DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

A novel approach of generating stereoscopic images using defocus

Tianteng Bi, Yue Liu, Dongdong Wong, Yongtian Wang; Beijing Engineering Research Center of Mixed Reality and Advanced Display， School of Optoelectronics, Beijing Institute of Technology, 5 South Zhongguancun Street, 100081; Beijing, China

Abstract

Direct 2D-to-3D conversion is an important task with various manual or automatic methods to satisfy the rapidly increasing requirements on stereoscopic contents. In order to construct stereoscopic images, defocus is adopted to estimate the depth from a single image and a novel layered structure is proposed in this paper. A simple searching and filtering-based method is proposed for finding the best layering result with a comparison of performances between our method and certain existing unsupervised learning methods to determine the optimum approach for automatic choice of the layered threshold, which can be used to generate a set of layered RGB images by combining with the depth map. Then the planar layered model is constructed and texture is mapped from image onto the planes. With the help of the proposed approach, a layered model can be created from a single defocused image to provide stereoscopic feelings. The proposed approach has also been expanded to construct layered stereoscopic panorama, which shows its great application potentials.

I. Introduction

With the growth of 3D hardware such as TV, mobile phone and video camera, more and more widespread 3D applications appeared in recent years and the requirements on 3D contents have also increased rapidly. Because there exist large amounts of 2D source, direct 2D-to-3D conversion methods become an important research issue. Many manual or automatic conversion methods have been proposed [1-5], in which the automatic methods are actually depth estimation to be used for generating a pair of right image and left image from a single still image.

Although depth estimation is a classical problem in computer vision, inferring the depth of a scene from a single image remains an extremely difficult problem. Most existing works on 3D reconstruction require the correspondence of multiple images. Stereo vision based approaches [6-7] use the computed disparities between a pair of images of the same scene taken from two different viewpoints to recover the depth. Shape from motion (SFM) [8-9] uses the correspondences between images to obtain the 2D motion field to recover the 3D motion and the depth. Depth from focus (DFF) [10-11] captures a set of images using different focus settings and measures the sharpness of image for each pixel, in which the depth of the pixel depends on the image that the pixel is selected from. These methods not only rely on correspondences between images, but also suffer from occlusion problem. More importantly, they cannot work for a single image scenario.

Humans are good at judging depth from a single image by combining such monocular cues as texture and defocus. For example, the texture of an object is different at different location. Among the single view depth cues, defocus is one of the strongest that allow humans to understand the order of the objects in a scene. This depth cue has been extensively investigated in depth estimation from a single viewpoint [12].

Most existing depth from defocus (DFD) methods require two or more images of the same scene which are taken at the fixed viewpoint with different focus and aperture settings [13-18] and the disadvantage of such methods is that the scene must be invariant during the long capture process.

Single-image based DFD approaches only need one image to compute the depth of the scene, which simplifies the capture procedure. Levin *et al.* proposed an algorithm using a coded aperture which is more sensitive to the depth variation [19]. The depth can be obtained by a set of calibrated blur kernels after a deconvolution process. Chen *et al.* represented the defocus blur amount by the energy spectra of the point spread function and detected the defocused step edge to recover depth with camera settings [20]. Zhuo's approach employed edge-detection methods to first estimate the defocus blur of the step edge based on Gaussian gradient ratio, then generated the dense defocus map by using interpolation [21]. The single-image based DFD approaches can be divided into step edge detection, defocus blur amount estimation and defocus map interpolation, during which many researchers use a parameterized model to formulate the edge blurred by the point spread function and recover the depth by estimating the parameters.

In this paper, we focus on generating visually-pleasing stereoscopic images based on depth obtained by using DFD from a single still image, during which the defocused image can be used to obtain depth of each pixel by computing the blur of pixels. Furthermore, inspired by Hoiem's pop-up method [22], we propose a layered structure for the purpose of generating stereoscopic models rather than a conventional pair of right image and left image, which converts the 2D images into a pop-book like images. At the same time, the depth data generated by DFD methods are clustered according to the number of the layers that is simply assigned to two in this paper.

For automatic classifying the pixels of an RGB image into different categories representing different layers in the stereoscopic image, we propose a simple searching and filtering-based algorithm to find the threshold of depth data. Certain unsupervised learning algorithms are adopted to classify the original RGB image into two layers according to the input depth map. A comparison of performances between our method and such unsupervised learning methods as K-means clustering and Gaussian Mixture Model is conducted to determine which method is the best fit for clustering the depth data. By performing the above-mentioned procedures, a set of layered RGB images can be generated and each layer belongs to certain part of the original RGB image.

Once the layered images are determined, the construction of the stereoscopic image of a scene is a simple matter of specifying plane positions and texture mapping from the images onto the planes. The proposed approach has also been expanded to construct layered stereoscopic panorama, which can be easily implemented by changing the conventional 2D image to panorama and using layered cylindrical model instead of the planes. The proposed approach can generate qualitatively correct stereoscopic

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

images and provide more stereoscopic feeling to a user when exploring the captured 2D images.

The rest of this paper is organized as follows. Section II presents the principle of DFD. Section III describes the layering process using the proposed method and unsupervised learning methods. Section IV shows the pop-up book like image and layered stereoscopic panorama to which the proposed methods are expanded. Finally the discussion and conclusion are presented in this section.

