Depth map generation using a single image sensor with phase masks

Jinbeum Jang, Sangwoo Park, Jieun Jo, and Joonki Paik*

Image Processing and Intelligent System Laboratory Graduate School of Advanced Imaging Science and Film, Chung-Ang University, Seoul 156-756, South Korea

*paikj@cau.ac.kr

Abstract: Conventional stereo matching systems generate a depth map using two or more digital imaging sensors. It is difficult to use the small camera system because of their high costs and bulky sizes. In order to solve this problem, this paper presents a stereo matching system using a single image sensor with phase masks for the phase difference auto-focusing. A novel pattern of phase mask array is proposed to simultaneously acquire two pairs of stereo images. Furthermore, a noise-invariant depth map is generated from the raw format sensor output. The proposed method consists of four steps to compute the depth map: (i) acquisition of stereo images using the proposed mask array, (ii) variational segmentation using merging criteria to simplify the input image, (iii) disparity map generation using the hierarchical block matching for disparity measurement, and (iv) image matting to fill holes to generate the dense depth map. The proposed system can be used in small digital cameras without additional lenses or sensors.

OCIS codes: (110.0110) Imaging systems; (100.6890) Three-dimensional image processing;(130.6010) Sensors; (110.6880) Three-dimensional image acquisition.

References and links

1. N. Yang, J. Lee, and R. Park, "Depth map generation from a single image using local depth hypothesis," in *Proceedings of IEEE International Conference on Consumer Electronics,* (IEEE, 2012), pp. 311–312.
2. A. Farooq and C. Won, "A survey of human action recognition approaches that use an RGB-D sensor," IEIE Trans. Smart Processing and Computing **4**, 281–290 (2015).
3. S. Kang, A. Roh, C. Eem, and H. Hong, "Using real-time stereo matching for human gesture detection and tracking," TechArt: Journal of Arts and Imaging Science **1**, 60–66 (2014).
4. H. Kim, J. Kang, and B. Song, "Depth-adaptive sharpness adjustments for stereoscopic perception improvement and hardware implementation," IEIE Trans. Smart Processing and Computing **3**, 110–117 (2014).
5. K. Denker and G. Umlauf, "Accurate real-time multi-camera stereo matching on the GPU for 3D reconstruction," Journal of WSCG **19**, 9–16 (2011).
6. S. Lee, M. H. Hayes, and J. Paik, "Distance estimation using a single computational camera with dual off-axis color filtered apertures," Opt. Express **21**, 23116–23129 (2013).
7. S. Kim, E. Lee, M. H. Hayes, and J. Paik, "Multifocusing and depth estimation using a color shift model-based computational camera," IEEE Trans. Image Processing **21**, 4152–4166 (2012).
8. S. Zhang, C. Wang, and S. C. Chan, "A new high resolution depth map estimation system using stereo vision and depth sensing device," in *Proceedings of IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA),* (IEEE, 2013), pp. 49–53.
9. Y. Kang and Y. Ho, "High-quality multi-view depth generation using multiple color and depth cameras," in *Proceedings of IEEE 9th International Conference on Multimedia and Expo (ICME),* (IEEE, 2010), pp. 1405–1410.
10. B. Yoon, K. Choi, M. Ra, and W. Kim, "Real-time full-view 3D human reconstruction using multiple RGB-D cameras," IEIE Trans. Smart Processing and Computing **4**, 224–230 (2015).

