Scientific Research Publishing

Depth Estimation from a Single Image Based on Cauchy Distribution Model

Ying Ming

College of Politics, National Defense University of PLA, Shanghai, China
Email: whumike@foxmail.com

Abstract

Most approaches to estimate a scene's 3D depth from a single image often model the point spread function (PSF) as a 2D Gaussian function. However, those methods are suffered from some noises, and difficult to get a high quality of depth recovery. We presented a simple yet effective approach to estimate exactly the amount of spatially varying defocus blur at edges, based on a Cauchy distribution model for the PSF. The raw image was re-blurred twice using two known Cauchy distribution kernels, and the defocus blur amount at edges could be derived from the gradient ratio between the two re-blurred images. By propagating the blur amount at edge locations to the entire image using the matting interpolation, a full depth map was then recovered. Experimental results on several real images demonstrated both feasibility and effectiveness of our method, being a non-Gaussian model for DSF, in providing a better estimation of the defocus map from a single un-calibrated defocused image. These results also showed that our method was robust to image noises, inaccurate edge location and interferences of neighboring edges. It could generate more accurate scene depth maps than the most of existing methods using a Gaussian based DSF model.

Keywords

Depth Estimation, Depth From Defocus, Defocus Blur, Gaussian Gradient, Cauchy Distribution, Point Spread Function (PSF)

1. Introduction

Estimation of 3D depth from the scene is a fundamental problem of computer vision and computer graphics applications including robotics, scene understanding, image deblurring and refocusing and 3D reconstruction. Conventional methods for 3D depth recovery have focused on stereovision [1], structure from

motion [2], and other methods that require two (or more) images. However, these algorithms often end up ignoring the numerous additional monocular cues that can also be used to obtain 3D information [1]. However, these methods suffer from the occlusion problem, or cannot be applied to dynamic scenes, which limits their applications in practice.

In recent work, several methods have been proposed to recover depths map from a single image, which do not suffer from the correspondence problem of multiple images matching. Their process is simple and fast, and therefore, they get more and more people's attention. However, depth estimation from a single image is a difficult task and requires that we take into account the global structure of the image, as well as use prior knowledge about the scene [3]. Currently, methods for single image depth restoration commonly use geometric depth information cues such as horizontal planes, vanishing points and edge surfaces, or monocular depth cues such as shading, color changes, perspective, texture variations, texture gradient, occlusion, hazy, sample objects, similar scenes, defocus, etc. [2]-[9]. These methods are still computing complex, difficult to apply in non-restricted scenarios.

The depth recovery method using monocular depth cues of a single defocused image is developed from the traditional method of Depth from Defocus (DFD) [10] which requires a pair of images of the same scene with different focus setting, including active illumination methods [11], coded aperture defocus depth methods [12] [13] and edge blur depth method [14]. Active illumination methods project sparse grid dots onto the scene and the defocus blur of those dots is measured by comparing them with calibrated images. Then the defocus measure can be used to estimate the depth of a scene. The coded aperture method changes the shape of camera aperture [12] or uses multiple color-filter aperture (MCA) [13] [15] to make defocus de-blurring more reliable. A defocus map and an all-focused image can be obtained after deconvolution using calibrated blur kernels. These methods require additional illumination or camera modification to obtain a defocus map from a single image.

In this paper, we focus on a more challenging problem of recovering the defocus map from a single image captured by an uncalibrated conventional camera, using edge blur defocus. The edge blur defocus depth methods are based on the amount of blur in the image with depth objects in the scene, defocus blurred image can be modeled as a convolution of clear image and PSF depth recovery from single-focus image. Elder and Zucker [8] used the first- and second-order derivatives of the input image to find the locations and the blur amount of edges. The defocus map obtained is sparse. Bae *et al.* [9] extend this work and obtain a full defocus map from the sparse map using an interpolation method.