II. The principle of DFD

DFD utilizes the blur amount of each pixel in the image to compute the depth of a corresponding point. Assuming that the optical system of a camera obeys the thin lens model [23], the relationship between a point and its image can be formulated as:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \tag{1}$$

where u is the distance from a point in the scene to the lens, v is the distance between the lens and the plane that the point is exactly focused and f is the focal length of the lens.

The principle of DFD can be explained as shown in Figure 1. Due to the depth of field of lens, only the objects at one particular distance along the optical axis (O.A.) from the lens to the scene are exactly focused. If the objects are away from this location, their images will be blurred.

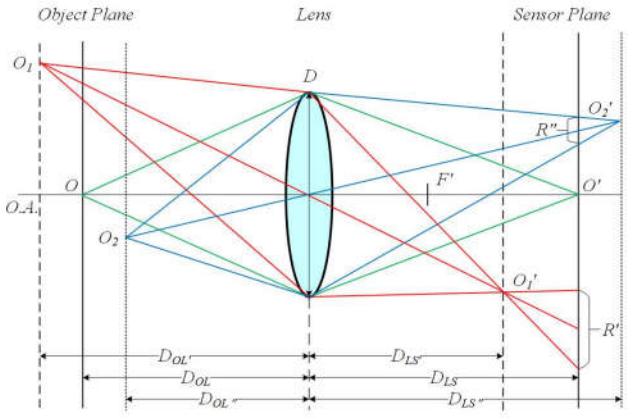

***Figure 1**. The geometrical optics principle of DFD.*

Figure 1 shows a simplified structure of a conventional camera whose distance between lens and sensor plane is invariant. For this kind of lens system, such points as point O at the location whose distance to the lens is D_{OL} can be perfectly focused on the sensor plane. Such points as point O_1 and point O_2 that are away from the plane at which the point O located cannot exactly be focused on the sensor, which means that they are imaged at place either closer or further to the sensor on which a blur circle is formed.

The amount of defocus blur denoted by the radius of blur circle is related with the distance between a point and the lens as shown in Figure 1. Therefore, if the radius of blur circle can be measured, the distance to the lens can be recovered with the system parameters of the camera.

According to the Eq. (1), we can see that

$$\frac{1}{D_{OL'}}+\frac{1}{D_{LS'}}=\frac{1}{F'} \tag{2}$$

where $D_{OL'}$ is the distance between point O_1 and the lens, $D_{LS'}$ is the distance between the lens and the plane where O_1 is exactly focused and F' is the focal length of the lens system.

As shown in Figure 1, let D be the diameter of the lens and R' the radius of the blur circle that the point O_1 forms on the sensor plane. According to the theorem of similar triangles, we have

$$\frac{D/2}{R'}=\frac{D_{LS'}}{D_{LS}-D_{LS'}} \tag{3}$$

After combining Eq. (2) and Eq. (3) we can obtain

$$D_{OL'}=\frac{D_{LS}F'}{D_{LS}-F'-2FR'} \tag{4}$$

where $F=F'/D$ is the F-number of the camera and D_{LS} is the invariant distance between the lens and the sensor plane.

For the other point O_2 which is closer to the lens than the point O_1, similarly we have

$$D_{OL''}=\frac{D_{LS}F'}{D_{LS}-F'+2FR''} \tag{5}$$

Eq. (4) and Eq. (5) describe the relationship between the radius of blur circle and the depth of the corresponding point in the scene under different circumstances.

The above-mentioned equations include such parameters of camera lens system as the diameter and the focal length of the lens which have an impact on the amount of the blur circle as shown in Figure 2 and Figure 3.

Figure 2 illustrates the variation of blur circle with the diameter of the lens. When the location of a point in the scene is fixed, its blur circle on the sensor is smaller when the diameter of the lens is smaller.

Figure 3 shows how the blur circle changes with the focal length of the lens. If the focal length is smaller than the right focal length that can exactly focus the point on the sensor plane, the radius of blur circle becomes smaller with the longer focal length. Once the focal length is larger than the right focal length, the radius of the blur circle will be larger when the focal length of the lens is longer. Thus, the radius of blur circle is a non-linear function of the focal length of lens.

According to the geometrical optics, the intensity distribution of blur circle is approximately uninform. When considering the diffraction effect, the blur circle is not a single circle but a centric circle with a set of alternately dark and bright rings. Moreover, most light energy concentrates on the centric circle and only little energy distributes on the set of dark and bright rings, which is

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

related to the wavelengths of light. Other factors such as chromatic aberration and distortion can also influence the energy distribution. Although the accurate description of this phenomenon is very complex, it can be best approximated to a two dimensional Gaussian function called point spread function (PSF) after considering all the elements. The two dimensional Gaussian function has the following form:

$$g(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(\frac{x^2+y^2}{2\sigma^2}\right) \tag{6}$$

where σ is the standard deviation called spread parameter that is proportional to the radius of blur circle.