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

11. N. Yang, J. Lee, and R. Park, "Depth map generation using local depth hypothesis for 2D-to-3D conversion," International Journal of Computer Graphics & Animation (IJCGA) **3**, 1–15 (2013).
12. S. Battiato, S. Curti, M. La Cascia, M. Tortora, and E. Scordato, "Depth map generation by image classification," in *Electronic Imaging 2004,* (International Society for Optics and Photonics, 2004), pp. 95–104.
13. F. Yu, J. Liu, Y. Ren, J. Sun, Y. Gao, and W. Liu, "Depth generation method for 2D to 3D conversion," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON),* (IEEE, 2011), pp. 1–4.
14. S. Zhuo and T. Sim, "Defocus map estimation from a single image," Pattern Recognition **44**, 1852–1858 (2011).
15. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," Pattern Analysis and Machine Intelligence, IEEE Transactions on **33**, 2341–2353 (2011).
16. L. Spinoulas, A. Katsaggelos, J. Jang, Y. Yoo, J. Im, and J. Paik, "Defocus-invariant image registration for phase-difference detection auto focusing," in *Proceedings of IEEE International Symposium on Consumer Electronics,* (IEEE, 2014), pp. 83–84.
17. J. Jeon, J. Lee, and J. Paik, "Robust focus measure for unsupervised auto-focusing based on optimum discrete cosine transform coefficients," IEEE Trans. Consumer Electronics **57**, 1–5 (2011).
18. Y. Yoo, J. Jang, J. Shin, and J. Paik, "Optimal PSF selection using second-order frequency analysis for digital autofocusing," TechArt: Journal of Arts and Imaging Science **2**, 81–86 (2015).
19. D. Kim, J. Shin, and J. Paik, "Real-time digital auto-focusing using prior PSF estimation," TechArt: Journal of Arts and Imaging Science **1**, 39–41 (2014).
20. J. Jang, Y. Yoo, J. Kim, and J. Paik, "Sensor-based auto-focusing system using multi-scale feature extraction and phase correlation matching," Sensors **16**, 5747–5762 (2015).
21. P. Śliwiński and P. Wachel, "A simple model for on-sensor phase-detection autofocusing algorithm," Journal of Computer and Communication **1**, 11–17 (2013).
22. R. Butler, "Exclusive: Fujifilm's phase detection system explained," *http://www.dpreview.com/articles/2151234617/fujifilmpd*.
23. R. Fontaine, "Innovative technology elements for large and small pixel CIS devices," in *Proceedings on International Image Sensor Workshop,* (IISS, 2013), pp. 1–4.
24. J. Ahn, K. Lee, Y. Kim, H. Jeong, B. Kim, H. Kim, J. Park, T. Jung, W. Park, T. Lee, E. Park, S. Choi, G. Choi, H. Park, Y. Choi, S. Lee, Y. Kim, Y. J. Jung, D. Park, S. Nah, Y. Oh, M. Kim, Y. Lee, Y. Chung, I. Hisanori, J. Im, D. K. Lee, B. Yim, G. Lee, H. Kown, S. Choi, J. Lee, D. Jang, Y. Kim, T. Kim, G. Hiroshige, C. Choi, D. Lee, and G. Han, "7.1 A 1/4-inch 8Mpixel CMOS image sensor with 3D backside-illuminated 1.12μm pixel with front-side deep-trench isolation and vertical transfer gate," in *Proceedings on IEEE International Solid-State Circuits Conference Digest of Technical Papers,* (IEEE, 2014), pp. 124–125.
25. J. Jang and J. Paik, "Dense depth map generation using a single camera with hybrid auto-focusing," in *Proceedings of IEEE International Conference on Consumer Electronics Berlin,* (IEEE, 2015), pp. 277–278.
26. D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," Communications on Pure and Applied Mathematics **42**, 577–685 (1989).
27. G. Koepfler, C. Lopez, and J.-M. Morel, "A multiscale algorithm for image segmentation by variational method," SIAM J. Numer. Anal. **31**, 282–299 (1994).
28. A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," IEEE Trans. Pattern Analysis and Machine Intelligence **30**, 228–242 (2008).
29. S. Mukherjee and R. M. R. Guddeti, "A hybrid algorithm for disparity calculation from sparse disparity estimates based on stereo vision," in *Proceedings of IEEE International Conference on Signal Processing and Communications (SPCOM),* (IEEE, 2014), pp. 1–6.
30. E. Z. Psarakis and G. D. Evangelidis, "An enhanced correlation-based method for stereo correspondence with subpixel accuracy," in *Proceedings of IEEE International Conference on Computer Vision,* (IEEE, 2005), pp. 907–912.
31. C. Çıgla, "Recursive edge-aware filters for stereo matching," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops,* (IEEE, 2015), pp. 27–34.

1. Introduction

Three dimensional (3D) information is acquired by measuring the distance between a lens and an object selected as a region of interest (ROI). Especially, a depth map is generated by analyzing all ROIs in an image based on recognizable features. For this reason, stereo matching technique is widely used in 3D applications [1–3].

There are three main approaches to obtain the depth map. The first approach performs stereo matching using two or more cameras [4, 5]. Because each image has a different viewing angle, the depth map is generated by estimating the disparity between the images. This approach provides high accuracy at the cost of complexity. However, it increases the cost and volume of

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

the system. The second approach estimates the disparity using a multiple color-filtered aperture (MCA) in a lens [6, 7]. The MCA-based approach is comparable with the proposed method since it estimates the depth map using a single image. The acquired image consists of three color channels each of which has differently located ROIs. Although the MCA-based approach has lower cost than the conventional stereo matching approach, the input image has color distortion and lower brightness. The last approach uses a infrared sensor-based camera to estimate the distance using the time-of-flight (ToF) method [8–10]. It computes the disparities quickly and accurately, but the range of estimated distance is restricted.

Several methods used a gradient map from a vanishing point that is generated by projecting light sources from the scene to the imaging sensor [11–13]. Zhuo *et al.* computed a defocus map by estimating point spread functions (PSFs) from the near- or far-focused image [14]. He *et al.* proposed the haze removal method and the depth map application using the dark channel prior [15]. These methods can accurately generate the depth map using a single image.