Namboodiri and Chaudhuri [14] model the PSF of defocus blur as a thermal diffusion process and use the inhomogeneous inverse heat diffusion to estimate defocus blur at the edge locations, and then apply a graph-cut based method to recover the scene's depth map. Zhuo *et al.* [16] use a Gaussian function to model

the PSF. The input image is re-blurred using a known Gaussian blur kernel and the ratio between the gradients of input and re-blurred images is calculated. The blur amount at edge locations can be derived from the ratio. They acquire better results of depth recovery than Namboodiri's. Fang *et al.* [17] use a DSF similar to Zhuo's, assuming the local depth is continuous. The depth of the other regions is interpolated from the depth of the inner edge by a local plane fitting. However, most of the existing Gaussian based PSF have the ambiguity problem between the hard edge and the soft edge of the scene [18]. In contrast, we estimate the defocus map in a different but effective way. The input image is re-blurred using a known Cauchy blur kernel and the ratio between the gradients of input and re-blurred images is calculated. We show that the blur amount at edge locations can be derived from the ratio. We then apply the matting interpolation to propagating the blur amount at edge locations to the entire image. We finally obtain a full depth map.

Inspired by [16] and [19], combined with our previous work [20], we propose an efficient blur estimation method based on the Cauchy PSF, and show that it is robust to noise, inaccurate edge location and interference from neighboring edges. Without any modification to cameras or using additional illumination, our method is able to obtain the defocus map of a single image captured by conventional camera. Our method can estimate the depth map of the scene with fairly good extent of accuracy.

2. Defocus Model

As the amount of defocus blur is estimated at edge locations, we must model the edge first. To estimate the amount of defocus blur at the edges of objects in an image, we adopt the ideal step edge model [16] which is

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

where $u(x)$ is the unit step function. A and B are the amplitude and offset of the edge, respectively. Note that the edge is located at $x = 0$.

We assume that focus and defocus obey the thin lens model. According to thin lens model, when an object is placed at the focusing distance d_f, the image will appear sharp [21], as shown in **Figure 1**. When the object is at other distance
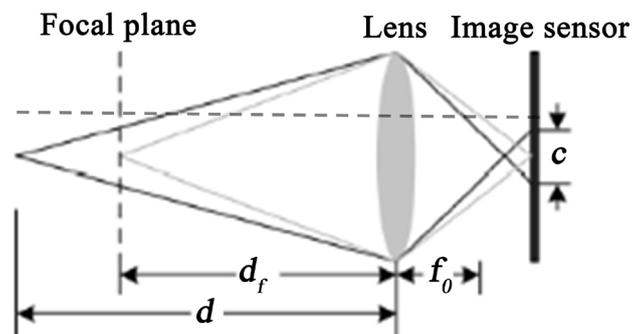
Figure 1. A thin lens model.

d, it results in a blurred image. The blurred pattern depends on the shape of aperture and is called the circle of confusion (CoC). The diameter of CoC c characterizes the amount of defocus which is a non-linear monotonically increasing function of the object distance d [21].

The defocus blur can be modeled as the convolution of a sharp image $f(x)$ with the point spread function (PSF) [10]. The PSF can be approximated by a Gaussian function $g(x, \sigma)$, where the standard deviation $\sigma = k \cdot c$ is proportional to the diameter of the CoC c and measures the defocus blur amount. We use σ as a measure of the depth of the scene, and call it the re-blur scale. A blurred edge $i(x)$ can be represented as

$$i(x) = f(x) \otimes g(x, \sigma) \tag{2}$$

According to [19], we know that a PSF is only required rotationally symmetric, and non-Gaussian model can be applied to a PSF. The shape of a Cauchy distribution function is similar to a Gaussian function, and drops more smoothly and heavier trailing. The previous work [20] also confirmed that the Cauchy distribution model is more robust to noise than Gaussian. So that we use a 2D Cauchy distribution function instead of 2D Gaussian. The scale parameter σ of a Cauchy distribution (as same as the standard deviation of a Gaussian σ) is used as a measure of the depth of the scene, then a defocus edge $i(x)$ can be given by

