SIViP (2014) 8:287–297
DOI 10.1007/s11760-013-0550-2

ORIGINAL PAPER

Efficient depth map generation with occlusion handling for various camera arrays

Woo-Seok Jang · Yo-Sung Ho

Received: 24 April 2012 / Revised: 27 August 2013 / Accepted: 27 August 2013 / Published online: 28 September 2013

Abstract In this paper, we propose a direct depth map acquisition method for the arc camera array as well as the parallel camera array. In conventional stereo matching algorithms, image rectification is necessary where disparity values are obtained by identifying correspondences in the identical horizontal line of stereo images. The acquired disparity values are then transformed to depth values. However, image rectification may fabricate unwanted outcomes related to the arc camera array. Thus, the proposed method excludes image rectification and directly extracts depth values using an epipolar constraint. In particular, occlusion detection and handling processes are inserted to increase depth map accuracy. Further, belief propagation-based energy optimization is employed to confirm occlusion regions. Then, reasonable depth values are assigned to obtained occlusion regions using distances and color differences of neighbor pixels. Experimental results show that compared to the conventional methods, the proposed method generate more stable depth maps with fewer limitations.

Keywords Epipolar constraint · Image rectification · Occlusion handling · Stereo matching

1 Introduction

Recently, interests in 3D entertainment systems have risen thanks to the financial success of numerous 3D films. The 3D image can be acquired by multiple cameras; multiview images are captured with wide viewing angles. In addition, depth images are necessary for a wider range of applications [1,2]. Depth images represent depth information of the 3D scene; depth information allows synthesis of non-existing views by employing depth image-based rendering (DIBR) [3,4].

Three approaches exist to acquiring depth information: active depth sensors, passive depth sensors, and hybrid depth sensors. Active depth sensors [5] directly measure depth data using a physical sensor, while passive depth sensors apply correlation of images captured from at least two cameras [6,7]. Hybrid depth sensors combine the two methods to cover their weaknesses [8]. Active and hybrid depth sensors can use only low-resolution images due to hardware limitations and require additional expensive physical sensors, e.g., depth cameras. Therefore, we acquire depth data using passive depth sensors at low price despite lengthy processing time and relatively low accuracy.

Stereo matching [9], a widely researched topic in computer vision, is one of the passive depth sensors. Fundamentally, two images of the same scene are given which are taken from slightly different viewpoints. Naturally, 3D data are acquired to find the corresponding points in the other image. In the human visual system, this is applied for guessing depth information from disparities obtained from both eyes. Stereo matching algorithms to date have been mainly developed for the parallel camera array [10]. Yet, the arc array is more actively used for the similar field of view when 3D image acquisition is processed, e.g., filming. In this case, the conventional algorithms used in the parallel arrays are not effective.

Most stereo matching algorithms rectify both images for the sake of simplicity and accuracy. Consecutively, corresponding points are found in the identical horizontal line of

W.-S. Jang (✉) · Y.-S. Ho
Gwangju Institute of Science and Technology (GIST),
123 Cheomdangwagi-ro, Buk-gu, Gwangju 500-712,
Republic of Korea
e-mail: jws@gist.ac.kr
URL: http://vclab.gist.ac.kr

Y.-S. Ho
e-mail: hoyo@gist.ac.kr

two images. However, image rectification comes with drawbacks in the arc array such as image distortion although which may be useful in the parallel array. This is due to the fact that while two images are originally on similar planes in parallel, their convergence angles are compulsorily transmitted to the same plane in the arc array.

In this paper, we discuss solving the problems of conventional methods that appear in the arc array. Initially, we examine image rectification, a prerequisite stage in conventional stereo matching methods. We propose a direct depth extraction method using an epipolar constraint to overcome the discussed issues. Further, occlusion detection and handling methods are also proposed to enhance the quality of depth maps. The proposed method displays less limitation and generates more flexible depth maps.

The remainder of this paper is organized as follows. In Sect. 2, we briefly explain the problem of 3D information acquisition via image rectification in the arc array. In Sect. 3, we explain our depth acquisition method based on an epipolar constraint. Section 4 analyzes experimental results of the proposed method. Finally, our conclusions are presented in Sect. 5.