This paper presents a novel depth map generation system using a dual pixel-type imaging sensor. This sensor acquires a set of stereo images using phase photo-diodes with different black masks. Next, variational segmentation-based stereo matching is performed using the acquired stereo images. After hierarchical block matching is performed to measure sub-pixel disparities, this system generates a dense depth map using the segmented image and the disparity edge map.

This paper is organized as follows. Theoretical background is introduced in section 2, and the proposed depth map generation method is presented in section 3. After summarizing experimental results in section 4, section 5 concludes the paper.

2. Image acquisition model of hybrid auto-focusing

Phase detection auto-focusing (PDAF) is the technique that automatically finds the best focusing position of a lens using a phase-difference information from a specially designed image sensor. This approach falls into the passive auto-focusing category and has an elaborate optical path generated by the relationship between the axis of light source and the separated lights [16]. However, PDAF needs the space to equip additional devices such as line sensors and half mirrors in the camera. Therefore, it is not appropriate in small, portable imaging systems despite the fast, accurate performance. Digital AF can solve this problem using contrast detection [17] or PSF [18, 19], but it has the high complexity.

A hybrid AF system is a variant of the existing AF system to solve this problem, and computes phase differences using a dual pixel-type complementary metal oxide semiconductor (CMOS) sensor [20, 21]. This sensor significantly reduces the cost of cameras since it can replace half mirrors, separating lenses, and line sensors. There are some types of the sensors for the hybrid AF. One is equipped with black masks on the color filters of some pixels as shown in Fig. 1(a) [22]. Phase masks with two directions are installed to interrupt light rays with a specific viewing angle. Another type uses an imaging sensor with special pixels containing two sub-lenses and photo-diodes as shown in Fig. 1(b) [23]. This sensor can simply acquire two sub-images by absorbing the left- and right-sided lights. The third type has isolation barriers between cells as shown in Fig. 1(c) [24]. By placing the barrier between two adjacent pixels, a pixel absorbs the light sources passed through a micro lens without the interference of neighboring lenses in each photo-diode. These sensors commonly generate two images with different viewing angles, and each pixel has different disparity.

The proposed system can use any types of an image sensor with phase masks. As shown in Fig. 2(a), each photo-diode has the phase mask that is installed in the right side of even columns, and in the left side of odd columns. Because a pair of phase pixels generates different disparities from objects with different distances as shown in Fig. 2(b), the proper amount phase difference

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

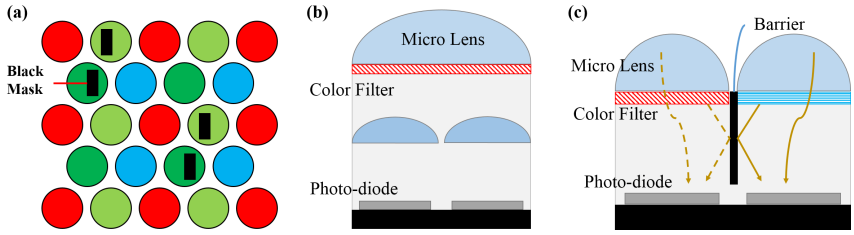

Fig. 1. Various imaging sensors. (a) an array of pixels with the black masks; (b) dual-pixel CMOS sensor; and (c) dual-pixel sensor with a barrier.

can give a clue how to move the lens to the right focusing position. Furthermore, sub-pixel phase differences are generated because the phase pixels have different viewing angles from each object with a different distance. When the left- and right-phase images are acquired using photo-diodes with right- and left-sided masks, the image acquisition model for hybrid AF is defined as [25]

$$g(x,y) = g^L(x,y) + g^R(x,y), \tag{1}$$

where

$$\begin{aligned} g^L(x,y) &= f(x,y) * h^L(x,y) + \eta(x,y) \\ g^R(x,y) &= f(x,y) * h^R(x,y) + \eta(x,y) \end{aligned}, \tag{2}$$

and $g(x,y)$ represents the sampled input image, $g^L(x,y)$ and $g^R(x,y)$ respectively right and left phase images, $h^L(x,y)$ and $h^R(x,y)$ respectively PSFs of $g^L(x,y)$ and $g^R(x,y)$, $f(x,y)$ the input scene in the real world, and $\eta(x,y)$ the additive noise in the image formulation process. In other words, $g(x,y)$ acquired from all phase photo-diodes is split into $g^L(x,y)$ and $g^R(x,y)$, and a set of phase images has phase differences because of different degraded models. Each photo-diode receives the different phase signal in the sensor. Therefore, hybrid AF system obtains the focused position of the lens using a single sensor accurately and quickly.

