Journal of Computers Vol. 29 No. 6, 2018, pp. 29-39
doi:10.3966/199115992018122906003

Recovering Depth from a Single Natural Image Based on Edge Blur Estimation

Feng-Yun Cao*, Xue-Jie Yang, Yan-Yu Qian, Pei-Bei Shi

Department of Information Engineering and Computer Science, Hefei Normal University, Hefei 230601, Anhui, China
{caofengyun, yangxuejie, qianyanyu, shipeibei}@hfnu.edu.cn

Received 12 July 2017; Revised 15 November 2017; Accepted 9 January 2018

Abstract. The traditional methods of Depth from Defocus (DFD) usually need to collect multiple defocus images, which are difficult to realize in practice. In this paper, the authors managed to solve this challenging problem with recovering the depth from a single image taken with an uncalibrated conventional camera. Different from all the existing depth recovering approaches, this approach avoides the collection of multiple images or the usage of deconvolution process, which provides a simple yet effective way to realize depth-recovering in a single image. The amount of defocus blur is obtained by the gradient magnitude ratio between the input and re-blurred images. Sparse blur map is obtained through the estimate of blur amount at the edge regions with the segmentation of images. Complete depth information is then recovered by propagating the sparse blur based on the local mean of edge blur. Experimental results on a variety of images show that the approaches in this paper can acquire a reliable estimation of the depth.

Keywords: defocus blur, depth information, gaussian gradient, image segment

1 Introduction

Depth from defocus plays an important role in computer vision and digital image processing with applications such as 3D information reconstruction, robot navigation, medical imaging, biometric identification, information extraction, industrial detection and so on. The depth information recovery is an important research topic in the field of computer vision in recent years. The depth information recovery algorithm consists of monocular cues (the vanishing point, shape, occlusion, texture change rate, air perspective, defocusing) or based on binocular cues (parallax). Depend on the acquisition of multiple image scene depth recovery method. The traditional stereo vision method [1] and motion vision method [14] is similar, use image or multiple images based on geometric relationships between the corresponding points in the scene depth recovery. Depth from focus [2] through a series of focus images by gradually and accurately determine the object points to the imaging system. The traditional methods of Depth from defocus [3] requires a pair of images of the same scene with different focus setting. It estimates the degree of defocus blur, which the depth of scene can be recovered based on the camera setting.. This kind of method to recover the scene depth is generally affected by the presence of complex image feature matching and occlusion and practical application of conditionality.

Today, the scene depth recovery using a single image, there are based on additional conditions, edge determi-nation, the texture and so on. Among them, the method of additional conditions includes active illumination method [4] and coded aperture [5], such methods need additional specific camera collecting lens, light source and conversion, the actual application of the binding. Study based on texture information [15] around the global, multiscale, and hierarchical levels, but this method can only be applied well in outdoor scenes, and only has good visuals, and actual depth difference. Saxena et al. [6] based on the application of machine learning methods, by constructing a spatial relationship model

* Corresponding Author

between the outdoor scenes and to restore scene depth, but because of its need for large scale machine learning, poor real-time. The method of the edge blur measurement [16], the image degradation process is modeled as a thermal diffusion process, depth recovery using inhomogeneous inverse heat conduction equation. Hu and Haan [17] and Zhuo and Sim [7] propagated the blur value from the edge regions to the whole image using matting and MRF, but because of its time is too expensive, poor real-time.

In this paper, the authors managed to solve this challenging problem with recovering the depth from a single image taken with an uncalibrated conventional camera. There are several problems in the closed work [17, 7] to ours. First, matting Laplace matrix [8] propagate information is quite limited and poor real-time. As a result, the depth of the flat area in the scene is obviously wrong. In addition, there is texture ambiguity in depth estimation from single image using defocus cue. The defocus measure we obtained may be due to a edge that is out of focus or that is a fuzzy texture. As shown in Fig. 1, the area marked by the red circle is actually blur texture of the pumpkin, but Zhuo's method [7] treats it as defocus blur, which results in error.

Fig. 1. Depth estimation using Zhuo's method [7]

To overcome these problems, this paper proposes a simple yet effective approach to realize depth-recovering from a single image. First of all, the amount of defocus blur is obtained by the gradient magnitude ratio between the input and re-blurred images. Next, with the segmentation image, a sparse blur map is obtained by estimating the blur amount at the edge regions. Complete depth information is then recovered by propagating the sparse blur based on the local mean of edge blur. Different from all the existing depth recovering approaches, this approach avoides the collection of multiple images or the usage of deconvolution process, which provides a simple yet effective way to realize depth-recovering in a single image.

