated in Fig. 1. This is a binary mask for the background component.

2.2. Reflection Edge Selection

Compared with the background, most reflection layers are related to smaller gradient magnitudes. Thus, in the gradient domain , we can obtain an initial reflection edge map based on the following threshold:

$$E'_R(i) = \begin{cases} 1, & \text{if } \tau_{r1} < g(i) < \tau_{r2} \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where $g(i)$ is the gradient value of the input mixture image on pixel i. The initial reflection edge map shown in Fig.1 includes some misclassified background edges. Having created the map indicating the regions of background components, we now use it to refine E'_R. To remove more tiny details from the background components, we create a mask by using an appropriate structuring element S to dilate the background edge map as

$$M = E_B \oplus S. \tag{9}$$

Then we can reduce the artifacts in the initial reflection edge map as follows:

$$E_R = \overline{M} \odot E'_R, \tag{10}$$

where $\overline{M}$ denotes not operation over M. Finally, the calculated reflection edge map E_R is shown in Fig. 1.

2.3. Layer Reconstruction

With the background and reflection edge maps generated before, the reflection and background layers can be separated based on the objective function proposed by Levin *et al.* [10] in this step, which is as follows:

$$\begin{aligned} J(I_B) = & \sum_{i,k} \rho(f_{i,k} \cdot I_B) + \rho(f_{i,k} \cdot (I - I_B)) \\ & + \gamma \sum_{i \in E_B, k} \rho(f_{i,k} \cdot I_B - f_{i,k} \cdot I) \\ & + \gamma \sum_{i \in E_R, k} \rho(f_{i,k} \cdot I_B), \end{aligned} \tag{11}$$

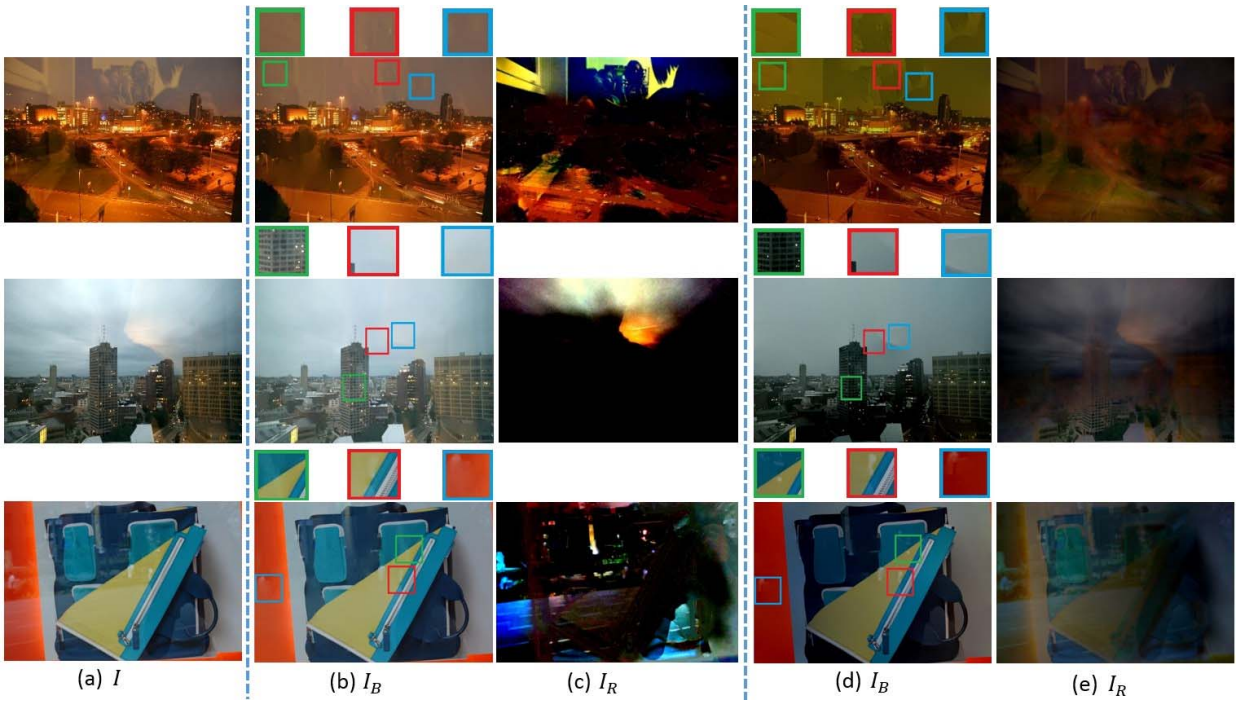

Fig. 3. Three examples of reflection removal results of our method and the one in [2]: (a) Input Images. (b) Background Images by our method. (c) Reflection Images by our method. (d) Background Images by [2]. (e) Reflections Image by [2]