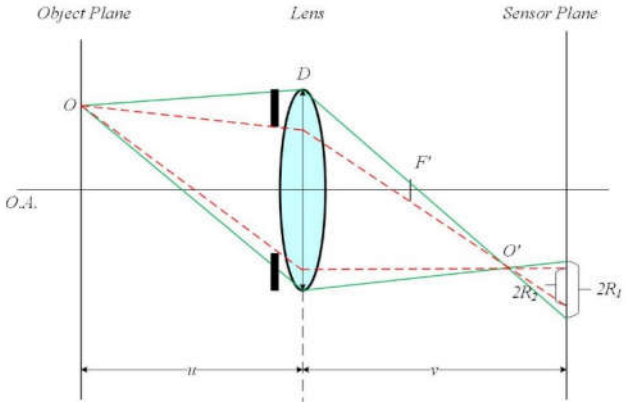

***Figure 2**. The relationship between the diameter of lens and the radius of blur circle.*

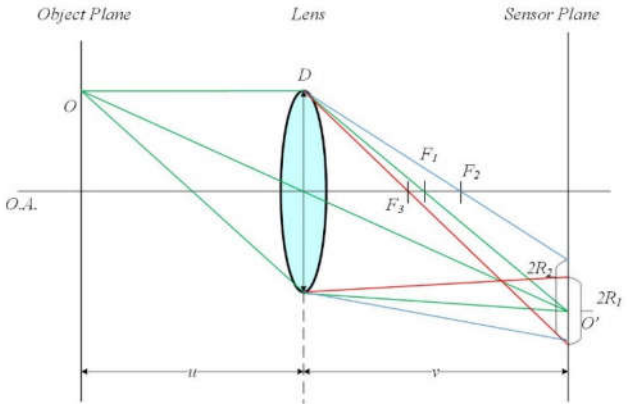

***Figure 3**. The changes of blur circle with the different focal length of lens.*

The radius of blur circle R can be formulated as:

$$R=k\sigma \tag{7}$$

where the proportionality constant k depends on the particular optical system.

In order to recover the absolute depth, k needs to be included in Eq. (4) as:

$$D_{OL'}=\frac{D_{LS}F'}{D_{LS}-F'-2Fk\sigma} \tag{8}$$

The defocus blur process can be modeled as a convolution operation of a focused image and the PSF. A blurred image is then given by:

$$i(x,y)=f(x,y)\otimes g(x,y) \tag{9}$$

where $i(x,y)$ is the blurred image, $f(x,y)$ is the full-focused image which does not exist reality. $g(x,y)$ is the same as the previous definition.

Ⅲ. Methods

This paper proposes an approach to generate the stereoscopic images by using defocus, which includes depth estimation and pixel classification. We use Zhuo's approach [21] to obtain a dense defocus map which is actually a relative depth map. In order to classify each pixel, the following methods are employed to partition each pixel into two layers.

A. Depth estimation

The DFD approach can obtain the defocus map of a single image captured by a conventional camera without calibration. The value of each pixel in the defocus map represents the degree of blur of the corresponding point in the scene, which actually represents the depth of the point in the scene. Although the absolute depth cannot be recovered with this algorithm, the relative relationship between the points can be determined with fairly good accuracy. The algorithm is applied in a simple way in which the captured image can be directly input to the program. The conventional camera without any particular processing is good enough for the algorithm and the computing efficiency depends on the size of the input image.

The proposed algorithm includes edge detection, edge defocus estimation and defocus propagation. The algorithm first re-blurs the step edge detected by canny operator [24] in the image. A Gaussian kernel is employed to re-blur the step edge and its standard deviation is determined manually. Then the gradient of the step edge and the re-blurred step edge is calculated respectively. Since the ratio between the gradient magnitude of the step edge and its blurred version is a non-linear function of location that achieves its maximum at the step edge, the defocus at the edge that is related to the ratio can be calculated.

The maximum value of the ratio function depends on both the standard deviation of the re-blur Gaussian kernel and the unknown spread parameter of the PSF. If the standard deviation of the re-blur Gaussian kernel is explicit, the maximum of the ratio function is only decided by the spread parameter of the PSF. The spread parameter evaluating the amount of blur can be calculated given the maximum value that is obtained by using the step edge map and the re-blurred one.

After computing the spread parameter, a sparse inaccurate defocus map is obtained due to the noise and weak edges. In order to improve the quality of the sparse defocus map, the joint bilateral filter [25] is used to modify the sparse defocus map on the edge location.

The last step is the defocus map interpolation which propagates the defocus from the step edge locations to the entire

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

map. This process can generate a dense defocus map instead of a sparse defocus map. The first issue to be solved is that the dense defocus map should be close to the sparse defocus map and the second issue is that the defocus blur discontinuities should be aligned to the edge. In this algorithm, the matting Laplace [26] is used to perform the defocus map interpolation formulated as the minimization of a cost function in reality.

After the above-mentioned procedure, a dense defocus map that shows the relative depth can be obtained. As shown in Figure 4, the proposed method is tested using the images downloaded from the Internet. The input images in the left column are conventional RGB images and those in the right column are the full defocus map. The darker part in the defocus map means the smaller depth.