Our work has three main contributions. Firstly, the paper propose an improved blur estimation method based on the gradient magnitude ratio [7]. Compared with the Zhuo's algorithm, our algorithm is more efficient in edge preserving denoising. Secondly, without any modification to the camera or using additional illumination, our blur estimation method combined with the local mean of edge blur can obtain the depth map of a scene by using only single defocused image captured by an uncalibrated conventional camera. As shown in Fig. 7, our method can extract a layered depth map of the scene with fairly good extent of accuracy. Finally, we discuss two kinds of ambiguities in recovering depth from a single image using defocus cue, one of which is usually overlooked by previous methods.

2 Related Work and Analysis

2.1 Defocus Image Degradation Model

Assuming that the edges are step edges and estimate the defocus blur information at the edges, then an ideal step edge model can be represented as

$$f(x) = Au(x) + B. \tag{1}$$

In the formula, $u(x)$ is a step function, A and B are corresponding to the amplitude and displacement of the edge position, and the edge position is x=0. Generally assume that focusing and defocusing follow the thin lens model [9], the scene is located at the focal point, and the light gathers in the focal plane to form a clear point. If the target deviates from the focus, the light convergence is not a clear point but a fuzzy surface, and its area increases with the increase of the distance. The blurred pattern depends on the shape of aperture and is called the circle of confusion (CoC) [9]. The diameter of CoC characterizes the amount of defocus and can be written as

$$r = \frac{|d - d_f|}{d} \frac{f_0^2}{N(d_f - f_0)}. \tag{2}$$

where f_0 and N are the focal length and the f-number of the camera respectively. The thin lens model for defocus blur is shown in Fig. 2.

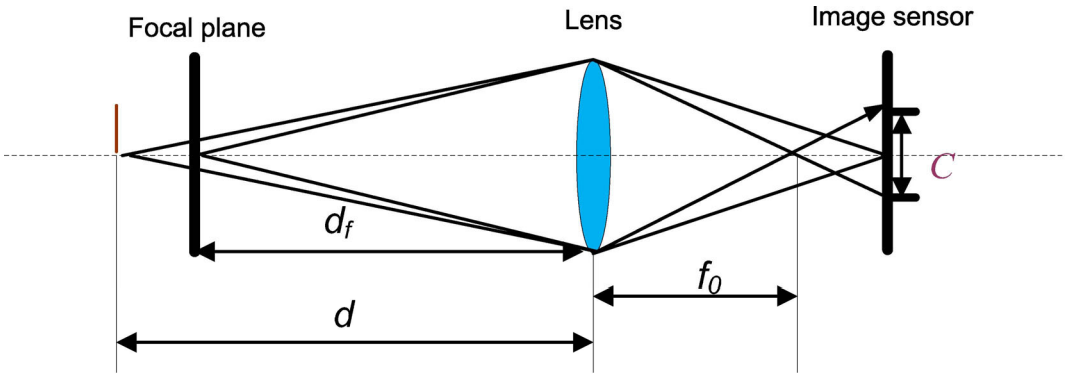


Fig. 2. Defocus model

Given the parameters d_f=500mm and f_0=80mm, change the distance d(2000mm~4000mm) and N(2,4,8) parameters, we can get

$$r = \frac{|d - 50|}{d} \frac{64}{42 * N}.$$
$$r = C * \left(1 - \frac{500}{d}\right). \tag{3}$$

As we can see, r is a non-linear monotonically increasing function of the object distance d. The defocus blur can be modeled as the convolution of a sharp image with the point spread function (PSF). The commonly used PSF are two-dimensional Gaussian functions $g(x, y, \sigma)$, where the standard deviation σ is directly proportional to the diameter of the blur circle, can be described as:

$$\sigma = k * r. \tag{4}$$

Where σ determines the effect of the PSF convolution, that is, the blurring of the image. Thus, the defocus degradation model can be represented as:

$$f(x,y) = f_0(x,y) * g(x,y,\sigma). \tag{5}$$

Where $f(x,y)$ is blurred image, $f_0(x,y)$ is sharp image and $g(x,y,\sigma)$ is Gaussian point spread function

$$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}. \tag{6}$$

2.2 Local Defocus Blur Estimation Algorithm

In this paper, we apply the local blur estimation method in [7] to obtain the rough blur map based on the sharp edge assumption. Fig. 3 shows the overview of this method in one-dimensional case. A step edge is re-blurred using a Gaussian kernel with know standard deviation, ⊗ and ▽ are the convolution and gradient operators respectively. The black dash line denotes the edge location.