2 Rectification problem in the arc array

In stereo matching, most recent algorithms are based on image rectification. Specifically, corresponding points are assumed to be on the identical horizontal line; this assumption is used for obtaining 3D information. Image rectification is a transform process which makes epipolar lines of stereo images parallel. The rectified images possess parallel epipolar lines on the coplanar image planes [11]. Thus, the corresponding points of two images share the same vertical coordinate. As a result, the only difference of the images becomes a mere horizontal displacement. Naturally, 3D information can be obtained by disparity-to-depth transformation. Figure 1 illustrates image rectification in the arc camera array. Two views located on distinct planes are transformed to coplanar image planes with epipolar lines of corresponding points becoming identical.

P_L and P_R represent arbitrary corresponding points in the left and right images. The real 3D position of the corresponding point is expressed as 'P.' l_L and l_R are epipolar lines of the corresponding point in the left and right images.

Image rectification offers advantages in obtaining accurate 3D information in the parallel camera array while causing problems in the arc camera array. Distortion of the original images occurs due to the transformation of image planes. Figure 2 illustrates the cases of the rectification problems. In Fig. 2, several pixels in the original plane transform to one pixel in the transformed plane. This induces information loss of some parts in the image. In order to try to solve

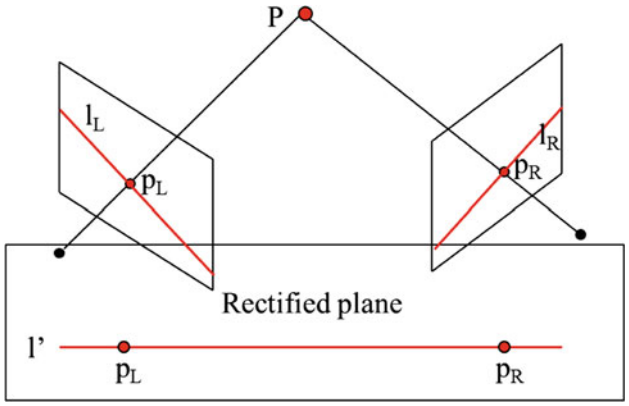

Fig. 1 Rectification in arc camera array. After image rectification, the corresponding points are found in the same horizontal line of two images. 3D information can be obtained by disparity-to-depth transformation

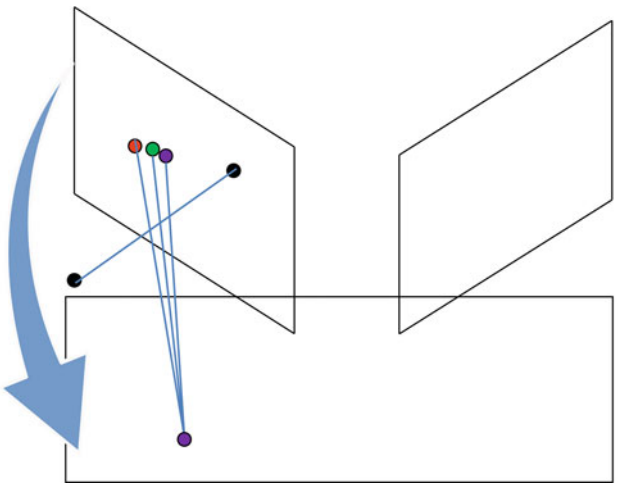

Fig. 2 Examples of image rectification problem. Several pixels are transformed to only one pixel position by image rectification. This causes loss of image information

the problem, image up-sampling or rectification filling can be applied. However, they are inaccurate processes including errors. In the proposed method, since error-free depth estimation is possible using original images, we do not need extra steps such as up-sampling and filling process of information loss.

Figure 3 shows the example of image rectification without filling process. Figure 3a and b represents the original and rectified images, respectively. Figure 3c represents transformation from the rectified image to the original image plane. Accurate depth estimation of information loss parts in Fig. 3c is impossible. Acquiring 3D information by preserving original images of the arc array cannot be solved by image rectification. Thus, we propose a depth acquisition method that excludes image rectification which fabricates image distortion.

