

# A Saliency-Guided Method for Automatic Photo Refocusing

Na Liu, Ran Ju, Tongwei Ren and Gangshan Wu

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China  
liunana1993@gmail.com, juran@smail.nju.edu.cn, {rentw, gswu}@nju.edu.cn

## ABSTRACT

With the prevalence of smart phones and pocket cameras, photo refocusing has become a basic editing and processing method for its power in interesting object emphasis and photo beautification. However, existing image refocusing methods can hardly be applied on mobile devices due to its high computational cost or the dependence on expensive hardware like light field camera. In this paper, we present a simple but effective method to perform image refocusing automatically and rapidly. The key of our method lies in the utilization of the characteristics of human visual systems. By leveraging current saliency detection methods, we locate the region of interest for a given photo rapidly. Then we calculate its depth map according to the frames captured before shooting. The original image is softly segmented into layers and blurred with different confusion sizes according to the depth map. At last, the blurred layers are softly combined into a refocused photo. Experimental results demonstrate that our method performs outstandingly both in automatic photo refocusing and computational complexity.

## Categories and Subject Descriptors

I.4.3 [Image Processing and Computer Vision]: Enhancement; I.4.9 [Image Processing and Computer Vision]: Applications

## Keywords

automatic photo refocusing, depth of field, salient object detection

## 1. INTRODUCTION

Nowadays, portable devices such as smart phones are preferred than clunky DSLRs in photographing for their convenience, improving lens and sufficient after effects [1]. Image refocusing, as one of the most popular tools for photo editing, generates photographs focused on different objects [2, 3, 4] with variable depth-of-fields [5, 6]. Fig. 1

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICIMCS '16, August 19-21, 2016, Xi'an, Shaanxi, China

© 2016 ACM. ISBN 978-1-4503-3528-7/15/08...\$15.00

DOI: [http://dx.doi.org/\\*\\*\\*\\*\\*](http://dx.doi.org/*****)



Figure 1: Examples of photo refocusing. To (a) and (b), (Left) Original photos; (Middle) Saliency maps; (Right) Refocusing results.

shows two examples of photo refocusing. The left of the examples are original photos, the middle are corresponding saliency maps which high-light the interesting region of photos, and the right are photos refocused on the interesting region. By photo refocusing, important regions are emphasized to capture viewer's attention [2]. Meanwhile, properly refocused photos offer more powerful aesthetic expression than all sharp ones. In fact, the technique has been, and continues to be widely applied in digital photography, even before digital photography age [7].

Current techniques for image refocusing [8] commonly rely on intense computation and human intervention, thus difficult to be applied on portable devices. To satisfy the user demand for rapid on-portable-device image refocusing, we propose a novel approach featured by low computation cost and outstanding depth-of-field blurring effect. Our method enables portable devices to generate refocused images in real time. Specifically, given an image with the frames taken before shooting, we firstly extract the depth structure and the saliency map. Then, the input image is segmented into salient objects and background components. Afterwards, a depth-aware blurring algorithm is performed on each layer. The salient region is kept unchanged while the background regions are blurred with different confusion kernels according to their depth. Finally, all the layers are synthesized into a complete image which focuses on the salient object. During the synthesization, we apply masks to different layers for soft matting.

Existing work as [9, 10] have done some research on image refocusing, but these work require multi-view images for the same scene. There are also some research dedicated to single-image refocusing, which is considered as a more challenging technique [5]. Levin [11] designed a coded aperture by inserting a patterned occluder within the aperture of the camera lens, depth and the all-focused image can be recovered from a photograph taken by this modified camera. Moreno [7] proposed a method based on

