SINGLE CAMERA-BASED FULL DEPTH MAP ESTIMATION USING COLOR SHIFTING PROPERTY OF A MULTIPLE COLOR-FILTER APERTURE

Seungwon Lee, Junghyun Lee*, Monson H. Hayes*, Aggelos K. Katsaggelos**, and Joonki Paik**

*Image Processing and Intelligent Systems Laboratory, Graduate School of Advanced Imaging Science, Multimedia, and Film, Chung-Ang University, Seoul, Korea
**Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA
superlsw@gmail.com, waajh86@gmail.com, mhh3@gatech.edu, aggk@eecs.northwestern.edu, and paikj@cau.ac.kr

ABSTRACT

A multiple color-filter aperture (MCA) camera can provide depth information as well as color and intensity in the single-camera framework, where the MCA generates misalignment between color channels depending on the distance of a region-of-interest. In this paper, we present a single camera-based estimation of the full depth map using the color shifting property of the MCA. For estimating the color shifting vectors (CSVs) among red, green, and blue color channels, edges are extracted at each color channel. At the edge, we estimate CSVs using normalized cross correlation combined with color shifting mask map. A full depth map is then generated by depth interpolation using the matting Laplacian method from sparsely estimated CSVs at an edge location. Experimental results show that the proposed method can not only estimate the full depth map but also correct the misaligned color image to generate photorealistic color images using a single camera equipped with MCA.

Index Terms— Depth estimation, computational camera, normalized cross correlation, depth interpolation, 3D image acquisition

1. INTRODUCTION

Various approaches to estimating three-dimensional (3D) depth information have been extensively studied for the past several decades because of its broad applications in areas such as robot vision, intelligent visual surveillance, 3D image acquisition, and intelligent driver assistant systems.

Most conventional depth estimation methods have relied on either multiple images such as stereo vision or on additional cues such as shading, focusing, and motion. Stereo matching is a depth estimation method using binocular disparity generated by a stereo camera. In spite of many advantages, it has a fundamental limitation that a pair of images of the same scene should be acquired by two cameras with both temporal and spatial synchronization [1].

As an alternative to binocular systems, monocular methods have also been studied. Depth from defocus is a single camera based depth estimation that measures the amount of defocus blur from a pair of images with different focus settings on the same scene. However, this approach is limited to still photography because a fixed camera view is required for taking multiple defocused images [2]. Zhuo [3] has used a single defocused image for depth estimation by considering the amount of blur measured by using the Gaussian gradient ratio at an edge as depth. However, it has the ambiguity problem between defocus and motion blurs.

Recently, various types of computational cameras have been developed to acquire additional modalities such as depth and light field that cannot be obtained with a conventional digital camera. A computational camera uses unconventional optics and software to produce new form of visual information. In combination with digital signal processing algorithms, these cameras have been used to solve a variety of imaging applications including refocusing, increased dynamic range, depth-guided editing, variable lighting, and reflectance reduction [4].

The multiple color-filter aperture (MCA) camera has been proposed for single camera-based multifocusing and depth estimation. The MCA is inserted between the lens and the imaging sensor to provide geometric information of an object from the color shifting property of the color filter array. Based on the MCA configuration, E. Lee has proposed a multi-focusing method using region-based depth estimation [5], and S. Lee has estimated the depth of an object by tracking the region-of-interest [6].

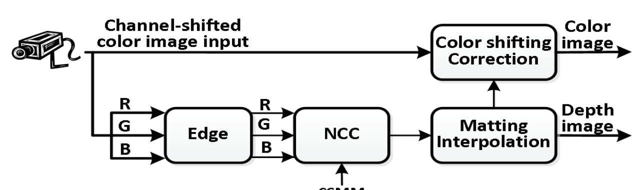


Fig. 1. The proposed MCA camera-based depth estimation system in the single-camera framework.

In this paper, we present the MCA camera-based estimation of the full depth map using matching among color channels and depth interpolation as shown in Fig. 1.

Although the MCA camera provides depth-dependent shifting or spatial disparity among color channels, conventional matching algorithms cannot easily estimate the spatial disparity because color channels have different intensity levels. To address this problem, we extract edge information within each color channel and estimate the color shift vectors (CSVs) between the red and the green (R-G) channels and between the red and the blue (R-B) channels within the edge regions. To estimate the color shift vectors, a normalized cross correlation (NCC) is used, which is a fast block matching method [7]. In addition, an a priori generated color shifting mask map (CSMM) is used to incorporate a feasibility constraint. The full depth map is then generated by propagating the sparsely estimated depth that is obtained from the CSVs within the edge regions to the entire image. The matting Laplacian is used for depth interpolation from the sparse depth map [3][8].