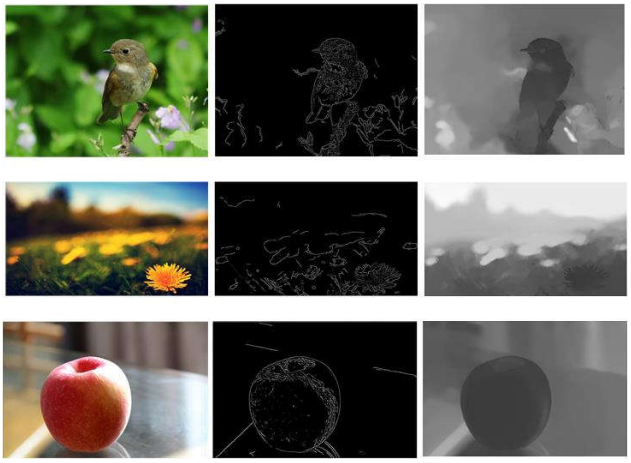

***Figure 4**. The test results of the depth estimation algorithm. The left column is the input images, the middle column is the sparse defocus map and the right column is the full defocus map.*

This algorithm can work well on most images and generate an acceptable depth map. However, the depth map may be incorrect at certain locations. For example, the region marked by a red rectangular in Figure 5 seems to be farther than the fruit with greater magnitude of blur, but the depth map shows a nearly equal value to the foreground fruit, which incorrectly shows that the wall is closer than the fruits to the camera. This may be caused by the textureless wall, which is the common problem of this kind of approach.

Some similar circumstances may also happen on the textured regions. As shown in the region marked by a blue rectangular in Figure 5, the blurred flower in the original RGB image is in front of the green hill, however, the gray level of the hill is lower than the one of flower according to the defocus map, which shows a wrong relative depth relationship.

B. Pixel classification

In this part, we propose a layered structure for the purpose of generating pop-up book like images, which requires the classification of each pixel in the RGB image into two layers according to depth data. Figure 6 shows the processing pipeline. First a conventional RGB image is used to recover the depth of the scene. The DFD method can then be employed to calculate the blur of each pixel to obtain the full depth map. The depth map shows the relative location relationship between the objects, which contributes to cut the image into some layers. At the top row of Figure 6 are the two new images that belong to certain parts of the original RGB image respectively. This step can be implemented by setting a depth threshold to distinguish the foreground and the background in the image such as the focused green orange and the farther blurred oranges in Figure 6.

In order to automatically classify the conventional RGB image into different layers according to the depth data, the depth threshold should be determined. A searching and filtering-based method to find the depth threshold is proposed that includes frequency domain filtering and depth threshold searching. In the proposed method, the data of the depth histogram is taken as the input of the algorithm and the output is the depth threshold. When the pixel value in the depth map is less than the threshold, the corresponding pixel in the RGB image is considered as belonging to the foreground. Otherwise, the pixel in the RGB image is assigned to the background.

***Figure 5**. Error locations in the depth map. The left column is the conventional images and the right column is the depth map.*

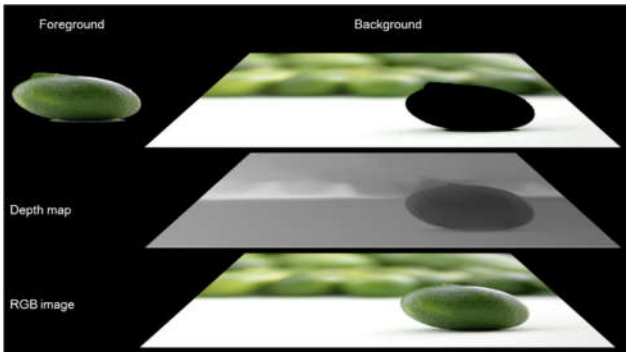

***Figure 6**. Pipelines of cutting image into two layers. From bottom to top: RGB image, depth map, layered images including foreground and background.*

Specifically, the depth histogram that is not smooth enough to search the depth threshold is first obtained on the basis of depth data. In the viewpoint of digital signal processing [27], the depth histogram can be regarded as a signal with high frequency noises as shown in Figure 7. In order to exactly search the depth threshold, we first filter the “Histogram signal” in the frequency domain to remove the high frequency noise, then generate a smooth curve used for searching the depth threshold.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

The depth threshold always appears as the minimal value of the curve, but not every minimal value can be a threshold. Based on the properties of the images, the most appropriate depth threshold must be at the middle of the two main high peaks which are found by filtering the depth histogram. Besides, during the searching process, we consider the difference between the peak-peak and the minimal value as a constraints to achieve optimal filter. After each filtering, let $\mathbf{d} = (d_1, d_2, ..., d_N)$ denote the values of the depth map from small to large, then the problem can be formulated as:

$$d_{opt} = \arg\min_{\hat{\mathbf{d}}} F(x) \tag{10}$$

where d_{opt} is the depth threshold, $F(x)$ represents the curve function and the value area of x is $\mathbf{d}$, $\hat{\mathbf{d}} = \left(\min\left(d_p^{(1)}, d_p^{(2)}\right), ..., \max\left(d_p^{(1)}, d_p^{(2)}\right)\right)$ is the values of depth map and $d_p^{(1)}$, $d_p^{(2)}$ denote the values of depth map corresponding to the peak-peak and the second peak.