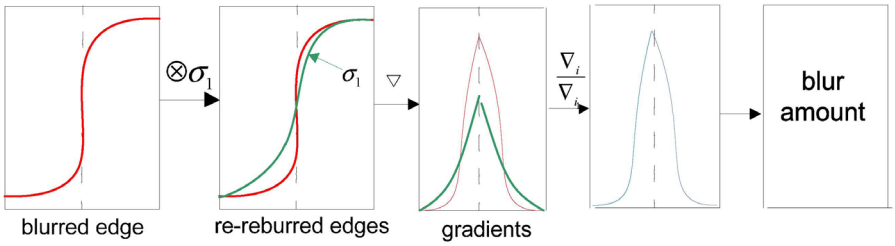


Fig. 3. Blur estimation approach [7]

For ease of description, we describe blur estimation method for 1D case. The gradient of the re-blurred edge is:

$$\begin{aligned}\nabla i_1(x) &= \nabla\left(\left(Au(x)+B\right) \otimes g(x,\sigma) g(x,\sigma_1)\right) \\ &= \frac{A}{\sqrt{2\pi\left(\sigma^2+\sigma_1^2\right)}} \exp\left(\frac{x^2}{2\left(\sigma^2+\sigma_1^2\right)}\right).\end{aligned} \tag{7}$$

where, σ_1 is the standard deviation of the re-blur Gaussian kernel, call it the re-blur scale. The gradient ratio at the edge of the original image and the re-blurred image is:

$$\frac{|\nabla i(x)|}{|\nabla i_1(x)|} = \sqrt{\frac{\sigma^2+\sigma_1^2}{\sigma^2}} \exp\left(\frac{x^2}{2\sigma^2} - \frac{x^2}{2\left(\sigma^2+\sigma_1^2\right)}\right). \tag{8}$$

It can be proved that the ratio is maximum at the edge location (x = 0) and the maximum value R is given by:

$$R = \frac{|\nabla i(0)|}{|\nabla i_1(0)|} = \sqrt{\frac{\sigma^2+\sigma_1^2}{\sigma^2}}. \tag{9}$$

Observation equations (7) and (9) can be seen, the edge gradient depends on both the edge amplitude A and the blur amount σ, while the maximum value R eliminates the effect of edge amplitude A and depends only on σ and σ_1 . Thus, given the maximum value R at the edge locations, σ can be calculated using

$$\sigma = \frac{1}{\sqrt{R^2-1}} \sigma_1. \tag{10}$$

As any direction of a 2D isotropic Gaussian function is a 1D Gaussian, the blur estimation is similar to

that in 1D case. We set the re-blurring $\sigma_1 = 1$ and use Canny edge detector [19] to perform the edge detection, the experimental parameters of edge extraction are set to 0.05. The blur scales are estimated at each edge location, forming a sparse depth map denoted by $\hat{d}(x)$. As shown in Fig. 4.

Fig. 4. Sparse map

However, due to noise or soft shadows, the blur estimates may contain some errors. To suppress these errors, Zhuo and Sim [7] apply joint bilateral filter [10] to solve this kind of problem. The output of the joint bilateral filter at each edge location can be defined as:

$$BF\left(\hat{d}(x)\right) = \frac{1}{W(x)} \sum_{y \in N(x)} G_{\sigma_s}\left(\| x - y \|\right) G_{\sigma_r}\left(\| I(x) - I(j) \|\right) \hat{d}(y). \tag{11}$$

where $W(x)$ is the normalization factor and $N(x)$ is the neighborhood of x given by the size of spatial Gaussian filter G_{σ_s}. However, in the actual process, it can be found that using joint bilateral filtering for edge location noise reduction is often not good to follow the original signal. To suppress these problems, we apply image guided filtering [11] replaced joint bilateral filter. Because it can achieve smooth filtering, it also has good edge preservation performance, and it can solve the interference of noise in local fuzzy estimation well. Image guided filtering is a linear shift filtering process, It includes the guide image I, the input image p and the output image q. Among them, the guide image I needs to be set according to the specific application, you can also take the input image p directly. For the first i pixels in an output image, the calculation method can be expressed as follows:

$$q_i = \sum_j W_{ij}(I) p_j. \tag{12}$$

where i and j are pixel tags, and is the filter kernel fuction. In document [11], is defined as follows:

$$W_{ij}(I) = \frac{1}{|\omega|^2} \sum_{k:(i,j) \in \omega_k} (1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \varepsilon}). \tag{13}$$

where ω_k is the window of the first k kernel, $|\omega|$ is the number of pixels in the window, μ_k and σ_k are the mean and variance of the guide image in the window, ε is the smoothing factor.

In order to better illustrate the performance advantage of the image guided filtering, which has the role of edge-preserving and denoising. This paper demonstrates it by comparing it with the joint bilateral filtering. Fig. 5 shows the application of the two filter in image enhancement using one-dimensional signals. Given the input signal (black), the output of the edge preserving and denoising is red, which is called the base layer. The difference between the input and output signals is expressed in blue, which is called the detail layer. The result of the bilateral filtering at the edge is that the reference layer will appear inconsistent with the input signal, because there are few pixels of the same color near the edge, and the weighted average is not enough. For the image guided filtering, however, the base layer is satisfied $\nabla q = \alpha \nabla I$ at the edge, so the gradient of the edge is unchanged. As can be seen from the diagram, the detail layer of bilateral filtering has a large fluctuation, which results in the phenomenon of gradient inversion at the edge of the enhanced image, while the image guided filtering can better deal with edge information.

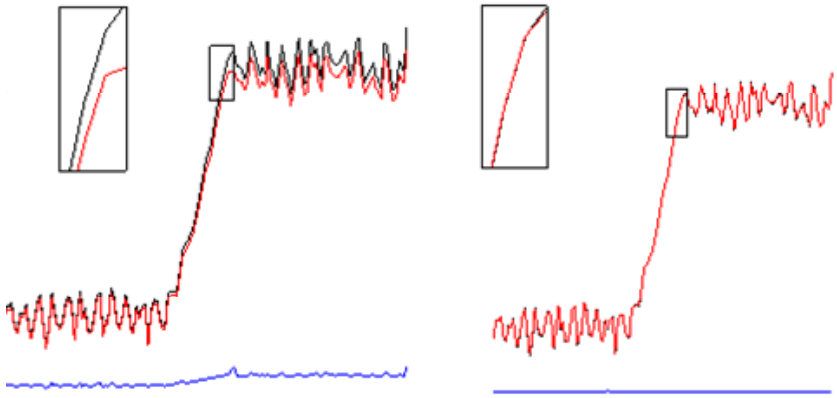


Fig. 5. Comparison of details for 1-D signal

The image guided filtering corrects some errors in the sparse depth map and avoid the propagation of these errors in the depth recovery process described in the next section.

3 Depth Recovery Algorithm

3.1 Overview

The overview of our algorithm is illustrated in Fig. 6. There are two steps of our depth estimation algorithm, which are initial estimation and refinement. The input of our algorithm is a single image. The input image is firstly smoothed using L0 gradient minimization [12]. Then a local blur estimation method is proposed to calcu-late the blur amounts at the boundaries of edges, where the obtained result is called sparse blur map. After that, a depth interpolation method is adopted to propagate the blur value from the boundaries to the unknown regions. In the next, based on a simple geometry prior of photography and Sky zone correction algorithm [13] is adopted to refine the depth map.