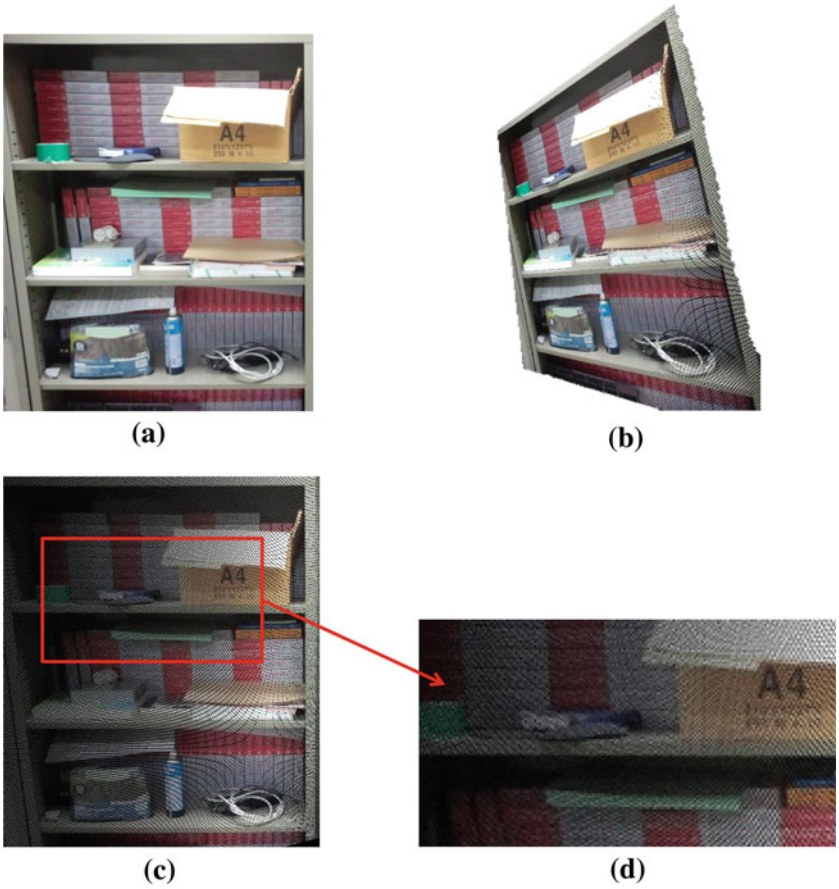

Fig. 3 Result of image rectification without filling process. Image rectification generates distortion and information loss. **a** Original image. **b** Rectified image. **c** Transformation from the rectified image to the original image plane. Accurate depth estimation lost information parts in **c** is difficult **d** Part of image rectification result

Related to stereo matching without rectification, Robert and Deriche [12] found corresponding points via camera parameters instead of searching in the identical horizontal line. However, their method induces high error rate since they use three images including aligned epipolar lines for depth estimation. The main contribution of their work is preservation of discontinuities. For this, they disallow regularizing and smoothing across such discontinuities. In the proposed method, we also consider this through smoothness strength of the energy function. Lopez et al. implement stereo matching using segmentation [13]. Since segments are used as the matching primitive, the performance of this method depends on the effectiveness of segmentation. Furthermore, the assumption of constant depth in each segment may cause problems since this is improper in practice.

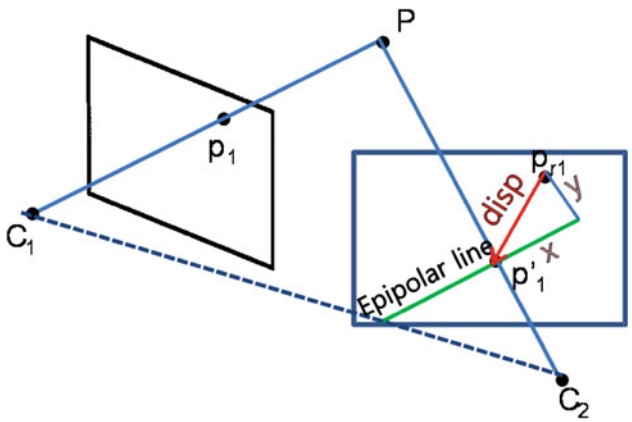

Fig. 4 Corresponding point found according to epipolar constraint from stereo images. Direct depth acquisition without depth transformation of disparity information can be processed by epipolar constraint