This constraint can avoid finding other undesirable minimal value. Let $F_p^{(1)}$ be the peak-peak and set k empirically from the integral 5 to 100, then the constraint can be presented as:

$$F_p^{(1)} - kF\left(d_{opt}\right) \geq 0 \tag{11}$$

By checking the constraint, the filtering result can be decided to be adopted or not. If the result cannot satisfy the constraint, the algorithm will start a new filtering and repeat the above-mentioned searching process. The filtering works according to an assigned parameter.

For the purpose of comparison, two classical unsupervised learning algorithms are introduced to cluster the depth map into two categories. The first one is K-means clustering and the other one is Gaussian Mixture Model (GMM). K-means clustering is a popular unsupervised machine learning method that partitions n observations to k clusters. GMM is a probabilistic model that is composed of the weighted sum of a set of Gaussian distributions, whose training process aims to determine the distributions representing the categories. Then the data is mapped to the distributions to calculate their probabilities used to classify the data.

As shown in Figure 8, the performance of the proposed method, K-means clustering and GMM on test images are evaluated. The depth map is firstly calculated and considered as the input image to partition the RGB image into two layers. In the bird image and orange image, our method can find the depth threshold precisely. The layered result is better than GMM and K-means clustering. In the layered image of the other two methods, there are some "noisy pixels" partitioned by mistake. Especially, GMM fails to classify the pixel of the orange image into two layers, probably because the depth value in the depth map doesn't satisfy the Gaussian distribution. Similar circumstance also happens in girl image and bridge image. For the girl image, K-means clustering achieves the best result compared with the proposed method that lost certain part of the face. In the bridge image, there are some blurred bright light spots whose depth is wrong, which leads to further rough clustering. All the three approaches can achieve good results in the flower image. GMM can obtain the white flower with clear outline, however there are still little "noisy pixels" in the center of the flower. Our method and K-means clustering can extract the flower perfectly, while our method has less pixels of green leaf in comparison with K-means clustering. Besides, expectation maximization (EM) algorithm is used to determine the parameters of each Gaussian distribution in GMM when employing GMM to classify the RGB image. The EM algorithm needs to run iteratively to achieve convergence, which means that it takes more time for GMM than the other two methods. Our method and K-means clustering are more appropriate for this application because K-means is more robust and our method is more accurate to certain extent. During the following steps, we mainly use our method and K-means clustering to layer the RGB images.

Ⅳ. Discussion and Conclusion

In order to apply layered models to generate stereoscopic feelings, we construct two planar models for two layered RGB images and use them as the texture of the plane. As shown in Figure 9, the 2D images are converted to stereoscopic images with two layers. The relative position of two layers will lead to the occlusion between foreground and background and show which part is closer in a direct way. Besides, the blank generated by removing the closer part from the image makes the foreground pop-up forward, which gives its observer stereoscopic feeling. Actually, the key is recovering the relative position relationship between the object and background. Once the relative relationship is obtained, the stereoscopic impression emerged naturally when setting the objects according to their relative position.

Because the DFD method we employ has no requirements about the scale of the input image, we can also test the complete process on the panorama image. As shown in Figure 11, we capture a cylindrical panorama of Dashuifa in Yuanmingyuan Garden (the former Summer Palace). 16 images are captured by a conventional camera with a tripod and stitched into a large panorama. Actually the layered result is not satisfactory and the main reason is there is little blur in the panorama image. Large scale outdoor scenes need short focal length to be captured, which leads to the long depth of field. When capturing images under this condition, the blur effects will not be obvious. Thus, the DFD methods cannot work in this scenario. It can be seen from Figure 10 that the depth histogram of our panorama only shows a single peak, which improperly indicates that all the objects in the image have the same depth. Therefore, all the three clustering methods cannot partition the pixels exactly.

In this paper, we propose a novel approach of generating stereoscopic images using defocus. Specifically, the DFD method is used for recovering the depth of RGB image by computing the blur of each pixel and some clustering algorithms are employed to classify the pixels into two images automatically by comparing the performance among K-means clustering, GMM and the proposed method to determine the optimum approach. Then a layered structure whose texture is mapped from images onto the planes are constructed for the purpose of generating stereoscopic images. With the help of the proposed approach, a layered stereoscopic model can be created from a single defocused image to provide stereoscopic feeling. The proposed approach has also been expanded to construct layered stereoscopic panorama, which shows its great application potentials.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

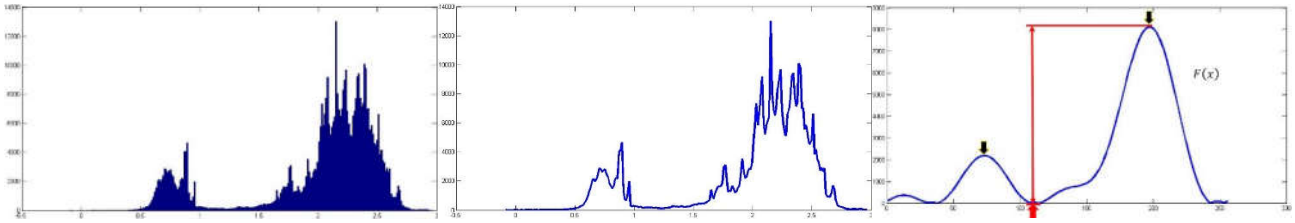

Figure 7. *Filter processing. From left to right: depth histogram, histogram curve and smooth curve after filtering. We find the depth threshold by searching the local minimal value based on the filtering result under the constraint of difference between the first peak value and the minimal value.*