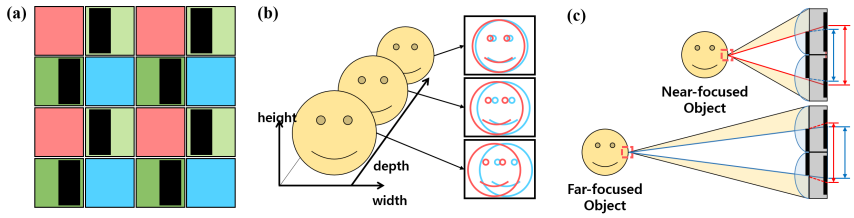

Fig. 2. The sensor-based PDAF system. (a) the sensor array with the black masks, (b) the phase difference of each object that has a different distance from the camera, and (c) a process of sub-pixel disparity generation.

3. Dense depth map generation using phase pixels

The proposed system generates a dense depth map using the dual pixel-type CMOS sensor with black masks. After sensing the raw data, an appropriate pre-processing steps, such as demosaicing and denoising, are needed to reduce the sensing noise. A sophisticated, accurate motion estimation is also needed to measure the disparities up to a sub-pixel precision.

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

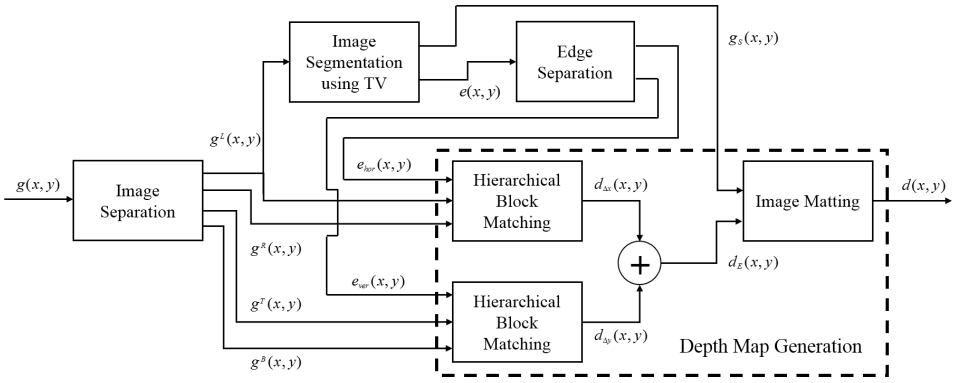

Fig. 3. Block diagram of the proposed dense depth map generation system.

To meet these requirements, the proposed method performs four steps as shown in Fig. 3: i) image separation from $g(x,y)$ using modified imaging sensor array, ii) image segmentation using an improved variational optimization method, iii) disparity map generation using the hierarchical block matching, and iv) dense depth map generation using image matting to fill holes. Elaborated description of each step is given in the following subsection.

3.1. Separation of stereo images using a single sensor

The proposed system acquires stereo images that are separated from an input image as the first step. Phase difference is equivalent to the disparity between the left image $g^L(x,y)$ and the right image $g^R(x,y)$ because objects have different phase differences as shown in Fig. 2(b). For this reason, phase images can be used as a pair of stereo images [25]. Also, the distance between the left- and right-side phase pixels is defined as the constant in the manufacturing process of the sensor. It means that rectification is very simple in the proposed system because vertical disparity is a constant in all pixels. As shown in Fig. 2(a), a left phase pixel is one pixel apart from the right pixel in the vertical and horizontal directions. When $g^L(x,y)$ and $g^R(x,y)$ are acquired from the phase pixels, the images have the disparities as shown in Fig. 4(a). Because the vertical disparity is constant, the measurement error of the disparities arises in the horizontal edge as shown in Fig. 4(b). Although the vertical disparity must be computed to obtain the accurate depth map in the horizontal edge, existing CMOS sensors cannot provide the vertical disparity data.

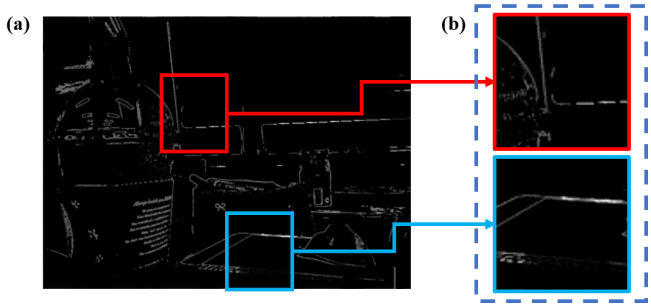

Fig. 4. Disparity map generation using the dual pixel-type imaging sensor as shown in Fig. 2(a). (a) the disparity map and (b) mis-measured horizontal disparities.