$$i(x) = f(x) \otimes c(x, y, x_0, y_0, \sigma), \tag{3}$$

$$\text{and} \quad c(x, y, x_0, y_0, \sigma) = \frac{1}{2\pi} \cdot \frac{\sigma}{\left[(x - x_0)^2 + (y - y_0)^2 + \sigma^2 \right]^{3/2}}, \tag{4}$$

where x_0 and y_0 is the location parameter, σ is the scale parameter, which affects the shape of Cauchy distribution dropping from the peak to low. For convenience and brevity, the following have taken x_0 and y_0 as 0, and omitted to write.

3. Edges Defocus Blur Estimate

A step edge is re-blurred twice using two known Cauchy kernels with scale parameter σ_1, σ_2, respectively. Then the ratio between the first re-blurred gradient magnitude of the step edge and its second re-blurred version is calculated. The ratio is maximum at the edge location. Using the maximum value, we can calculate the amount of the defocus blur of an edge.

For convenience and simplicity, we describe our blur estimation algorithm for 1D case firstly and then extend it to 2D image. The gradient of the first re-blurred edges is

$$\nabla i_1(x) = \nabla \left(i(x) \otimes c(x, \sigma_1) \right) = \nabla \left\{ \left[\left((Au(x) + B) \otimes c(x, \sigma) \right) \right] \otimes c(x, \sigma_1) \right\}. \tag{5}$$

Depending on the nature of convolution, the Equation (5) can be rewritten as

$$\nabla i_1(x) = \left\{ \left[\left(A\nabla[u(x)] + B \right) \right] \otimes c(x, \sigma) \otimes c(x, \sigma_1) \right\}. \tag{6}$$

We know that the derivative of the unit step function is a unit impulse function $\delta(x)$, then (6) becomes

$$\nabla i_1(x) = \left\{ A\delta(x) \otimes c(x,\sigma) \otimes c(x,\sigma_1) \right\}. \tag{7}$$

Take the Fourier transform of both sides

$$\begin{aligned} \mathcal{F}\left[\nabla i_1(x)\right] &= \mathcal{F}\left[A\mathcal{E}(x)\right] \cdot \mathcal{F}\left[c(x,\sigma)\right] \cdot \mathcal{F}\left[c(x,\sigma_1)\right] \\ &= A \cdot \mathcal{F}\left[c(x,\sigma)\right] \cdot \mathcal{F}\left[c(x,\sigma_1)\right]. \end{aligned} \tag{8}$$

Since $c(x,\sigma) = \dfrac{1}{\pi} \cdot \dfrac{\sigma}{x^2+\sigma^2}$, and $c(x,\sigma_1) = \dfrac{1}{\pi} \cdot \dfrac{\sigma_1}{x^2+\sigma_1^2}$, so that

$$\mathcal{F}\left[c(x,\sigma)\right] \cdot \mathcal{F}\left[c(x,\sigma)\right] = A \cdot \mathcal{F}\left[\frac{1}{\pi} \cdot \frac{\sigma}{x^2+\sigma^2}\right] \cdot \mathcal{F}\left[\frac{1}{\pi} \cdot \frac{\sigma_1}{x^2+\sigma_1^2}\right] \tag{9}$$

According to the Fourier pair, $e^{-a|x|} \leftrightarrow \dfrac{2a}{a^2+\omega^2}$ [22], and the symmetry of the Fourier transform, we can get

$$\frac{2a}{a^2+x^2} \leftrightarrow 2\pi e^{-a|\omega|}. \tag{10}$$

Following the linear properties of the Fourier transform, the two Fourier transform terms on the right side of Equation (9) are given as follows

$$\mathcal{F}\left[\frac{1}{\pi} \cdot \frac{\sigma}{x^2+\sigma^2}\right] = e^{-\sigma|\omega|}, \text{ and } \mathcal{F}\left[\frac{1}{\pi} \cdot \frac{\sigma_1}{x^2+\sigma_1^2}\right] = e^{-\sigma_1|\omega|}. \tag{11}$$