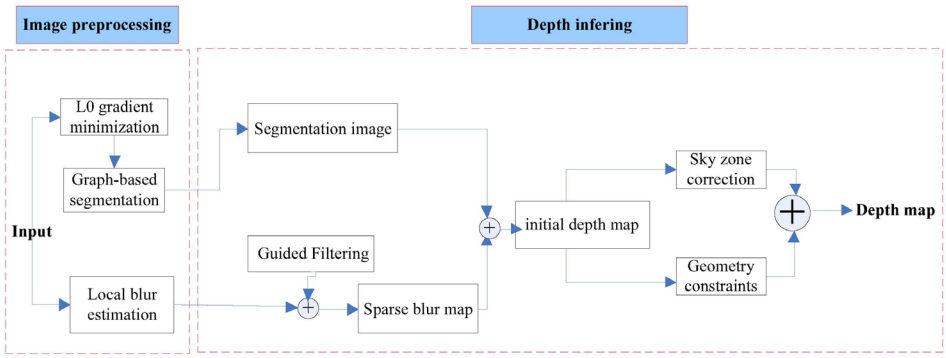


Fig. 6. Flowchart of our algorithm

3.2 Depth Interpolation Based on Local Blur Estimation

Our blur estimation methods yields a sparse depth map $\hat{d}(x)$ with depth estimates at edge locations. To obtain a full depth map d(x), we need to propagate these values from edge locations to the entire image. Through the understanding of the defocused imaging model in the previous section, we can clearly know the target in the scene, the object with respect to the defocus distance corresponding to different degree of the camera is not the same, but in the local small area depth is consistent. In this paper, based on the local neighborhood consistency, the image segmentation algorithm is used to realize the expansion of the local blur number. For its low complexity, good controllability and flexibility, we adopt the graph-based image segmentation [18] proposed by Felzenszwalb and Huttenlocner. This segmentation method is based on

graph. The image matrix is constructed to be an undirected graph $G=(V,E)$, where each node $v_i \in V$ corresponds to a pixel in the image. The correlation between pixels is represented by a weighted value $w(v_i,v_j)$, which is a non-negative measure of the dissimilarity between neighboring elements vi and vj.S to represent the results of segmentation, written as

$$s=\left\{C_i \mid C_i \in V, \cup_i C_i \in V, C_i \cap C_j=\varnothing, i \neq j\right\}. \tag{14}$$

where Ci index disjoint subsets of segmentation. In a partition of a subset, the difference between the elements is expressed as the following

$$Int(C)=\max_{e \in MST(C,E)} w(e). \tag{15}$$

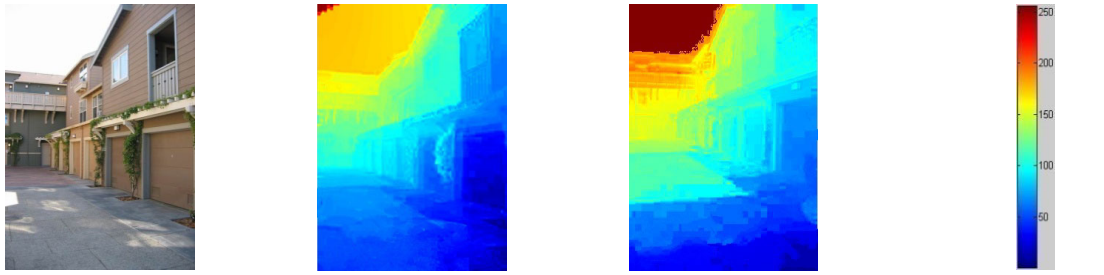
where $MST(C,E)$ index Minimum spanning tree for many partition subsets. Difference between different subsets, written as

$$Dif(C_1,C_2)=\min_{v_i \in C_1, v_j \in C_2, (v_i,v_j) \in E} w\left((v_i,v_j)\right). \tag{16}$$

For the need of a merger between subsets, using the value to determine the following,

$$D(C_1,C_2)=\begin{cases} T, Dif(C_1,C_2) > MInt(C_1,C_2) \\ F, other \end{cases}. \tag{17}$$

we need to set three parameters σ, k and min_size. σ is the parameter of Gauss filter Prof. Felzenszwalb proposed for smoothing image and is set around 0.8 for most cases. Min_size is used to control the pixels' number of the smallest patch. Additionally k is a threshold and a larger k causes a preference for larger components while it is not a minimum component size. In this paper, k=70 and Min_size =10, as shown in Fig. 7, the edge of the local sparse fuzzy quantity is good to expand from the local neighborhood, obtain the target depth information of the scene results have good depth change trend.