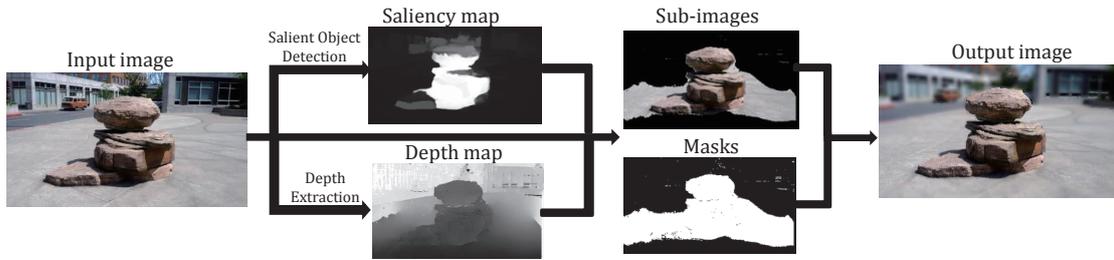


Figure 2: An overview of the proposed method.

computational photography to refocus from a photograph with the aid of a grid of dots projected on the scene. These methods rely on additional optical elements or devices to provide more information. Bando [3] proposed an interactive method to tackle the image refocusing problem, which requires a lot of user intervention. Zhang [5] presented a postprocessing method which can accomplish the tasks of focus-map estimation and image refocusing and defocusing. In this paper, the refocusing process is fully automatic and do not need any additional devices or human intervention.

To show the power of our method, we conduct experiments on the tiny dataset [2]. The results demonstrate that our method performs well both in refocusing effect and computational time.

## 2. SALIENCY-GUIDED IMAGE REFOCUSING

Fig. 2 shows an overview of our method. For an input raw image, the corresponding saliency map and depth map are firstly calculated. Then the input image is segmented into layers and blurred with different confusion. Finally, the output image is generated by combining the blurred layers and will refocus automatically on salient objects.

To be specific, our method essentially contains three main procedures. Firstly we differentiate each part of the raw input image according to the distance from the lens, namely the depth. Therefore, the salient object to be focused as a part of the image, is extracted from its background. There have been numerous methods for depth map estimation [12, 13] and saliency detection [14, 15, 16, 17], especially saliency detection by combining color and depth cues [18, 19, 20]. We reconstruct the 3D information according to the frames captured before shooting [2]. The use of multiple images together with Structure from Motion (SfM) approaches can help reduce uncertainty. Besides, long range connections are applied to depth estimation to improve dense reconstruction. Afterwards, we leverage the state-of-the-art highly efficient salient object detection algorithm which is based on *Minimum Barrier Distance* (MBD) transform [21] to generate saliency map. Secondly the image is segmented into multiple sub-images and blurred to different extent according to different depth ranges. Notably, the salient object will keep unchanged in order to stand out against the blurred background. Finally our method re-combines all the processed sub-images with an augmented approach to keep the stitching edges smooth, yielding the final refocused image. Through the three steps, a qualified refocused image is produced automatically. We will explain the vital procedures in detail in the following

subsections.

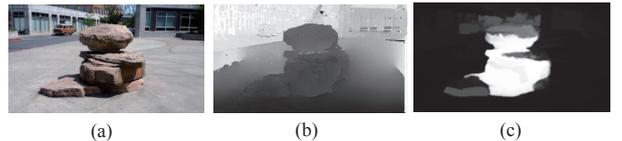


Figure 3: Input image and corresponding maps. (a) Input image. (b) Depth map. (c) Saliency map.

### 2.1 Adaptive Segmentation and Blurring

On the basis of the depth map and saliency map obtained in previous steps, the input image is segmented into  $k$  sub-images according to their depth ranges. Choosing proper  $k$  is an open question which varies according to different images. Too large value will leads to low efficiency while small value will make poor refocus results. Here for simplicity of illustration and without loss of generality, we restrict  $k$  to 5. We find that this value can achieve the balance between the efficiency and effectiveness. Therefore the whole depth space is separated into 5 different depth ranges by 4 thresholds denoted as  $D_1$  to  $D_4$ . Therefore the depths of sub-images can be represented as  $[0, D_1]$ ,  $[D_1, D_2]$ ,  $[D_2, D_3]$ ,  $[D_3, D_4]$ ,  $[D_4, MAX\_GRAY\_VALUE]$ . The region within a specific interval belongs to one sub-image. It should be noted that the middle two thresholds ( $D_2$  and  $D_3$ ) are determined by the saliency map as follows: we extract the salient object and compute its depth range  $[S\_MIN, S\_MAX]$ , and assign  $S\_MIN$  to  $D_2$  and  $S\_MAX$  to  $D_3$ . It means region in the interval  $[D_2, D_3]$  should be refocused. The other thresholds are selected adaptively by the gray distribution of depth map. For example, Fig. 4(a) shows the gray-level histogram of the input image,  $D_1$  and  $D_4$  are located in the bottoms which satisfy:

$$\forall D' \in [D_i - \sigma, D_i + \sigma], \mathcal{F}(D') \geq \mathcal{F}(D_i), i \in \{1, 4\} \quad (1)$$

where  $\mathcal{F}$  is the frequency of gray value  $D_i$ . It means  $D_1$  and  $D_4$  are local minimums which can ensure that the blurring degree for one object is the same as much as possible. Obviously the salient object to be focused, which falls into a specific depth range, is extracted as one sub-image, while the background is divided into other four sub-images with different depth ranges. Thus the input image is segmented to five sub-images according to these depth intervals. Fig. 4(b) shows a segmented image example. Then we perform a segment-level blurring to each sub-image except the middle one ( the middle one is the part should

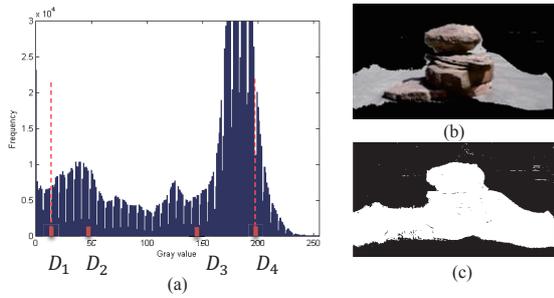


Figure 4: (a) Gray-level histogram. (b) Segmented image. (c) Corresponding mask.

be refocused). The degree of blurring depends on the depth range of the sub-image, the farther from salient objects, the deeper degree of the blurring.

## 2.2 Augmented Re-combination

The re-combination procedure is shown in Fig. 5. After five sub-images processed completely, we can simply combine them together to obtain the final refocused image. However, this trivial approach will lead to the roughness of the stitching edges. To solve this problem, we multiply each sub-image with a corresponding mask to generate the output image:

$$\text{Output image} = \sum_{i=1}^n \frac{1}{N_i} \text{Sub-image}_i \times \text{Mask}_i. \quad (2)$$

where  $N_i$  is the normalization coefficient. For a specific sub-image, we construct its corresponding mask according to the following steps. We first construct a binary matrix with the size equivalent to the input image. The retained part in sub-image is set to value 1 while the discarded is set to 0. Then we perform a blurring which is similar to the corresponding sub-image on the mask to make a transition of the splitting boundary. For example, the corresponding mask of Fig. 4(b) is shown in Fig. 4(c). After that, for each sub-image, we multiply with its corresponding mask. Finally we add all the multiplied sub-images up as the output image, the stitching edges of which are more smooth than before.

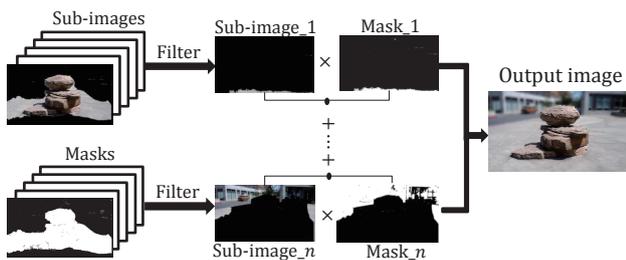


Figure 5: Flowchart of the re-combination process.

## 3. EXPERIMENTS

To validate the performance of our method, we compare it with the method in [3] and the more accurate algorithm which blurs pixels one by one. We will also analyze the execution efficiency of our method in this section.

## 3.1 Dataset and Experimental settings

We evaluate our method on a real dataset [2] which is collected by a smartphone camera, and is captured in the video mode at 24 framed per second. All the scenes are captured by a variety of users, which makes the dataset more close to people’s lives and thus reflecting the practical application of our method to real world.

We implement our method in MATLAB and carry out all the experiments on a machine with a 2.60GHz dual-core CPU and 4GB memory.