By substituting (11) into (9), we get

$$\mathcal{F}\left[c(x,\sigma)\right] \cdot \mathcal{F}\left[c(x,\sigma_1)\right] = e^{-(\sigma+\sigma_1)|\omega|}. \tag{12}$$

Substituting (12) into Equation (8), there are

$$\mathcal{F}\left[\nabla i_1(x)\right] = A \cdot e^{-(\sigma+\sigma_1)|\omega|}. \tag{13}$$

After performing the inverse Fourier transform of (13), we can get

$$\nabla i_1(x) = \frac{2A}{\pi} \cdot \frac{\sigma+\sigma_1}{x^2+(\sigma+\sigma_1)^2}. \tag{14}$$

Similarly, we can get the second re-blurred gradient magnitude of the step edge

$$\nabla i_2(x) = \nabla\left(i(x) \otimes c(x,\sigma_2)\right) = \frac{2A}{\pi} \cdot \frac{\sigma+\sigma_2}{x^2+(\sigma+\sigma_2)^2}, \tag{15}$$

where σ is the original image of the scale parameter for the Cauchy distribution function; σ_1 and σ_2 are two re-blurred scale parameters. The gradient magnitude ratio between the twice re-blurred edges R is

$$R(x) = \frac{\left|\nabla i_1(0)\right|}{\left|\nabla i_2(0)\right|} = \frac{\sigma+\sigma_1}{\sigma+\sigma_2} \cdot \left[\frac{x^2+(\sigma+\sigma_2)^2}{x^2+(\sigma+\sigma_1)^2}\right] \tag{16}$$

It can be proved that the ratio $R(x)$ is maximum at the edge location ($x = 0$), assumed $\sigma, \sigma_1, \sigma_2 > 0$ and $\sigma_1 < \sigma_2$. The maximum value $R_{\max}$ is given by

$$R_{\max} = \frac{|\nabla i_1(0)|}{|\nabla i_2(0)|} = \frac{\sigma + \sigma_2}{\sigma + \sigma_1}. \tag{17}$$

Thus, given the maximum $R_{\max}$ and let $\sigma_1 < \sigma_2$, the unknown blur amount σ can be calculated by

$$\sigma = \frac{\sigma_2 - R_{\max}\sigma_1}{R_{\max} - 1}. \tag{18}$$

In order to achieve a 2D image blur estimation, we use 2D isotropic Cauchy distribution function to re-blur the input image, and blur estimation is similar to 1D case. In the 2D image, the gradient magnitude can be calculated as follows:

$$\|\nabla i(x, y)\| = \sqrt{\nabla i_x^2 + \nabla i_y^2}, \tag{19}$$

where ∇i_x and ∇i_y are gradient in x and y directions.

4. The Whole Scene Depth Map Extraction

After obtaining edge position blur amount estimation, we get a sparse depth estimation map $\hat{d}(x)$. In order to get the full depth map $d(x)$ of the entire image, we need to propagate the sparse depth estimation map $\hat{d}(x)$ from edge locations to the entire image. To achieve this and compare with other PSF model, we apply the matting Laplacian to perform the defocus map interpolation, same as [16]. Formally, the depth interpolation problem can be formulated as minimizing the following cost function:

$$d = \frac{\lambda D \hat{d}}{L + \lambda D} \tag{20}$$

Here, $\hat{d}$ and d are sparse depth map vector representation of $\hat{d}(x)$ and full depth map $d(x)$. D is a diagonal matrix, λ is the balance parameter, L is a matting Laplacian matrix. For the detailed explanation of the expansion process and parameters, readers can refer to [16].

5. Results

We test the proposed method on a PC with a 2.5 GHz Intel Core i5 Processor. As for contrastive comparison, the Zhuo and Sim's method [16] and Fang *et al.* method [17] are used to calculate the blur map for the same images.