(a)input image (b) initial depth (c)our final depth (d)colorbar

Fig. 7. The depth recovery result of our method

3.3 Ambiguities in Depth Recovery

There are two kinds of ambiguities in depth recovery from single image using defocus cue. The first one is the focal plane ambiguity. When an object appears blur in the image, it can be on either side of the focal plane, as shown in Fig. 8(a). To remove this ambiguity, most of the depth from defocus methods assume all objects of interest are located on one side of the focal plane. When taking images, they just put the focus point on the nearest/farthest point in the scene. However, it is very inaccurate to adopt such simple constraints in practical applications. In practical applications, the farther away from the focus, the greater the degree of ambiguity, on the contrary, the smaller the degree of ambiguity. Therefore, in this paper, we select the minimum point in the global fuzzy graph, and approximately consider it as the focus position, then, when it is located in front of the focal plane, at the back of the focal plane. The method avoids the simple irrationality existing in the previous algorithm, and the experimental results show that it can solve the focal plane ambiguity.

Simulation experiments focus position change are carried out in this paper to further reflect the ambiguity of the two focal plane, as shown in Fig. 8(a) for the original image, the middle position of the focus is on the teapot, teapot and are due to defocus phenomenon occurred from focusing. When we change the focus position, in front of the teapot and the rear teapot, the corresponding defocus blur occurs as the specific object changes.

Fig. 8. Simulation image for focus positin change

The second ambiguity is called the blur texture ambiguity. The defocus measure we obtained may be due to a sharp edge that is out of focus or a blur edge that is in focus. This ambiguity is often overlooked by previous work and may cause some artifacts in our result. One example is shown in Fig. 9. The region indicated by the white rectangle is actually blur texture of the flower, but Zhuo's method [7] treats it as sharp edges due to defo-cus blur, which results in error estimation of the depth in that region.

Fig. 9. The blur texture ambiguity [7]

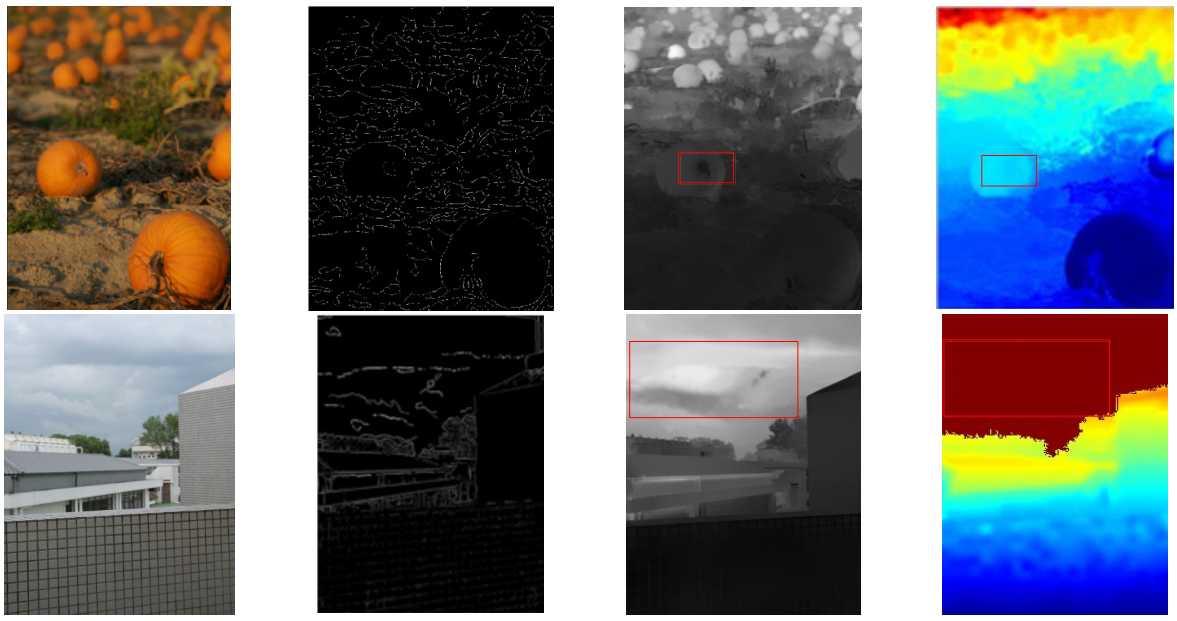
3.4 Sky Zone Correction

As in the blur estimation, the results will be influenced by the interference of color information, especially in the area of the sky in a natural scene, in order to eliminate the inaccurate results sky area interference information brings. This paper introduces the separation method of sky region based on statistics in the literature [13]. The sky region is separated out, as shown in Fig. 10. After separation, the sky two value map can be obtained. This paper will separate the sky two map and depth map with the initial value, the actual situation of the camera in the real world when shooting sky area distance lens distance is much larger than other scenes based on the sky area depth correction, to eliminate the interference of color information, to recover the exact target depth information of the scene, as shown in Fig. 7(c) shown in the sky region is very good depth correction.