## 3.2 Results Evaluation

Fig. 6 presents some examples of performance validation results. As is shown in the third row, the salient objects stand out from the background regions. Besides, the blurring degrees in different depth ranges of the background are also different. In terms of execution efficiency, our method achieves linear time complexity. More specifically, regardless of the size of blurring kernel, the time complexity of the algorithm is  $O(N)$ , where  $N$  is the number of pixels.

We first validate the effectiveness by comparing our method with the refocusing algorithm in [3] as is shown in the forth row of Fig. 6. We can see that our method outperforms [3] in three aspects. Firstly, the salient object in our method stands sharper against its background. Secondly, our method works better in processing of edges between segments. Moreover, our method is automatic while algorithm in [3] is not. In summary, we achieve better refocusing effect, yielding enhanced user experience.

Our method blurs the background according to depth ranges, pixels falling in the same depth range are blurred with the same kernel. We also make a comparison with the algorithm that blurs pixels one by one. In other words, each pixel is blurred with a specific kernel according to its depth. Fig. 6 shows the results in the last row, which demonstrates that our method has no obvious difference with the accurate algorithm. However, it is worth mentioning that the accurate algorithm is much slower than our method.

## 4. CONCLUSION

In this paper we presented a simple but effective method to perform image refocusing automatically and rapidly which can be applied to portable devices. For a given photo, the region of interest is located by leveraging saliency detection methods. Then its depth map is calculated, the original photo is segmented into layers and blurred with different confusion sizes based on the depth map. Finally the blurred layers are re-combined into a refocused photo.

## 5. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation of China (61321491, 61202320), Research Project of Excellent State Key Laboratory (61223003), Research Fund of the State Key Laboratory for Novel Software Technology at Nanjing University (ZZKT2016B09), and Collaborative Innovation Center of Novel Software Technology and Industrialization.

## 6. REFERENCES

- [1] B. Zhang, R. Ju, T. Ren, and G. Wu. Say cheese: Personal photography layout recommendation using 3d aesthetics estimation. In *PCM*. Springer, 2016.