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

The proposed sensor has one of four directional black masks in each pixel of the 2×2 pixel array to measure both horizontal and vertical disparities as shown in Fig. 5. From (1) and (2), the image acquisition model is modified as

$$g(x,y) = g^L(x,y) + g^R(x,y) + g^T(x,y) + g^B(x,y), \tag{3}$$

where

$$\begin{aligned} g^T(x,y) &= f(x,y) * h^T(x,y) + \eta(x,y) \\ g^B(x,y) &= f(x,y) * h^B(x,y) + \eta(x,y) \end{aligned}, \tag{4}$$

and $g^T(x,y)$ and $g^B(x,y)$ respectively represent top and bottom phase images, and $h^T(x,y)$ and $h^B(x,y)$ respectively the corresponding blur functions of $g^T(x,y)$ and $g^R(x,y)$. The proposed system generates four stereo images from a set of 2×2 phase pixels. Therefore, it can measure the vertical disparity as well as the horizontal one.

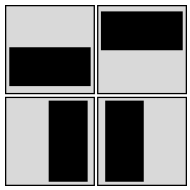

Fig. 5. The proposed sensor array to generate the dense depth map.

3.2. Segmentation using the improved variational merging criteria

Given a set of four stereo images, the left phase image $g^L(x,y)$ is simplified by image segmentation under assumption that an object has the same disparity in a segment. The segmented image provides boundaries of a depth map.

The proposed segmentation method is based on the variational optimization that was originally proposed by Mumford *et al.* using a piecewise constant smoothness [26]. The segmentation process is performed by minimizing the energy functional

$$E(g_S(\mathbf{x}), L) = \min\left[\int_i \int_\Omega \{g_{S_i}(\mathbf{x}) - g^L(\mathrm{x})\}^2 \mathrm{d}\mathbf{x} + \nu\|B\|\right], \tag{5}$$

where

$$g_{S_i}(\mathbf{x}) = \frac{\int_{\Omega_i} g^L(\mathbf{x})\mathrm{d}\mathbf{x}}{\int_{\Omega_i} \mathrm{d}\mathbf{x}}, \tag{6}$$

and $\mathbf{x} = (x,y)$, $g_{S_i}(\mathbf{x})$ the i-th segment of the segmented image $g_S(\mathbf{x})$, B the segmentation boundary, Ω_i the i-th segmented region that satisfies $\bigcup_{i=1}^N \Omega_i = \Omega \subset R^2$ and $\Omega_i \cap \Omega_j = \emptyset$, for $i \neq j$, where N is the number of segments, ν a regularization factor of the segmentation, and $\|.\|$ the length of a curve. An optimal solution is obtained by assuming that all $g_{S_i}(\mathbf{x})$ has a constant smoothness. Furthermore, an appropriate value of ν suppresses a noisy segmentation. However, the original version of the optimization-based segmentation has too high computational complexity to be implemented in a real-time application.

For a simple implementation, the proposed segmentation method uses the variational merging criterion proposed by Koepfler *et al.* [27]. This method defines the merging criterion using

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

the expansion property of boundaries. If $g_{S_i}(\mathbf{x})$ expands, the energy of the current boundary B decreases as

$$\begin{aligned} 0 &\leq E(g_S(\mathbf{x}),B') - E(g_S(\mathbf{x}),B) \\ &\leq \min(\|A_i\|,\|A_j\|) \times \left[\sup\{g^L(\mathbf{x})\} - \inf\{g^L(\mathbf{x})\}\right]^2 - \nu\|\partial(A_i,A_j)\|, \end{aligned} \tag{7}$$

where B' represents the previous boundary, A_i and A_j respectively i-th and j-th area of the previous segmented result, and $\sup(.)$ and $\inf(.)$ respectively the supremum and infimum. From (7), the merging criterion is defined as

$$E(g_S(\mathbf{x}),B') - E(g_S(\mathbf{x}),B) = \frac{\|A_i\|\|A_j\|}{\|A_i\| + \|A_j\|}\left|\sum_{A_i} g_S(\mathbf{x}) - \sum_{A_j} g_S(\mathbf{x})\right| - \nu\|\partial(A_i,A_j)\|. \tag{8}$$

This implies that the Koepfler's segmentation method is easily implemented and generates the optimal $g_S(x,y)$ without noise as shown in Fig. 6(c). But this method is very sensitive to the brightness of $g_S(x,y)$, and it merges week edges despite the boundary between two regions.

In order to solve these problems, an additional brightness-invariant criterion is used in the proposed method. Before initializing segmentation regions, the edge image $g_E^L(x,y)$ is generated from $g^L(x,y)$ using the Laplacian of Gaussian filtering to clearly segment objects with a proper threshold as shown in Fig. 6(b). Next, two regions are merged under the condition of (8). If the first merging criterion does not satisfy, the merging of the regions is performed when $g_E^L(x,y) \neq 0$. The proposed method repeats the merging process to get the simplified image $g_S(x,y)$ as shown in Fig. 6(d), and the final edge image $e(x,y)$ is used to compute disparities quickly.