4 Contrast Experiment and Result Analysis

In Fig. 11, we compare our method with the Zhuo's method [7]. The Zhuo's method needs complex matting algorithm, and is poor in real time. The region indicated by the red rectangle is actually blur texture of the pumpkin image, but Zhuo's method [7] treats it as sharp edges due to defocus blur, which results in error estima-tion of the depth in that region.

Fig. 10. Examples separation region of the sky

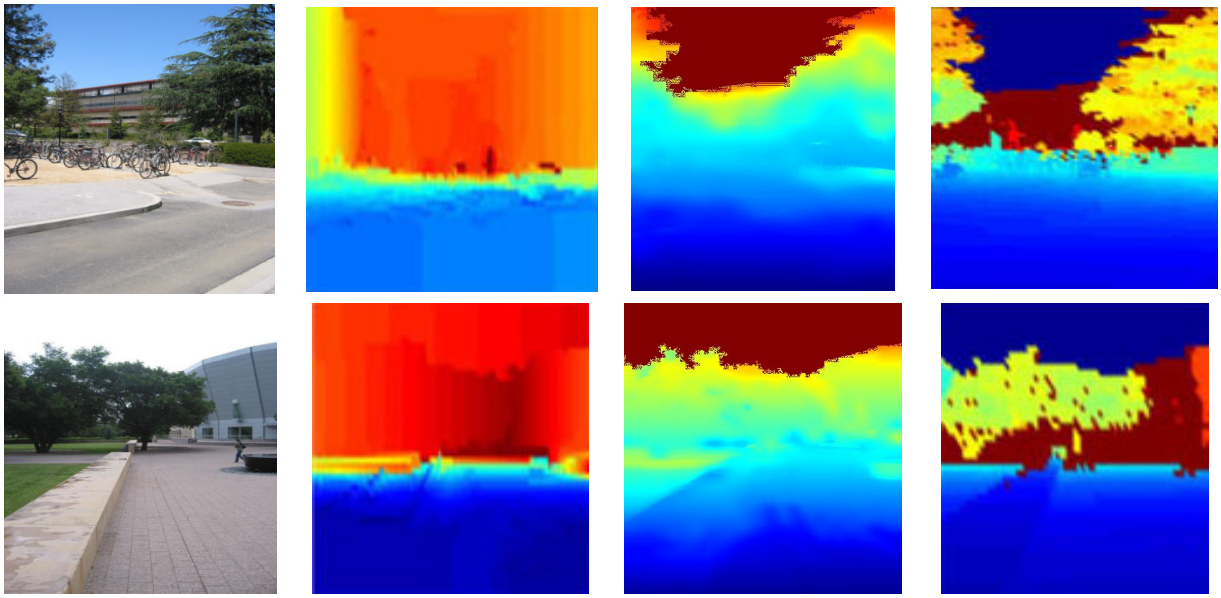


(a)input image (b) Sparse depth map (c) Zhuo's result (d) our depth map

Fig. 11. Comparison of Zhuo's method

In the building image, the scene mainly contains three layers: the wall, house and sky layers. Ours and Zhuo's method can be produce depth maps accurately representing those depth layers. But in the red rectangle mark area, it is obvious that our method is able to accurately recover the depth of a scene from a single defocused image.

In Fig. 12, we compare our method with the Saxena's method. we perform contrastive experiments on the well-known Make3D depth-map database [20]. Note that most of the images in this database do not include adequate defocus cues. However, our method can also obtain fairy good depth estimation results. The experimental results are shown as Fig. 12. The images shown in the first column are origin images. The images shown in the second is our depth map. The images shown in the third and fourth column are Make3D's results and the ground-truth depth-map. To better illustrate our algorithm, real time contrast results are shown in Table 1.