The different steps of our proposed algorithm for the white flower image are displayed in **Figure 2**. The color in each color bar changed continuously from blue to red represents a number of small to large, also represents the depth from near to far (the same figure). The foreground objects in the white flower image are three white flowers. The focus point is on the white petals on the bottom of the image. The depth of the scene changes continuously from the bottom to the top of the image. As shown in **Figure 2**, the sparse depth map (**Figure 2(d)**) gives an accurate and reasonable measure of amount of edge blur. The depth map (**Figure 2(e)**) accurately captures the continuous change of the depth in this scene image. The foreground and background are well separated.

As shown in Figure 3, we compare our method with the Zhuo *et al.*'s method [16]. Both methods generate reasonable layered depth maps. The depth map reflects the continuous change of the depth. In the building image, there are mainly 3 depth layers in the scene: the wall in the nearest layer, the buildings in the middle layer, and the sky in the farthest layer. However, our method has higher accuracy in local estimation and thus, our depth map captures more details of the depth in Figure 3(c). As shown in the figure, the difference in the depth of the left and right arms can be perceived in our result. In contrast, the Zhuo *et al.*'s method does not recover this depth difference in Figure 3(b).
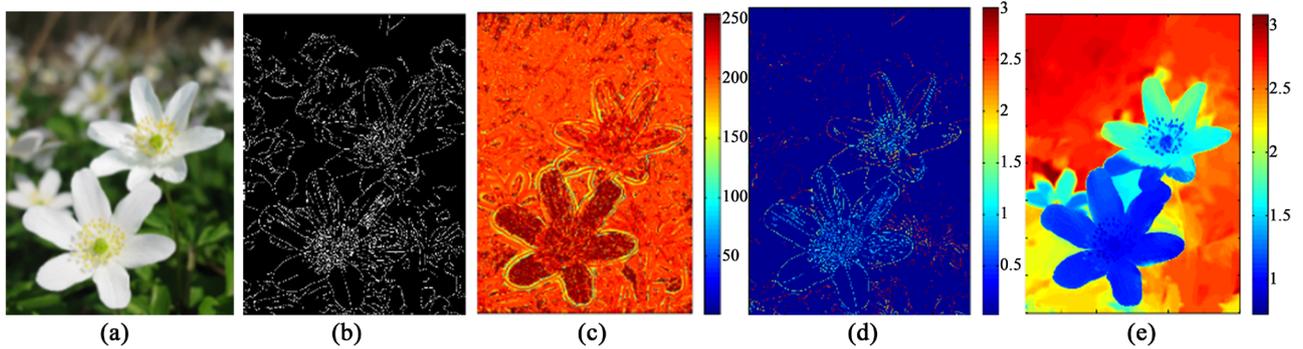
Figure 2. The different steps of our proposed algorithm for the white flower image. (a) Input image; (b) Edge; (c) Ratio of gradient; (d) Sparse depth map; (e) Full depth map.
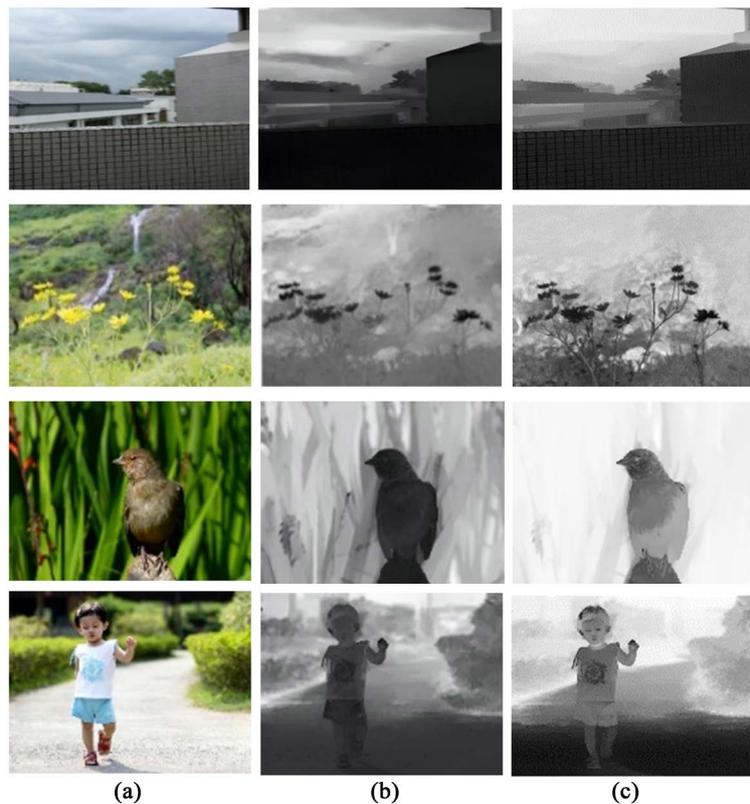
Figure 3. Comparison of our method with the Zhou's method in some different scenes. (a) Input image; (b) Zhou's result; (c) Our result.