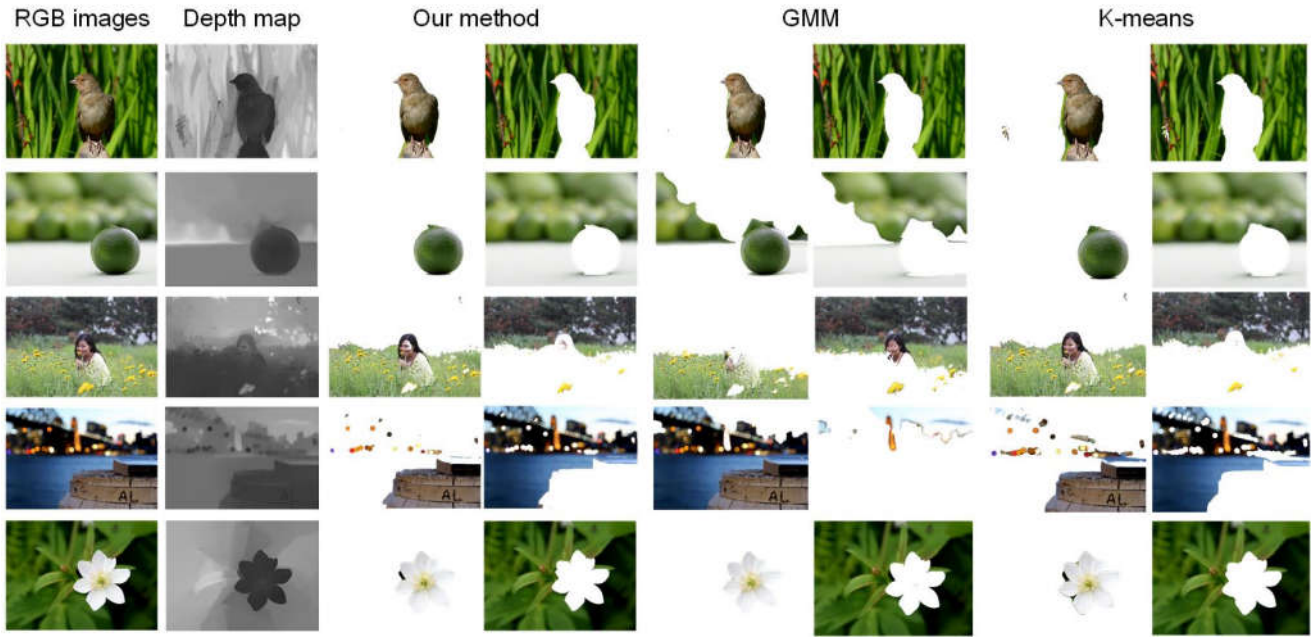

Figure 8. *Comparison of experiment result. From left to right: input image, depth map estimated by Zhuo's approach [16], pixel clustered by our method, by GMM, by K-means cluster. The layered results clustered by our method and by K-means are more accuracy.*

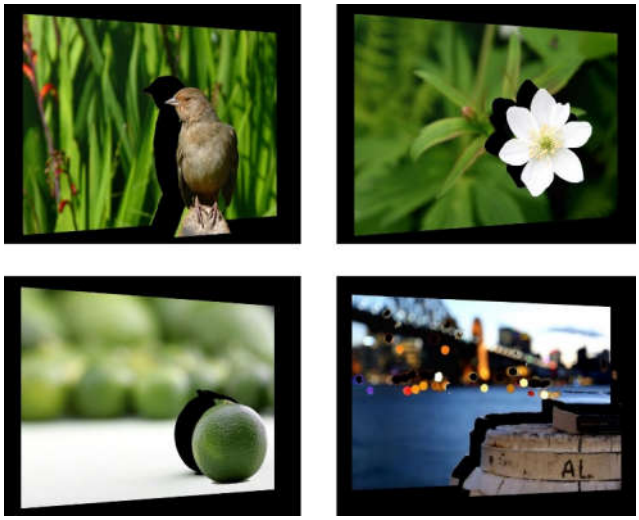

Figure 9. *Stereoscopic images generated by layered structure. The occlusion and blank can provide the stereoscopic feelings.*

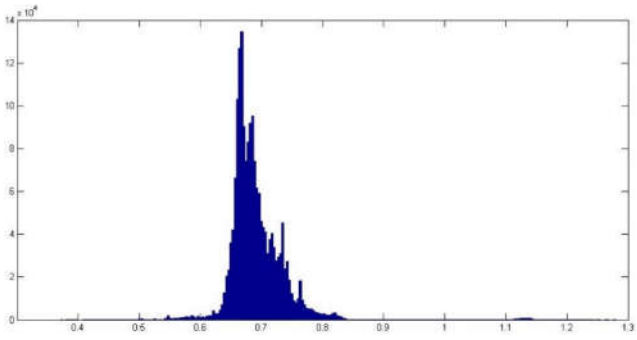

Figure 10. *Depth histogram of our panorama image. The depth of the pixels appear in an interval, leading to the incorrect cluster.*

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

The real layered model can give people the stereoscopic impression based on the relative location relationship between objects in the image. The proposed method converts a 2D image to a pop-up book like structure that is different from the conventional stereoscopic process. The key of our approach is the depth estimation algorithm. Thus, the accurate and stable depth estimation method will be the research emphasis in the future.