where $f_{i,k}$ is the k-th derivative filter. E_B and E_R are two sets of background and reflection edges obtained before, respectively. The first term ensures the sparsity of gradients of the two layers. The last two terms enforce the agreement with the labeled gradients. More details can be found in [10].

3. EXPERIMENTS

To show the performance of our method, we compare our method with the method proposed in [2], which also uses only a single image as input. For the threshold values τ_{r1} and τ_{r2}, we set them to 0.1 and 0.3, respectively. λ in Eq. (4) is set to 0.4 in our experiment. These parameters are set empirically. The images in our experiments are from [7, 13] and Internet.

We show three examples in Fig. 3. As we can observe, our method can generate a clear separation with much less residual. Considering the regions highlighted by rectangles of Fig. 3, our algorithm can remove the majority of reflections which is better than the ones generated by [2] where the results still contain visible residual edges. Moreover, in the second row of Fig. 3(d), the reflections are not removed but the details in the non-reflection areas are over-smoothed. On the other hand, the method in [2] causes a color change. From the three results generated by [2], we can find that the results are darker than the original image and this phenomenon is very obvious in the second and third row of Fig. 3(d).

4. CONCLUSION

A method to remove reflections with a single image is proposed. With the multi-scale inference scheme proposed here, we can generate the background edge map. Using this map as a mask, we introduce a framework to refine the initial reflection edges. This framework can well classify the background and reflection edges. With the two well classified edge maps, our experimental results show the effectiveness of our method. However, our method cannot deal with image containing tiny artifacts or details including the small texture in some objects and something like grass. This is an ubiquitous problem existed in the methods based on the algorithm proposed in [10]. This problem will be taken into account in our future study.

5. ACKNOWLEDGEMENT

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Singapore, under its Interactive Digital Media (IDM) Strategic Research Programme. Boxin Shi is supported by a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

6. REFERENCES

[1] A. Levin, A. Zomet, and Y. Weiss, "Learning to perceive transparency from the statistics of natural scenes," in *Advances in Neural Information Processing Systems (NIPS), 2002*, pp. 1247–1254.

[2] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. of Computer Vision and Pattern Recognition (CVPR), 2014*, pp. 2752–2759.

[3] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," in *ACM Trans. on Graphics*, vol. 24, no. 3, 2005, pp. 828–835.

[4] H. Farid and E. H. Adelson, "Separating reflections from images by use of independent component analysis," *Journal of the Optical Society of America*, vol. 16, no. 9, pp. 2136–2145, 1999.

[5] Y. Y. Schechner, J. Shamir, and N. Kiryati, "Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface," in *Proc. of International Conference on Computer Vision (ICCV), 1999*, vol. 2, pp. 814–819.

[6] K. Gai, Z. Shi, and C. Zhang, "Blind separation of superimposed moving images using image statistics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 19–32, 2012.

[7] Y. Li and M. S. Brown, "Exploiting reflection change for automatic reflection removal," in *Proc. of International Conference on Computer Vision (ICCV),2013*, pp. 2432–2439.

[8] X. Guo, X. Cao, and Y. Ma, "Robust separation of reflection from multiple images," in *Proc. of Computer Vision and Pattern Recognition (CVPR),2014*, pp. 2195–2202.

[9] O. Le Meur, T. Baccino, and A. Roumy, "Prediction of the inter-observer visual congruency (iovc) and application to image ranking," in *Proceedings of the 19th ACM International conference on Multimedia, 2011*, pp. 373–382.

[10] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1647–1654, 2007.

[11] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," in *Proc. of Computer Vision and Pattern Recognition (CVPR), 2014*, pp. 2965–2972.

[12] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. of Computer Vision and Pattern Recognition (CVPR), 2007*, pp. 1–8.

[13] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, "A computational approach for obstruction-free photography," *ACM Trans. on Graphics*, vol. 34, no. 4, p. 79, 2015.