3 Depth map acquisition using epipolar constraints

3.1 Epipolar constraints in stereo images

As discussed above, finding corresponding points in the same horizontal line through image rectification generates visual distortion. Further, obtaining disparity values in the whole region is burdensome. Nonetheless, unconditionally scanning both horizontal and vertical directions is ineffective in terms of time and accuracy. In order to reduce the correspondence problem to one-dimensional search, epipolar constraints can be beneficial. Figure 4 illustrates corresponding point finding by using epipolar constraints. For the stereo view denoted by C_1 and C_2, we have a 3D position P and

two image points—p_1 and p'_1—corresponding points on each view image [14]. p_{r1} is the collocated position of p_1 in the right image. The epipolar plane is defined by the three points: P, C_1, and C_2. Thus, p'_1, the corresponding point of p_1, is on the epipolar plane. The epipolar line is defined as the line that intersects the image plane and epipolar plane. Therefore, the corresponding point of p_1 is located on the epipolar line in the right image plane [14]. Defining the disparity between corresponding points without rectification is challenging since the disparity direction is not only horizontal but also vertical. Figure 4 illustrates that the disparity between collocated position 'p_{r1}' and the corresponding point 'p'_1' has two directions 'x' and 'y.' Furthermore, disparity values may be negative-signed.

Since common stereo matching algorithms employ one-dimensional disparity, disparity information can be easily transformed to depth information by (1).

$$d = \frac{f \times l}{\text{disp}} \tag{1}$$

d, f, l, and disp represent depth value, focal length, gap between cameras, and disparity value, respectively. Detection of corresponding points by epipolar constraints does not require one-directional search due to the omission of rectification. Thus, we present direct depth acquisition without depth transformation of disparity information.

3.2 Direct depth value extraction

We use a global method to extract depth values. Principally, an energy function in regard to matching is defined by maximum a posterior Markov random field (MAP-MRF), solved through optimization [15]. In general, compared to local methods, global methods provide higher quality results at the cost of lengthier execution [16]. Since accuracy is more prioritized, we adopt a global method. We solve the high-complexity problem in stereo matching by means of a hierarchical structure. The energy function for matching is defined as in (2) [15].

$$E(x, y, d) = E_{\text{data}}(x, y, d) + \lambda_{\text{smooth}} E_{\text{smooth}}(x, y, d) \tag{2}$$

where x and y are horizontal and vertical coordinates of the reference image while d represents the depth value. The energy function, $E(x, y, d)$, comprises a data term and a smoothness term. The matching costs for data and smoothness terms are defined in (3). The smoothness strength λ_{smooth} is adaptively refined by considering discontinuity regions [7].

$$E_{\text{data}}(x, y, d) = |I_{\text{L}}(x, y) - I_{\text{R}}(x', y', d)| \tag{3}$$

$$E_{\text{smoothness}}(x, y, d) = \sum_{(p,q)\in N} W(p, q) \tag{4}$$

$I_{\text{L}}(x, y)$ is the pixel value in the left image given (x, y) coordinate. $I_{\text{R}}(x', y', d)$ is the matched pixel value in the right image given the depth value at (x, y) in the left image, denoted by d. N and $W(p, q)$ represent neighboring pixels and their depth value difference, respectively. In order to find the matched pixel in $I_{\text{R}}(x', y', d)$, two processes for 3D projective transformation are used. First, the left image pixel is backprojected to the 3D space based on the camera parameters and depth information. Then, the backprojected pixel in the 3D space is projected to the right image [17]. 3D projective transformation is performed as follows.

$$(x, y, z)^T = R_{\text{src}} A_{\text{src}}^{-1} (u, v, 1)^T d_{u,v} + t_{\text{src}} \tag{5}$$

$$(l, m, n)^T = A_{\text{dst}} R_{\text{dst}}^{-1} (x, y, z)^T - t_{\text{dst}} \tag{6}$$

$$(u', v') = (1/n, m/n) \tag{7}$$

where A_{src}, R_{src}, and t_{src} are internal, rotation, and translation parameters in the left image, respectively. Similarly, A_{dst}, R_{dst}, and t_{dst} are those in the right image. $d_{u,v}$ values are the depth candidates at (u, v) coordinate in the image. The left image pixel is sent to the 3D space by (5) and projected to the right image by (6). (u', v') in (7) represents the projected coordinate to the right image. Through such processes, $I_{\text{R}}(x', y', d)$ in (3) is determined and the energy function, $E(x, y, d)$, is solved by the optimization method provided from Constance Space Belief Propagation (CSBP), [18] considering complexity. The obtained depth map provides initial values for generating more accurate depth information.