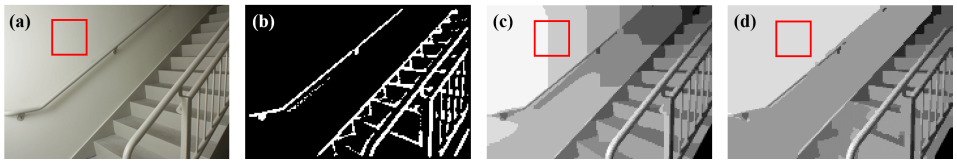

Fig. 6. Segmentation Result of the proposed segmentation method. (a) an input image, (b) the edge image of (a), (c) segmentation result without (b), and (d) segmentation result with (b).

3.3. Depth map generation using motion estimation and matting

Given the segmentation result, the proposed system generates a depth map using fine motion estimation and image matting. This system needs a floating-point disparity estimation from a set of stereo images because of the masks shown in Fig. 2. Jang *et al.* presented a hierarchical phase correlation method to detect floating-point phase difference in [20]. But the phase correlation matching is not efficient because of multiple Fourier transforms and the corresponding search of the peak point. Moreover, vertical and horizontal disparity maps must be generated from $g^L(x,y)$ and $g^R(x,y)$, and $g^T(x,y)$ and $g^B(x,y)$ respectively.

To obtain vertical and horizontal disparities, $e(x,y)$ is split into $e_{ver}(x,y)$ and $e_{hor}(x,y)$. Next, the proposed system measures the disparity using the multiple-scale, hierarchical block matching method based on the sum of absolute difference (SAD) proposed in [20] to reduce the computation. Block matching using SAD evaluation performs just simply operations by pixel unit without transformation and convolution, and also considers tiny brightness change. If $e_{ver}(x,y) > 0$ and $e_{hor}(x,y) > 0$, each motion value is defined as

$$\Delta x = \arg\min_{x'}\left[\sum_x\sum_y \left|g_{roi}^L(x+x',y) - g_{roi}^R(x,y)\right|\right], \tag{9}$$

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

and

$$\Delta y = \arg\min_{y'} \left[\sum_x \sum_y \left| g^T_{roi}(x, y+y') - g^B_{roi}(x,y) \right| \right], \tag{10}$$

where Δx and Δy represent disparities in the x- and y-coordinates, respectively, and $g^L_{roi}(x,y)$, $g^R_{roi}(x,y)$, $g^T_{roi}(x,y)$, and $g^B_{roi}(x,y)$ respectively the ROI images of $g^L(x,y)$, $g^R(x,y)$, $g^T(x,y)$, and $g^B(x,y)$. The proposed system crops four stereo images to the size of 64×64, and then linear interpolation is performed to compute elaborate disparities from $g^L_{roi}(x,y)$ and $g^T_{roi}(x,y)$. The disparity range of [-2, 2] is used, and the proposed method repeats interpolation until Δx and Δy of three decimal points are obtained. Consequently, the proposed system obtains final disparity map as

$$d_E(x,y) = d_{\Delta x}(x,y) + d_{\Delta y}(x,y), \tag{11}$$

where $d_E(x,y)$ denotes the final disparity map, and $d_{\Delta x}(x,y)$, $d_{\Delta y}(x,y)$ each the vertical and horizontal disparity map.

Finally, dense depth map is generated using image matting from $g_S(x,y)$ and $d_E(x,y)$. To fill holes of $d_E(x,y)$ from $g_S(x,y)$, the proposed system uses the map interpolation method by Zhuo [14]. This method generates the depth map using matting Laplacian matrix from the reference image and sparse edge image [28]. In the proposed method, each image is replaced by $g_S(x,y)$ and $d_E(x,y)$. Thus, the following linear system generates the depth map [14]

$$(L + \alpha D_E)D = (\alpha D_E D_D), \tag{12}$$

where each D represents the depth map in vector form, D_E the vectorized version of $d_E(x,y)$, D_D the diagonal matrix to decide the edge, L the matting Laplacian function created by $e(x,y)$, and α the parameter that determines the smoothness of D. Therefore, the proposed system generates the depth map $d(x,y)$ by converting D to 2D matrix version.

4. Experimental results

In order to evaluate the performance of the proposed system, a set of stereo images was acquired using a camera equipped with a F1.8 lens and the sensor array shown in Fig. 2(a). The test camera module is shown in Fig. 7. A set of stereo images with vertical disparity was captured under a normal illumination condition, and another set was acquired by rotating the camera by 90°.

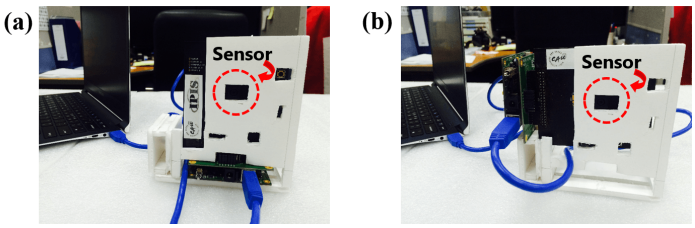

Fig. 7. A prototype camera to acquire four stereo images. (a) a setup to acquire the horizontal disparity and (b) the 90° rotated version to acquire the vertical disparity.