Acknowledgement

This work has been supported by the National Natural Science Foundation of China (Grant No. 61370134) and National Key Technology R&D Program (Grant No. 2012BAH64F01, 2012BAH64F02, 2012BAH64F03, 2012BAH64F04). The authors would like to thank Jie Hao for his suggestions and helps for the realization of the proposed algorithm.

References

[1] L. J. Angot, W.-J. Huang, and K.-C. Liu, "A 2D to 3D video and image conversion technique based on a bilateral filter," in IS&T/SPIE Electronic Imaging, 2010, pp. 75260D-75260D-10.

[2] M. Guttmann, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo extraction from video footage," in Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 136-142.

[3] J. Konrad, G. Brown, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee, "Automatic 2d-to-3d image conversion using 3d examples from the internet," in IS&T/SPIE Electronic Imaging, 2012, pp. 82880F-82880F-12.

[4] M. Liao, J. Gao, R. Yang, and M. Gong, "Video stereolization: Combining motion analysis with user interaction," Visualization and Computer Graphics, IEEE Transactions on, vol. 18, pp. 1079-1088, 2012.

[5] R. Phan, R. Rzeszutek, and D. Androutsos, "Semi-automatic 2D to 3D image conversion using scale-space random walks and a graph cuts based depth prior," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 865-868.

[6] S. T. Barnard and M. A. Fischler, "Computational stereo," ACM Computing Surveys (CSUR), vol. 14, pp. 553-572, 1982.

[7] U. R. Dhond and J. K. Aggarwal, "Structure from stereo-a review," IEEE transactions on systems, man, and cybernetics, vol. 19, pp. 1489-1510, 1989.

[8] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondence," in Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, 2000, pp. 557-564.

[9] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," International Journal of Computer Vision, vol. 9, pp. 137-154, 1992.

[10] N. Asada, H. Fujiwara, and T. Matsuyama, "Edge and depth from focus," International Journal of Computer Vision, vol. 26, pp. 153-163, 1998.

[11] S. K. Nayar and Y. Nakagawa, "Shape from focus," Pattern analysis and machine intelligence, IEEE Transactions on, vol. 16, pp. 824-831, 1994.

[12] P. Favaro and S. Soatto, 3-d shape estimation and image restoration: Exploiting defocus and motion-blur: Springer Science & Business Media, 2007.

[13] A. P. Pentland, "A new sense for depth of field," Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 523-531, 1987.

[14] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, pp. 406-417, 2005.

[15] M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," International Journal of Computer Vision, vol. 27, pp. 203-225, 1998.

[16] A. N. Rajagopalan and S. Chaudhuri, "An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 21, pp. 577-589, 1999.

[17] M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," International Journal of Computer Vision, vol. 13, pp. 271-294, 1994.

[18] P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, pp. 518-531, 2008.

[19] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in ACM Transactions on Graphics (TOG), 2007, p. 70.

[20] C.-W. Chen and Y.-Y. Chen, "Recovering depth from a single image using spectral energy of the defocused step edge gradient," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 1981-1984.

[21] S. Zhuo and T. Sim, "Defocus map estimation from a single image," Pattern Recognition, vol. 44, pp. 1852-1858, 2011.

[22] D. Hoiem, A. A. Efros, and M. Hebert, "Automatic photo pop-up," ACM Transactions on Graphics (TOG), vol. 24, pp. 577-584, 2005.

[23] E. Hecht, "Optics 4th edition," Optics, 4th Edition, Addison Wesley Longman Inc, 1998, vol. 1, 1998.

[24] J. Canny, "A computational approach to edge detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 679-698, 1986.

[25] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," ACM transactions on graphics (TOG), vol. 23, pp. 664-672, 2004.

[26] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, pp. 228-242, 2008.

[27] A. Antoniou, Digital signal processing: McGraw-Hill Toronto, Canada, 2006.

Author Biography

Tianteng Bi *received his B.S. degree in optical information science and technology from Taiyuan University of Technology, Taiyuan, Shanxi Province, China, in 2012, where he is currently working toward the Ph.D. degree in optical engineering, Beijing Institute of Technology. His primary research interest is depth estimation from a single image, computer vision and virtual reality.*

Yue Liu *received his Ph.D. degree in telecommunication and information system from Jilin University, Jilin Province, China in 2000.He is currently a Professor of optical engineering at the School of Optoelectronics, Beijing Institute of Technology, Beijing. His research interests include human*

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

computer interaction, virtual and augmented reality, accurate tracking of the pose of camera, 3D display system and camera calibration etc. He has published more than 100 technical papers. Dr. Liu is a member of council of China Society of Image and Graphics, a member of China Society of Optics, and he also serves on the editorial board of the Optical Technique.