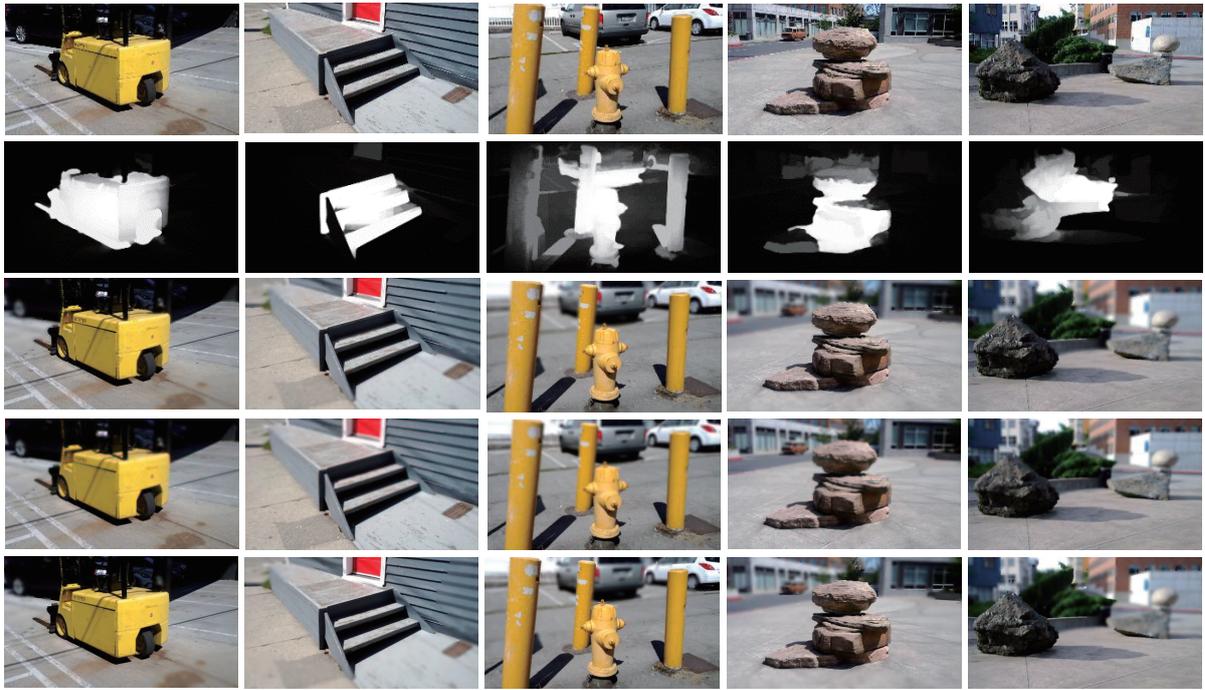


Figure 6: Examples of comparison. (Top row) Input images. (Second row) Saliency maps. (Third row) Results of our method. (Forth row) Results of algorithm in [3]. (Bottom row) Results of accurate method.

- [2] F. Yu and D. Gallup. 3d reconstruction from accidental motion. In *CVPR*, pages 3986–3993. IEEE, 2014.
- [3] Y. Bando and T. Nishita. Towards digital refocusing from a single photograph. In *PG*, pages 363–372. IEEE, 2007.
- [4] D.T. Vu, B. Chidester, H. Yang, M.N. Do, and J. Lu. Efficient hybrid tree-based stereo matching with applications to postcapture image refocusing. *TIP*, 23(8):3428–3442, 2014.
- [5] W. Zhang and W.-K. Cham. Single-image refocusing and defocusing. *TIP*, 21(2):873–882, 2012.
- [6] C. Sun and G. Wu. Depth extraction from a light field camera using weighted median filtering. In *PCM*. Springer, 2016.
- [7] F. Moreno-Noguer, P.N. Belhumeur, and S.K. Nayar. Active refocusing of images and videos. 26(3):67, 2007.
- [8] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *CSTR*, 2(11):1–11, 2005.
- [9] A. Kubota and K. Aizawa. Reconstructing arbitrarily focused images from two differently focused images using linear filters. *TIP*, 14(11):1848–1859, 2005.
- [10] J. Yang and D. Schonfeld. Virtual focus and depth estimation from defocused video sequences. *TIP*, 19(3):668–679, 2010.
- [11] A. Levin, R. Fergus, F. Durand, and W.T. Freeman. Image and depth from a conventional camera with a coded aperture. 26(3):70, 2007.
- [12] A. Saxena, M. Sun, and A.Y. Ng. Make3d: Learning 3d scene structure from a single still image. *TPAMI*, 31(5):824–840, 2009.
- [13] R. Ju, Y. Yang, X. Xu, C. Xia, and G. Wu. A complementary aggregation approach for local stereo matching using color and correlation cues. In *PCM*, pages 372–383, 2013.
- [14] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *TPAMI*, 33(2):353–367, 2011.
- [15] Z. Liu, W. Zou, and O. Le Meur. Saliency tree: A novel saliency detection framework. *TIP*, 23(5):1937–1952, 2014.
- [16] T. Ren, Y. Liu, R. Ju, and G. Wu. How important is location information in saliency detection of natural images. *MTAP*, 75(5):2543–2564, 2016.
- [17] S.-H. Zhong, Y. Liu, T.-Y. Ng, and Liu Y. Perception-oriented video saliency detection via spatio-temporal attention analysis. *NEUCOM*, 2016.
- [18] J. Guo, T. Ren, J. Bei, and Y. Zhu. Salient object detection for rgb-d image via saliency evolution. In *ICIMCS*. ACM, 2015.
- [19] R. Ju, Y. Liu, T. Ren, L. Ge, and G. Wu. Depth-aware salient object detection using anisotropic center-surround difference. *SPIC*, 38:115–126, 2015.
- [20] J. Guo, T. Ren, and J. Bei. Salient object detection for rgb-d image via saliency evolution. In *ICME*. IEEE, 2016.
- [21] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mèch. Minimum barrier salient object detection at 80 fps. In *ICCV*, pages 1404–1412.