Experimental results show that the proposed method can not only estimate the full depth map but also correct color misalignment using the MCA-based single camera.

2. PRINCIPLE OF THE MCA: A REVIEW

The principle of the MCA [5,6] is briefly reviewed in this section. The aperture of an optical system is the opening that adjusts the amount of light entering the camera. The center of an aperture is generally aligned with the optical axis of the lens, and the convergence pattern on the image plane will form either a point or a circular region depending on the distance of the object from the plane of focus. When the center of the aperture is not aligned with the optical axis, the convergence point will be shifted away from the optical axis by an amount that is a function of the distance of the object from the plane of focus of the camera. The MCA camera has red (R), green (G), and blue (B) filters that allow light to enter the camera through an aperture that is off the at different locations in the optical plane. The main advantage of the MCA camera is that it provides additional depth information that can be estimated from the direction and amount of color deviation from the optical axis. optical axis of the camera. Each filter thus focuses the light.

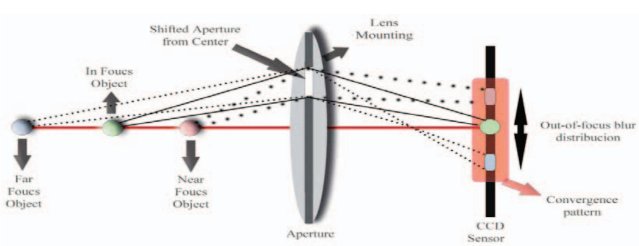


Fig. 2. The convergence patterns of the CSM according to objects at different distances.

3. COLOR SHIFTING PROPERTY BASED FULL DEPTH MAP ESTIMATION

In a color image acquired by the MCA camera, CSVs between R-G and R-B channels are estimated within the edge regions by NCC combined with CSMM. A full depth map is then generated using matting Laplacian-based depth interpolation of the sparsely estimated depth.

3.1. Sparse depth map estimation

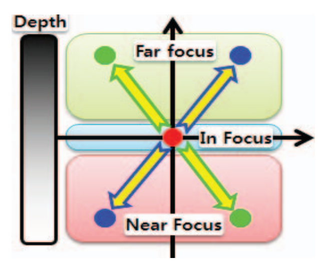
For the initial sparse depth map estimation, we use normalized cross-correlation for fast block matching [7]. Specifically, with $f_1(x,y)$ a block in the red color channel and $f_2(x,y)$ a block in the green or blue color channel, then the NCC is given by

$$C_N(u,v) = \frac{\sum_{x,y}\{f_1(x,y)-\overline{f_1}\}\{f_2(x-u,y-v)-\overline{f_2}\}}{\sqrt{\sum_{x,y}\{f_1(x,y)-\overline{f_1}\}^2}\sqrt{\sum_{x,y}\{f_2(x-u,y-v)-\overline{f_2}\}^2}} \quad (1)$$

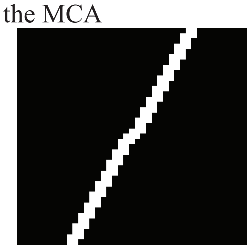
which can be efficiently evaluated using the FFT.

The disparity measured by edge-based NCC is vulnerable to error because of erroneously detected edges and different intensity level between the color channels. To obtain a more accurate disparity estimate, we apply an *a* priori constraint on the feasible pattern of CSVs, which enforces the color shifting property of the MCA in the form of a mask called CSMM.

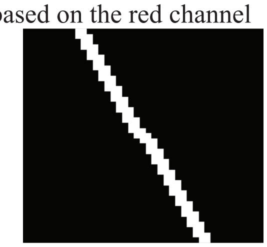
(a) Geometric configuration of the MCA



(b) Color shifting property based on the red channel



(c) CSMM for R-G channels



(d) CSMM for R-B channels

Fig. 3. Color shifting property of the MCA configuration and the Color Shifting Mask Map.