Using the camera shown in Fig. 7, four stereo images of size 1024×1024 were acquired as shown in Figs. 8(a)-8(d). A same scene was captured to make four stereo images. The distances of the nearest and farthest objects are respectively 20 and 110 cm, and all objects have the same interval of 10 cm. From Figs. 8(a)-8(d),the disparity tends to decrease as the distance between the imaging sensor and object increases as shown in Fig. 8(e). Since x and y disparity curves

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

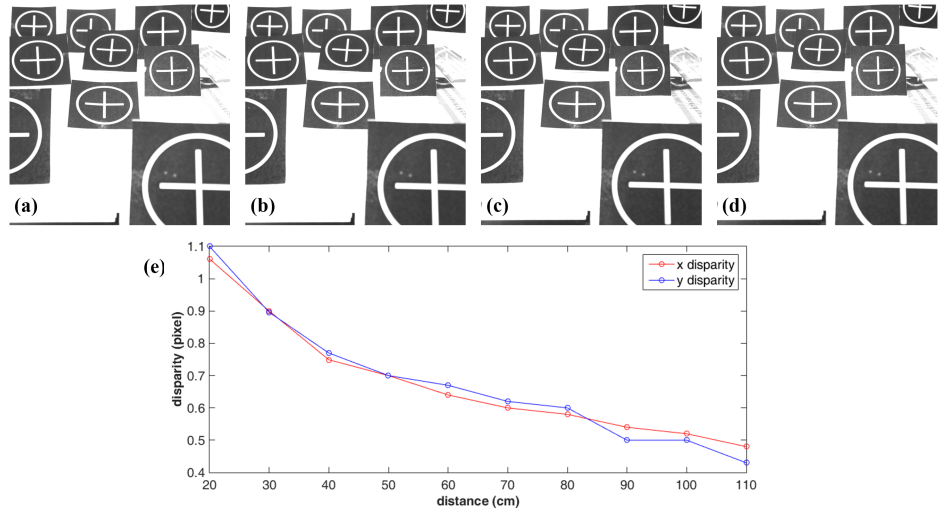

Fig. 8. Acquired four stereo images using the proposed sensor array. (a) left image, (b) right image, (c) top image, (d) bottom image, and (e) the disparity measurement results of (a)-(d).

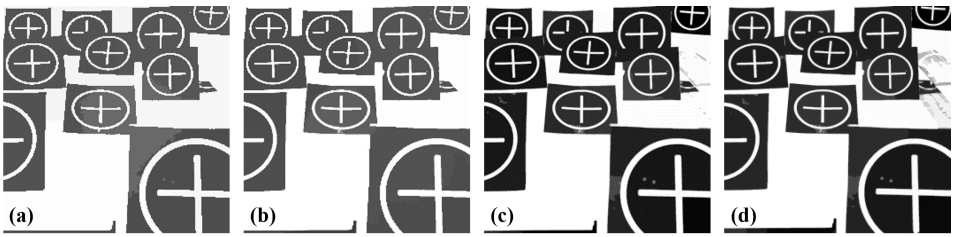

Fig. 9. Segmentation results of Fig. 8(a) using different algorithms. (a) k-means clustering, (b) mean-shift, (c) variational method by Koepfler [27], and (d) the proposed segmentation method.

are almost identical, the proposed sensor can estimate the disparity in the entire range between 20 and 100 cm.

Segmentation results of Fig. 8(a) using four different algorithms are shown in Fig. 9. In this experiment, the proposed method was compared with k-means clustering, mean-shift, and variational segmentation [27] methods. As shown in Figs. 9(a)-9(c), three existing methods produce incorrectly segmented results because of the sensitivity to the brightness change. Especially, existing methods fail to segment the ruler as shown in the right side of Fig. 8(a). On the other hand, the proposed method can accurately segment the ruler as shown in Fig. 9(d).

To evaluate the depth map generation performance of the proposed system using the sensor array shown in Fig. 5, five existing stereo matching methods were compared with the proposed method: the PSF by Zhuo [14], region-based stereo matching by Mukherjee [29], the enhanced normalized cross correlation by Psarakis [30], recursive edge-aware filters by Çığla [31], hierarchical phase correlation using the sensor of Fig. 2(a) [25] as shown in Figs. 10(a)-10(e). The proposed system can generate more reliable depth map than other methods as shown in Fig. 10(f).