3.3 Occlusion detection and handling

Since stereo images are captured from different positions, occlusion occurs; some pixels are visible only in one image [7]. Therefore, obtaining an accurate depth map is difficult solely by the energy function of depth extraction defined in Sect. 3.2. In order to solve this problem, occluded pixel detection and reasonable depth value designation are critical.

We propose several constraints for occlusion detection based on the uniqueness constraint [7]. Firstly, we project all pixels in the left image to the right image using the left depth map to generate the left occlusion map. In this case, two candidates for the occlusion region exist. The first candidate is the group of pixels which are projected to outside the right image plane. They are regarded as occlusion due to the unmatching pixels in the right image. The energy function is defined in (8).

$$E_{\text{O}}(d_{u,v}) = \sum_{u,v} |o_{u,v} - G_{\text{O}}(u, v; d_{u,v})| \tag{8}$$

$G_{\text{O}}(u, v; d_{u,v})$ is a binary map constructed as the result of projection. If the pixel is projected to inside the right image

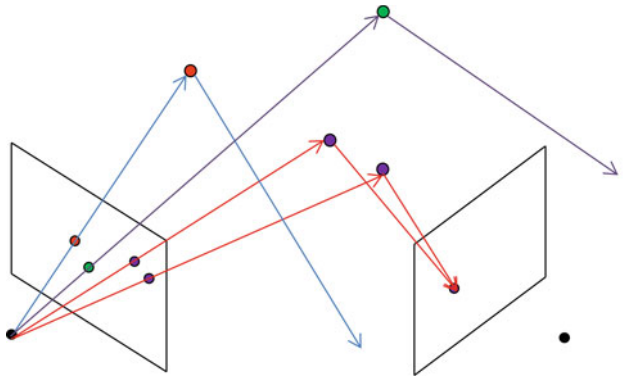

Fig. 5 Occlusion detection using projection. Unmatching and many-to-one mapping pixels are regarded as occlusion

plane, $G_{\mathrm{O}}(u, v; d_{u,v})$ is set to '0,' otherwise, '1.' We express the occlusion value $o_{u,v}$ in binary, '1' and '0' representing occlusion and non-occlusion, respectively.

The second candidate is the case that many pixels in the left image are projected to the same pixels in the right image. In such a case, several pixels are actually occluded by other pixels in the 3D space. The energy function regarding this is defined in (9).

$$E_{\mathrm{W}}(d_{u,v}) = \sum_{u,v} |o_{u,v} - G_{\mathrm{W}}(u, v; d_{u,v})| \tag{9}$$

$G_{\mathrm{W}}(u, v; d_{u,v})$ is set to '1' in the case of many-to-one mapping. Among the matching pixels in the left image, the nearest point from the right camera is more likely to be visible. Thus, the weighting factor applied to the nearest pixel from the right camera is different from those applied to the others. Figure 5 illustrates occlusion detection method through projection.

The other constraint for occlusion detection checks whether corresponding pixels from interim depth maps are located at the same positions in the 3D space. If a particular pixel in the left image is not occluded, the 3D position of this pixel is identical to that of the corresponding pixel in the right image. The energy function for this constraint is defined as follows.

$$E_{\mathrm{c}}(d_{u,v}, d_{u',v'}) = \sum_{u,v} |o_{u,v} - G_{\mathrm{c}}(u, v; d_{u,v}, d_{u',v'})| \tag{10}$$

If the positions of corresponding points from the two images are equal, $G_{\mathrm{c}}(u, v; d_{u,v}, d_{u',v'})$ is set to '0,' otherwise, '1.' Figure 6 illustrates this process.