In **Figure 4**, we test our method on the pumpkin image, and compare our method with both Zhuo *et al.*'s method [16] and Fang *et al.*'s method [17]. In the Pumpkin image (**Figure 4(a)**), the depth of the scene changes continuously from the bottom to the top of the image. Our method is able to produce defocus maps corresponding to those layers. As shown in **Figures 4(b)-(d)**, we see that, the result of Zhuo *et al.* is a grayscale image, its intensity changes gradation from black to white, represents the depth changing from near to far. In the color images, the meaning of the color is same as the former. All three methods can generate a reasonable layered depth map. But Zhuo's result at the strong edge appeared estimation error, as shown in **Figure 4(b)**, the stem of the pumpkin at the left side of the middle in this scene. The depth estimation there is a significant error. As shown in **Figure 4(c)**, the method of Fang *et al.* eliminates the estimation error of Zhuo *et al.*'s method, but the shape of objects in the scene do not be recognized, and the depth layer changes also significantly rough. In contrast, our method is able to produce a more accurate and continuous defocus map. The proposed method not only identifies the shape of pumpkins, but also more accurately restored both objects and a detail continuously change in this scene.

A comparison of our method with the focal stack method [23] is shown in **Figure 5**. Depth recovery from this image is quite challenging due to the complex structure of the scene. The focal stack method uses 14 images with different
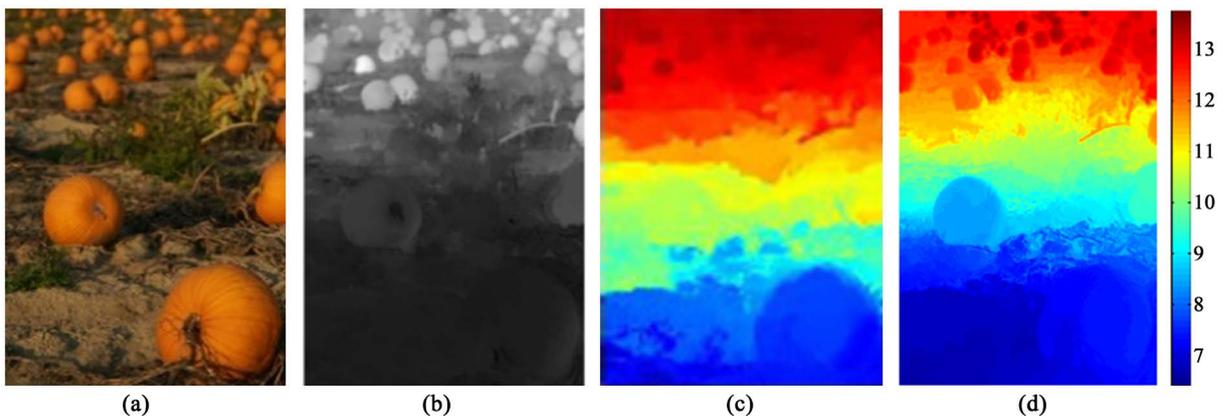
Figure 4. Comparison of our method with both the Zhou's and Fang's method. (a) Input image; (b) Zhou's result; (c) Fang's result; (d) Our result.
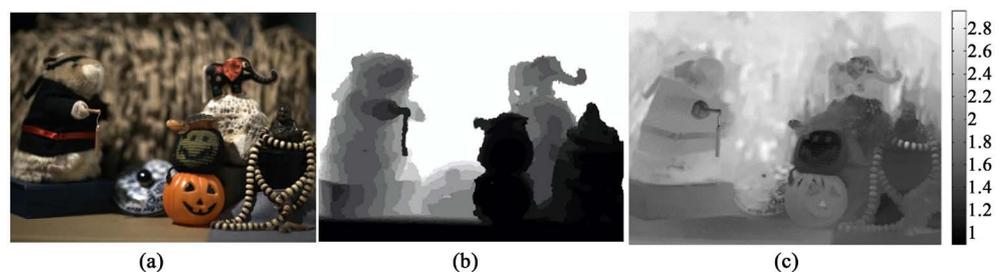
Figure 5. Comparison of our method and focal stack method. (a) Input image; (b) The result of focal stack method; (c) Our result.