(a)input image (b) Saxena's result (c) our depth map (d) The ground truth depth map

Fig. 12. Comparison of Saxena's method

Table 1. Real time comparisons with the results of the Saxena's and Zhuo's algorithms

Method	time-consuming
Saxena's	17s
Zhuo's	35s
Ours	9s

5 Conclusion

In this paper, a new method is presented to recovery the depth information from a single natural scene. The key point for recovering depth information is the blur amount at edge locations based on the Gaussian gradient ratio. A full depth map is then recovered using the the local mean of edge blur. We show that our method can generate more accurate scene depth maps compared with existing methods. What is more important is that the our method overcomes the high computational complexity caused by deconvolution and matting operation in the previous algorithm, and greatly improves the efficiency of the algorithm.

Since our blur map estimation method is based on the edge prior, the blurring subtle textures may bring in the estimation errors in the blur map. It is difficult to determine whether this is a sharp edge that is out of focus or a blur edge that is in focus. Although We discuss ambiguities arising in recovering depth from single images using defocus cue and propose some possible ways to remove the ambiguities, the method still may be occasionally invalid. In the future, we would like to extend our method to work on more depth cues, such as junction, vanish-ing lines and texture gratitude, would be combined with the defocus cue to eliminate the estimation errors.

Acknowledgements

This work was supported by key project of natural science research of Anhui Province and lateral research funds "Application of low order feature in depth recovery of single image, Grant No. KJ2017A926", "Research on the difficulty of pedestrian detection based on machine learning" and "Research on the modeling method of human body digital model driven by finite element method, Grant No. KJ2017A927". The author would like to acknowledge Qin Tong and Shi Peibei for their discussion and useful suggestion.

References

[1] U.R. Dhond, J.K. Aggarwal, Structure from stereo-a review, Systems Man & Cybernetics IEEE Transactions 19(6)(1989) 1489-1510.

[2] S.K. Nayar, Y. Nakagawa, Shape from focus, IEEE Transactions on Pattern Analysis & Machine Intelligence 16(8)(1989) 824-831.

[3] P. Favaro, S. Soatto, A geometric approach to shape from defocus, IEEE Transactions on Pattern Analysis & Machine Intelligence 27(3)(2005) 406-17.

[4] F. Moreno-Noguer, P. N. Belhumeur, S.K. Nayar, Active refocusing of images and videos, Acm Transactions on Graphics 26(99)(2007) 67.

[5] A. Levin, R. Fergus, W.T. Freeman, Image and depth from a conventional camera with a coded aperture, Acm Transactions on Graphics 26(3)(2007) 70.

[6] A. Saxena, M. Sun, A.Y. Ng, Make3D: learning 3D scene structure from a single still image, IEEE Transactions on Pattern Analysis & Machine Intelligence 31(5)(2008) 824-840.

[7] S. Zhuo, T. Sim, Defocus map estimation from a single image, Pattern Recognition 44(9)(2011) 1852-1858.

[8] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Transactions on Pattern Analysis & Machine Intelligence 30(2)(2008) 228.

[9] E. Hecht, Optics, 4th ed., Addison-Wesley, Boston, 2001.

[10] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, M. Hoppe, Digital photography with flash and no-flash image pairs. <http://www.hhoppe.com/flash.pdf>, 2004.

[11] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Transactions on Pattern Analysis & Machine Intelligence 35(6)(2013) 1397.

[12] L. Xu, C. Xu, Y. Xu, J. Jia, Image smoothing via L 0 gradient minimization, Acm Transactions on Graphics, 30(6)(2011) 1-12.

[13] D. Hoiem, A. A. Efros, M. Hebert, Automatic photo pop-up, Acm Transactions on Graphics 24(3)(2005) 577-584.

[14] F. Dellaert, S.M. Seitz, C.E. Thorpe, S.Thrun, Structure from motion without correspondence, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2000.

[15] D. Hoiem, A. A. Efros, M. Hebert, Geometric context from a single image, in: Proc. Tenth IEEE International Conference on Computer Vision IEEE Computer Society, 2005.

[16] V.P. Namboodiri, S. Chaudhuri, Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[17] H. Hu, G.D. Haan, Adaptive image restoration based on local robust blur estimation, in: Proc. International Conference on Advanced Concepts for Intelligent Vision Systems Springer-Verlag, 2007.

[18] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, International Journal of Computer Vision 59(2)(2004) 167-181.

[19] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8(6)(1986) 679-698.

[20] A. Saxena, S.H. Chung, A.Y. Ng, Learning depth from single monocular images, in: Proc. International Conference on Neural Information Processing Systems, 2005.