So far, we have presented occlusion detection method using several constraints. The final energy function combining these constraints for occlusion detection is defined as

$$E_{\mathrm{OD}} = \sum_{u,v} [(1 - o_{u,v}) E_{\mathrm{data}}(x, y, d_{u,v}) + \lambda_{\mathrm{a}} o_{u,v}]$$

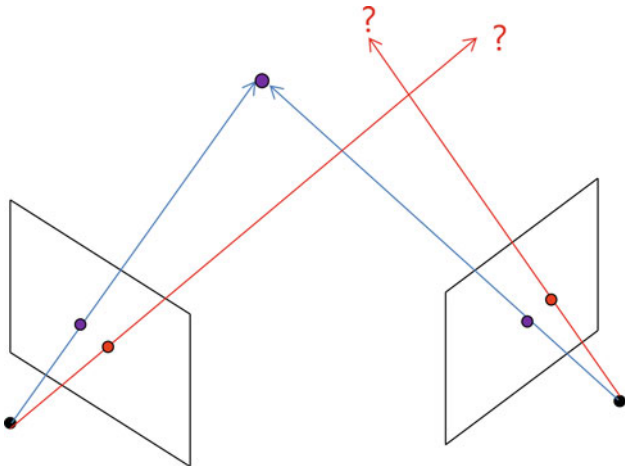

Fig. 6 Occlusion detection using cross-check in 3D space. Occlusion is determined by checking consistency between stereo images

$$+\lambda_{\mathrm{O}} E_{\mathrm{O}}(d_{u,v}) + \lambda_{\mathrm{w}} E_{\mathrm{w}}(d_{u,v}) + \lambda_{\mathrm{c}} E_{\mathrm{c}}(d_{u,v}, d_{u',v'}) + \sum_{u,v} \sum_{p,q \in N} \lambda_{\mathrm{s}} |o_{\mathrm{p}} - o_{\mathrm{q}}|. \tag{11}$$

If the probability that the pixel at (u, v) being occluded is high, each binary map from (8), (9), and (10) is set to '1.' Otherwise, the binary map is set to '0.' Thus, when the hypothesized occlusion value and binary map are equal, the final function (11) has low energy. (11) also includes the color difference for data term in addition to the discussed constraints, coming from the assumption that large color differences generate mismatches, although a particular pixel may be regarded as visible. The last represents the smoothness term for the energy function of occlusion detection. The final energy function (11) is optimized with respect to $o_{u,v}$ by belief propagation. Occlusion value in position (u, v) is computed as follows.

$$O_{u,v} = \arg\min_{O_{u,v}} E_{\mathrm{OD}}. \tag{12}$$

After occlusion detection, reasonable depth values should be assigned to occluded pixels. Since occlusion is visible only in one image, determining accurate depth values is impractical. Yet, depth values of occlusion are similar to those of non-occlusion in the background. Hence, by using initial depth values from neighboring pixels in the non-occluded region, depth can be effectively estimated in the occluded region. Initial depth map was already obtained by (2). In the proposed method, we perform occlusion handling by propagating depth values of non-occluded pixels to the occlusion. For this, we propose a potential energy function as a measure of likeness which is defined by (13) according to the distances and color differences from neighbor pixels. The depth value of the most analogous region is selected as the optimal depth value of the current occluded pixel.

Table 1 Parameters used in experiments

	λ_a	λ_O	λ_w	λ_c	λ_s	σ
Value	7.5	12	3	12	4.2	7

$$E_{OH}(u, v, d_{u,v}) = \sum_{(p,q)\in N_v} (1 - o_{p,q}) \frac{1}{\text{dist}[(u, v), (p, q)]} \times \exp\left(-\frac{\text{diff}[(u, v), (p, q)]}{\sigma^2}\right) \quad (13)$$

N_v means neighboring pixels whose spatial distances from occluded pixels are less than the predefined value. In this paper, '41' is used as the predefined distance value. dist$[(u, v), (p, q)]$ is the spatial distance between occluded pixel (u, v) and neighboring visible pixel (p, q), while diff$[(u, v), (p, q)]$ is the color difference. The depth value, which has the maximum value of (13), is determined as the optimal depth value for the pixel (u, v). This process assigns the optimal disparity by finding the similar region to the occlusion part according to the weighting of distance in a pixel-by-pixel manner. The occlusion handling process works at only occluded pixels which are near non-occluded pixels. Thus, it completely handles thin or small occlusion. However, wide and large occlusion is processed at only near non-occluded pixels. In order to solve this problem, we apply the potential energy function (13) for occlusion handling repeatedly until all occluded pixels are filled. Finally, we apply post-processing to the acquired depth map to improve the quality [19].