focus settings to produce the layered depth map. Our method is able to generate a comparable result using just one of the 14 images.

6. Conclusion

In this paper, we presented a new method to calculate the blur amount at edge locations based on the Cauchy gradient ratio. A full defocus map is then produced using the matting interpolation. Experimental results on some real images show that our method can accurately recovery depth from an un-calibrated single defocused image. It demonstrates that our method is robust to noise, inaccurate edge location and interferences of neighboring edges and is able to generate more accurate defocus maps compared with existing Gaussian based PSF methods. It also shows the non-Gaussian PSF Model is feasibility as that is pointed out by Ens [18] and Subbarao [19]. In the future, we would like to extend our method to recover depth by combining our method with other monocular cues, e.g., geometric cues or textures change etc. to further improve the accuracy of the depth recovery.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Wu, C., Yang, Y. and You, F. (2006) A Depth Measurement Approach Based on Integral Imaging and Multiple Baseline Stereo Matching Algorithm. *Acts Electronics Sinica*, **34**, 1090-1095.

[2] Hassner, T. and Basri, R. (2006) Example Based 3D Reconstruction from Single 2D Images. *Conference on Computer Vision and Pattern Recognition Workshop*, *CVPRW'06*, New York, 17-22 June 2006, 15. https://doi.org/10.1109/CVPRW.2006.76

[3] Saxena, A., Schulte, J. and Ng, A. (2007) Depth Estimation Using Monocular and Stereo Cues. *Proceedings of the* 20*th International Joint Conference on Artificial Intelligence*, Hyderabad, 6-12 January 2007, 7 p.

[4] Liu, B., Gould, S. and Koller, D. (2010) Single Image Depth Estimation from Predicted Semantic Labels. 2010 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), San Francisco, 13-18 June 2010, 1253-1260. https://doi.org/10.1109/CVPR.2010.5539823

[5] Su, C., Cormack, L., Bovik, A., *et al.* (2013) Depth Estimation from Monocular Color Images Using Natural Scene Statistics Models. 2013 *IEEE* 11*th IVMSP Workshop*: 3*D Image/Video Technologies and Applications* (*IVMSP* 2013), Yonsei, 10-12 June 2013, 4. https://doi.org/10.1109/IVMSPW.2013.6611900

[6] Yuan, H., Wu, S., An, P., Zheng, Y. and Xu, L. (2014) Object Guided Depth Map Recovery from a Single Defocused Image. *Acta Electronica Sinica*, **42**, 2009-2015. (In Chinese)