Dongdong Wong *received his Ph.D. degree in optical engineering from Beijing Institute of Technology, Beijing, China in 2006. He is currently an associate professor of optical engineering at the School of Optoelectronics, Beijing Institute of Technology, Beijing. His research interests include virtual reality and augmented reality technology and application, human computer interaction, new media entertainment theme park and precise location tracking algorithm and the corresponding devices etc.*

Yongtian Wang *received the B.Sc. degree in precision instrumentation from Tianjin University, China, in 1982, and the Ph.D. degree in optics from the University of Reading, U.K., in 1986. He is currently a Yangtze River Scholar of the Chinese Ministry of Education, a professor and the director of the Center for Research on Optoelectronics and Information Technology in Beijing Institute of Technology. Dr. Wang is a Fellow of the Optical Society of America and the International Society for Optical Engineers. His research interests include optical design and CAD, optical instrumentation, image processing, virtual reality (VR) and augmented reality (AR) technologies and applications.*

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-436

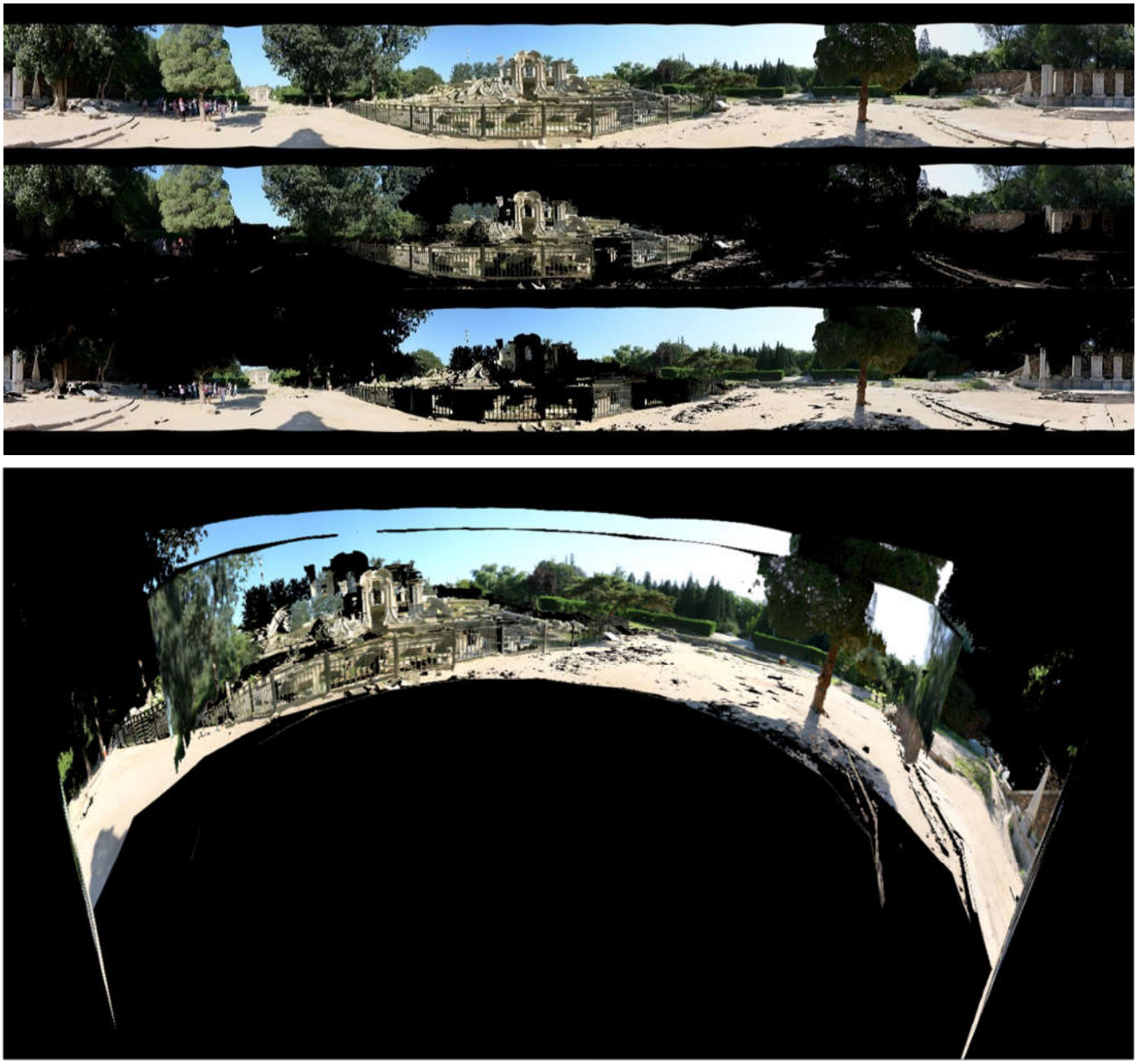

Figure 11. *Layered panoramas and stereoscopic panorama model. From top to bottom: panorama image, foreground, background and cylindrical stereoscopic panorama. The layered structure still can give people stereo impression in a certain extent in spite of some mistakes in the layered panorama images.*