4 Experimental results and analysis

In order to evaluate the performance of the proposed method, we tested five stereo images with two kinds of camera arrangement. These test data are *Newspaper* and *Café* in the parallel array and *Fitness1*, *Fitness2*, and *Friends* in the arc array. The resolutions of the test images are 1,920 × 1,080 except for *Newspaper*. *Newspaper* is captured at 1,024 × 768 resolution. Color correction was applied to these data sets for accuracy of matching [20]. The camera parameters of these images are estimated by camera calibration [21] for image rectification and 3D projective transformation. Table 1 lists parameter values used in the proposed method. These are acquired by experiments to balance energy terms.

Figure 7 shows the initial depth map directly obtained by epipolar constraints, and the occlusion map acquired by the proposed occlusion detection in the parallel camera array.

Figure 8 shows the results of the proposed method for direct depth value extraction and occlusion detection in the arc camera array.

Figures 9, 10, 11, 12, 13 illustrate visual comparisons of depth maps—results of the proposed method and the conventional methods. The methods of (a), (b), and (c) in the figures generate the depth image by transforming disparity map obtained through rectification to depth map by (1). The results of (a), (b), and (c) in the figures are obtained by applying accelerated belief propagation (BPA) [22], synchronous belief propagation (BPS) [22], and CSBP. The results

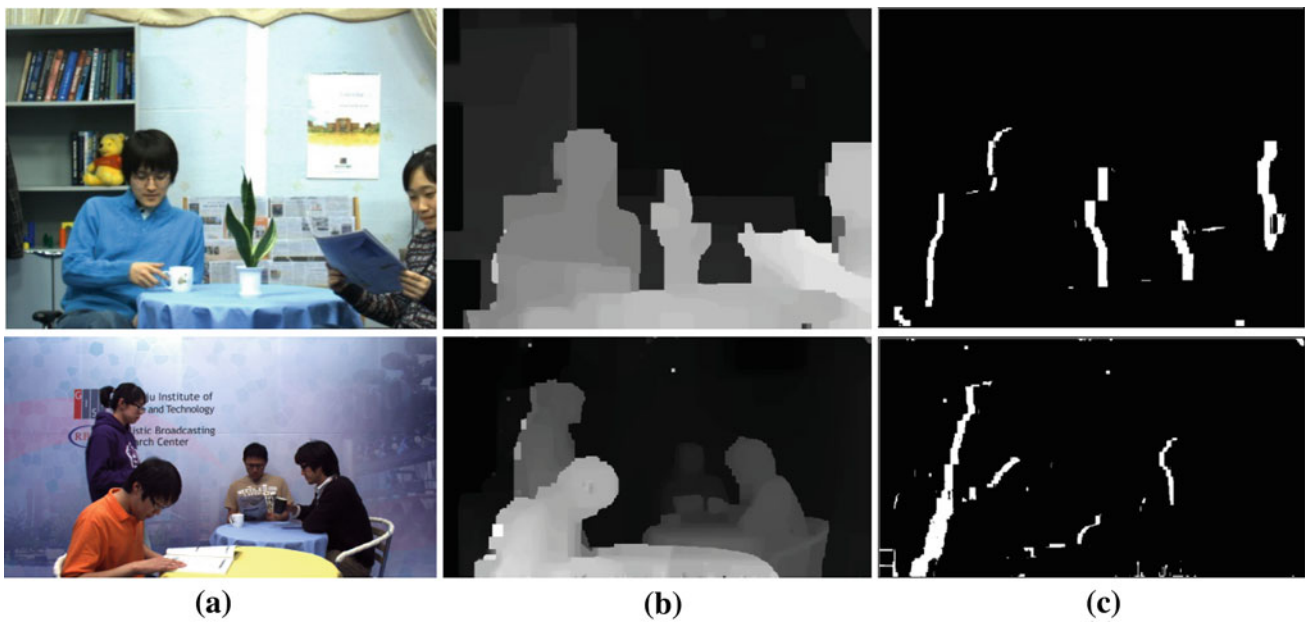

Fig. 7 Initial depth and occlusion map obtained by the proposed method in the parallel camera array. **a** Original image. **b** Initial depth map. **c** Result of occlusion detection