[7] Hoiem, D., Efros, A. and Hebert, M. (2011) Recovering Occlusion Boundaries from an Image. *International Journal of Computer Vision*, **91**, 328-346. https://doi.org/10.1007/s11263-010-0400-4

[8] Elder, J. and Zucker, S. (1998) Local Scale Control for Edge Detection and Blur Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 699-716. https://doi.org/10.1109/34.689301

[9] Bae, S. and Durand, F. (2007) Defocus Magnification. *Computer Graphics Forum*, **26**, 571-579. https://doi.org/10.1111/j.1467-8659.2007.01080.x

[10] Pentland, A. (1987) A New Sense for Depth of Field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, 523-531. https://doi.org/10.1109/TPAMI.1987.4767940

[11] Moreno-Noguer, F., Belhumeur, P.N. and Nayar, S.K. (2007) Active Refocusing of Images and Videos. *ACM Transactions on Graphics*, **26**, 67-75. https://doi.org/10.1145/1276377.1276461

[12] Levin, A., Fergus, R., Durand, F., *et al.* (2007) Image and Depth from a Conventional Camera with a Coded Aperture. *SIGGRAPH* 2007, San Diego, 5-8 August 2007, 9 p. https://doi.org/10.1145/1275808.1276464

[13] Veeraraghavan, A., Raskar, R., Agrawal, A., *et al.* (2007) Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing. *SIGGRAPH* 2007, San Diego, 5-8 August 2007, 12 p. https://doi.org/10.1145/1275808.1276463

[14] Namboodiri, V. and Chaudhuri, S. (2008) Recovery of Relative Depth from a Single Observation Using an Uncalibrated (Real-Aperture) Camera. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Anchorage, 24-26 June 2008, 1-6. https://doi.org/10.1109/CVPR.2008.4587779

[15] Lee, S., Lee, J., Hayes, M., *et al.* (2012) Single Camera-Based Full Depth Map Estimation Using Color Shifting Property of a Multiple Color-Filter Aperture. *IEEE ICASSP*, Kyoto, 25-30 March 2012, 801-804. https://doi.org/10.1109/ICASSP.2012.6288005

[16] Zhuo, S. and Sim, T. (2011) Recovering Depth from a Single Defocused Image. *Pattern Recognition*, **44**, 1852-1858. https://doi.org/10.1016/j.patcog.2011.03.009

[17] Fang, S., Qin, T., Cao, Y., *et al.* (2013) Depth Recovery from a Single Defocused Image Based on Depth Locally Consistency. *ACM* 5*th International Conference on Internet Multimedia Computing and Service*, Huangshan, 17-18 August 2013, 56-61.

[18] Ens, J. and Lawrence, P. (1993) An Investigation of Methods for Determining Depth from Focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 97-108. https://doi.org/10.1109/34.192482

[19] Subbarao, M. and Gurumoorthy, N. (1988) Depth Recovery from Blurred Edges. *Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition*, *CVPR* '88, Ann Arbor, 5-8 June 1988, 498-503. https://doi.org/10.1109/CVPR.1988.196281

[20] Ming, Y. and Jiang, J. (2008) Moving Object Detection of Infrared Video Based on Cauchy Distribution. *Journal of Infrared, Millimeter, and Terahertz Waves*, **27**, 65-72. (In Chinese) https://doi.org/10.3724/SP.J.1010.2008.00065

[21] Hecht, E. (2001) Optics. 4th Edition, Addison Wesley, Boston.

[22] Guan, Z. and Xia, G. (2004) Signals and Linear System Analysis. 4th Edition, Advance Education Press, Beijing. (In Chinese)

[23] Hasinoff, S. and Kutulakos, K. (2008) Light-Efficient Photography. 10*th European Conference on Computer Vision*, *ECCV* 2008, Marseille, 12-18 October 2008, 45-59. https://doi.org/10.1007/978-3-540-88693-8_4