Figure 3 shows the color shifting property of the MCA, where three apertures are located at vertices of the equilateral triangle as shown in Fig. 3(a). Figure 3(b) shows the movement in the green and blue channels relative to the read channel when an object moves away from a position within the plane of focus. Based on the color shifting property, CSMM is generated as shown in Figs. 3(c) and 3(d), and the CSV is estimated by maximizing the NCC subject to the CSMM constraint as follows:

$$CSV(x,y) = \arg\max_{u,v} C_N(u,v), \text{ subject to } CSMM(u,v) = 1 \quad (2)$$

In addition to the inherent color shifting property, we can select the CSV containing higher matching rate out of the two from R-G and R-B channels. The depth of a point at (x, y) is finally determined as

$$D(x,y) = -sign(v) \times \sqrt{u^2 + v^2} \quad (3)$$

where (u, v) represents the CSV estimated at (x, y).

3.2. Generation of the full depth map

Given the sparse depth map within the edge regions, it is then necessary to fill in the rest of the image by depth interpolation for generating the full depth map. The proposed depth interpolation is inspired by the matting Laplacian, which is used for recovering the sparse depth map in single defocused image by Zhuo [3][8].

Let $\hat{d}$ and d be the sparse and the full depth maps, respectively. The depth interpolation is performed by minimizing the following energy function

$$E(d) = d^T L d + \lambda (d - \hat{d})^T A (d - \hat{d}) \quad (4)$$

where L denotes the matting Laplacian matrix, and A represents a diagonal matrix whose element A_{ii} is 1 if the i-th pixel is an edge, and 0 otherwise. The constant λ is a parameter that controls the effect between the fidelity to the sparse depth map and smoothness of interpolation. L is defined as

$$L(i,j) = \sum_{k|(i,j)\in w_k} \left(\delta_{ij} - \frac{1}{|w_k|} \left(1 + (I_i - \mu_k)^T \left(\Sigma_k - \frac{\varepsilon}{|w_k| U_3} \right)^{-1} (I_j - \mu_k) \right) \right) \quad (5)$$

where δ_{ij} and U represent the Kronecker delta function and a 3x3 identity matrix, respectively. In addition, μ_k and Σ_k are the mean and covariance matrix of the colors within window w_k, I_i and I_j are the colors of the input image I at pixels i and j respectively, ε denotes a regularization parameter, and $|w_k|$ is the size of window w_k. The closed-form solution for minimizing (4) is given as

$$d = (L + \lambda A)^{-1} \lambda A \hat{d} \quad (6)$$

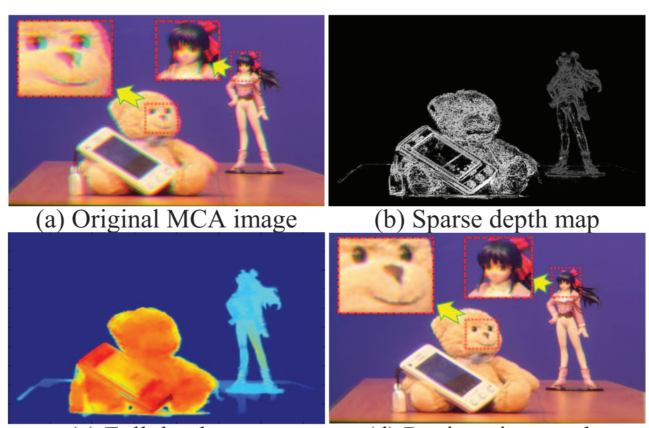
For further compensating erroneously interpolated depth, the image is segmented using mean-shift segmentation [9]. The interpolation within regions where edge pixels are less than 1% is not performed .

In order to align the color-shifted channels in the MCA camera, color channel registration is implemented to produce a color-aligned image. We can calculate the full CSVs using the trigonometric functions of the sine and the cosine and the angle of the color-filter aperture from the full depth map. The green and blue channels are shifted at each pixel by inverted CSVs in the entire image and the bicubic interpolation is implemented at the holes generated by pixel shifting.

4. EXPERIMENTAL RESULTS

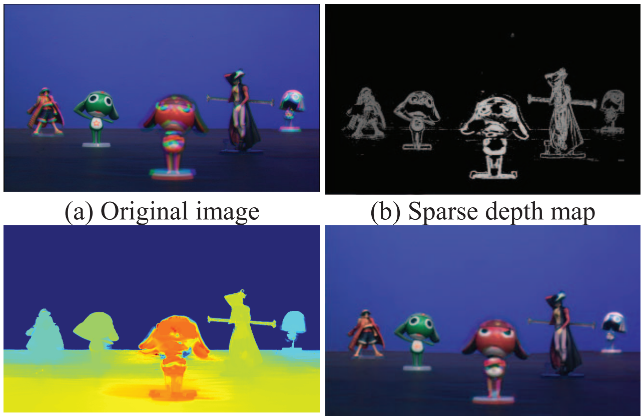
For the experiment, we captured test images by the prototype MCA camera. We used Laligant's edge detector that has noise cancellation function with very low computational complexity. Depth was estimated at edge location having values more than 0.04 in the normalized edge range from 0 to 1 [10]. λ takes a value in the range between 0.01 and 0.1, and the block size of the FFT for NCC is 32×32.