The depth maps were estimated from three sets of stereo images using the proposed system

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

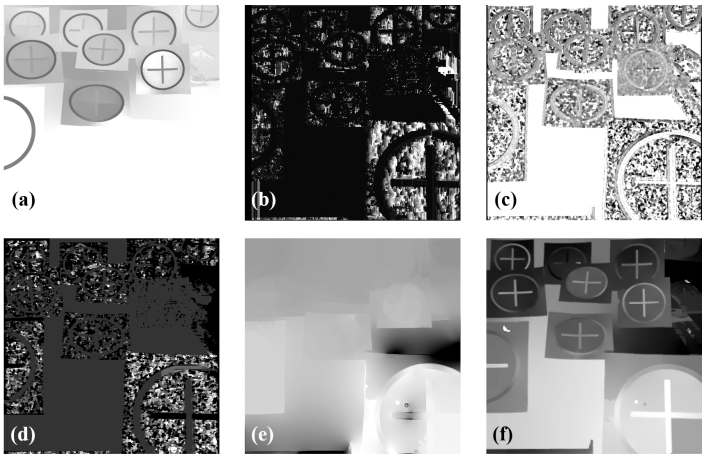

Fig. 10. Results of dense depth map generation: (a) depth map using defocus map [14], (b) region-based stereo matching [29], (c) normalized cross correlation [30], (d) recursive edge-aware filters [31], (e) hierarchical phase correlation [25], and (f) the proposed system.

as shown in Fig. 11. Input images of Figs. 11(a)-11(c) have the background of distance of 100, 230, and 350 cm respectively, and all objects have the distance within 100 cm. Although some incorrect disparities are observed in the background region as shown in Figs. 11(b) and 11(c), the proposed system generates the dense depth map with an acceptable accuracy.

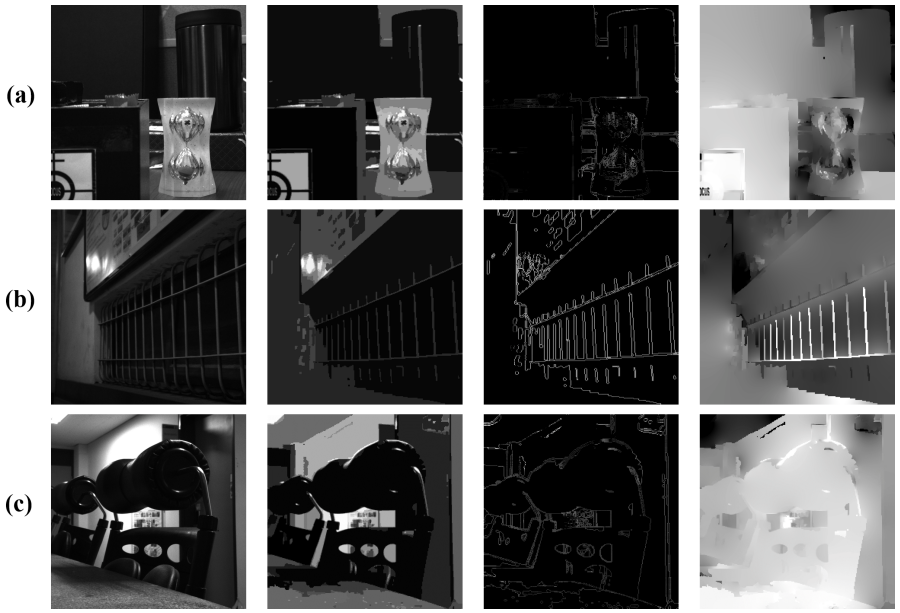

Fig. 11. Depth map generation results of various input images: (a) results of 'Desk', (b) 'Hallway', and (c) 'Chair'.

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016

5. Conclusions

In this paper, a novel single sensor-based depth map generation method is presented. Existing stereo matching systems require bulky, expensive hardware to measure the disparities because of an additional camera. The proposed system uses the dual pixel-type imaging sensor with black masks that generate stereo images to estimate the disparity. Since the proposed sensor array does not generate the horizontal error, it can accurately obtain a depth map using multiple disparities. The proposed variational segmentation method is simply implementable and generates an acceptable segmentation result without noise amplification. Moreover, the proposed system generates depth map by computing the disparities using hierarchical block matching. The proposed motion estimation method measures the sub-pixel disparity in response to a very fine angle of view. Consequently, the proposed system can generate the depth map using a single image sensor with multiple objects and additive noise.

Acknowledgments

This work was supported by Institute for Information & Communications Technology Pro-motion (IITP) grant funded by the Korea government (MSIP) (B0101-1-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis), the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16- 1018) supervised by the IITP (Institute for Information & communications Technology Promotion), the Technology Innovation Program (Development of Smart Video/Audio Surveillance SoC & Core Component for Onsite Decision Security System) under Grant 10047788.

#263896 Received 25 Apr 2016; revised 26 May 2016; accepted 29 May 2016; published 3 Jun 2016