Figs. 4 and 5 show the results of the full depth map generation and the corresponding color channel registration. More specifically, results of the sparse depth estimation at edge location are shown in Figs. 4(b) and 5(b). The interpolated full depth maps are shown in Figs. 4(c) and 5(c). Color misalignment of objects was successfully removed for obtaining photo-realistic images as shown in Figs. 4(d) and 5(d). Fig. 6 shows the result of depth estimation with more complex environment. The estimated depth of an object is affected by adjacent objects as shown in Fig 6(c). As a result, color channel registration fails in some occluded regions because depth continuity is not considered in the color channel matching step, which can be solved by applying the full depth map information.



(a) Original MCA image (b) Sparse depth map
(c) Full depth map (d) Registration results

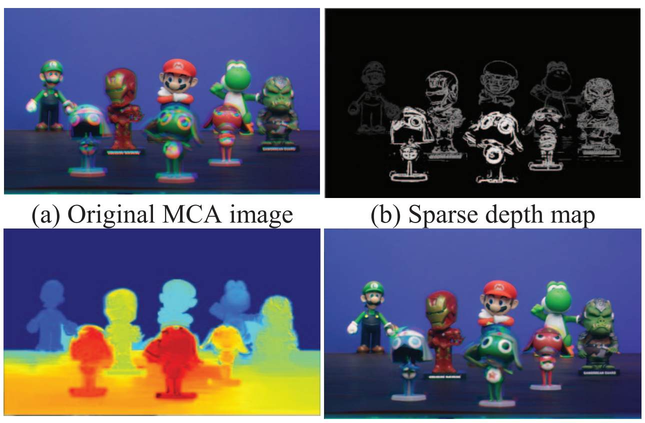
Fig. 4. Experimental results of the proposed method using the MCA camera.



(a) Original image (b) Sparse depth map

(c) Full depth map (d) Registration results

Fig. 5. Experimental results of the proposed method using the MCA camera.



(a) Original MCA image (b) Sparse depth map

(c) Full depth map (d) Registration results

Fig. 6. Experimental results of the proposed method using the MCA camera with more objects than in Figure 5.

5. CONCLUSION

An MCA camera can provide depth information as well as color and intensity in the single-camera framework. In this paper, we proposed a single camera-based full depth map estimation using efficient color channel matching based on color shifting property of the MCA. For block matching between RGB color channels, edge was extracted at each color channel. The CSVs between R-G and R-B color channels are then estimated using FFT-based NCC combined with CSMM. The full depth map was generated by depth interpolation using the matting Laplacian method from sparsely estimated CSVs at edge.

Experimental results show that the proposed method can not only estimate the full depth map but also correct color misalignment to make photo-realistic color images using the MCA camera.

7. ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059) and by this research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the HNRC(Home Network Research Center) –ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2011-C1090-1111-0010)

6. REFERENCES

[1] U. Dhond and J. Aggarwal, "Structure from stereo-a review," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1498-1510, November 1989.

[2] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 406-417, March 2005.

[3] S. Zhuo and T. Sim, "Recovering depth from a single defocused image," *International Journal of Pattern Recognition*, submitted, 2011.

[4] S. Nayar, "Computational cameras: redefining the image," *IEEE Computer Magazine, Special Issue on Computational Photography*, pp. 30-38, August/June 2006.

[5] E. Lee, W. Kang, S. Kim, and J. Paik, "Color shift model-based on a multiple color-filter aperture camera," *IEEE Trans on Consumer Electronics*, vol. 56, no. 2, pp. 217-323, May 2010.

[6] S. Lee, J. Lee, and J. Paik, "Simultaneous object tracking and depth estimation using color shifting property of a multiple color-filter aperture camera," *Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1401-1404, 2011.

[7] D. Padfield, "Masked FFT Registration," *Int. Conf. Computer Vision and Pattern Recognition*, pp. 2918-2925, 2010.

[8] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228-242, February 2008.

[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 603-619, May 2002.

[10] O. Laligant and F. Truchetet, "A nonlinear derivative scheme applied to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 242-257, February 2010.