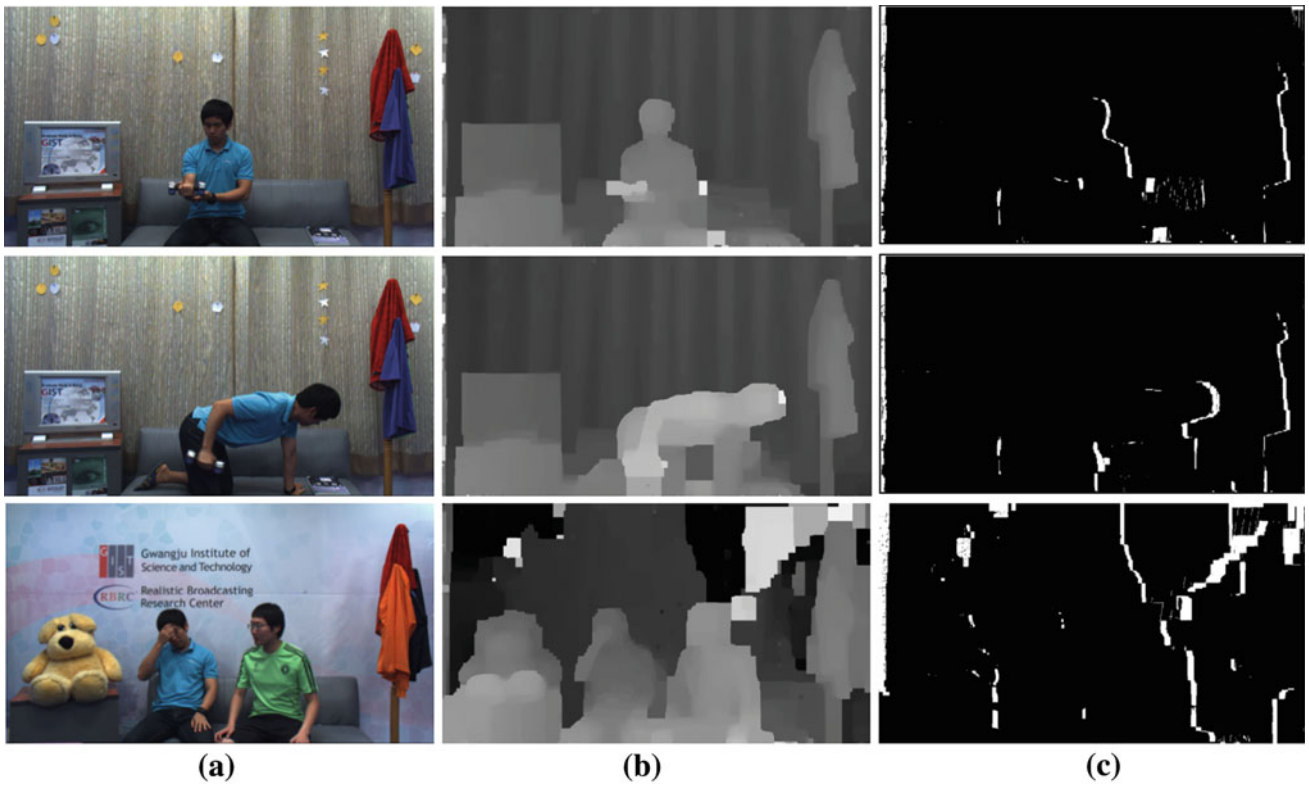

Fig. 8 Initial depth and occlusion map obtained by the proposed method in the arc camera array. **a** Original image. **b** Initial depth map. **c** Result of occlusion detection

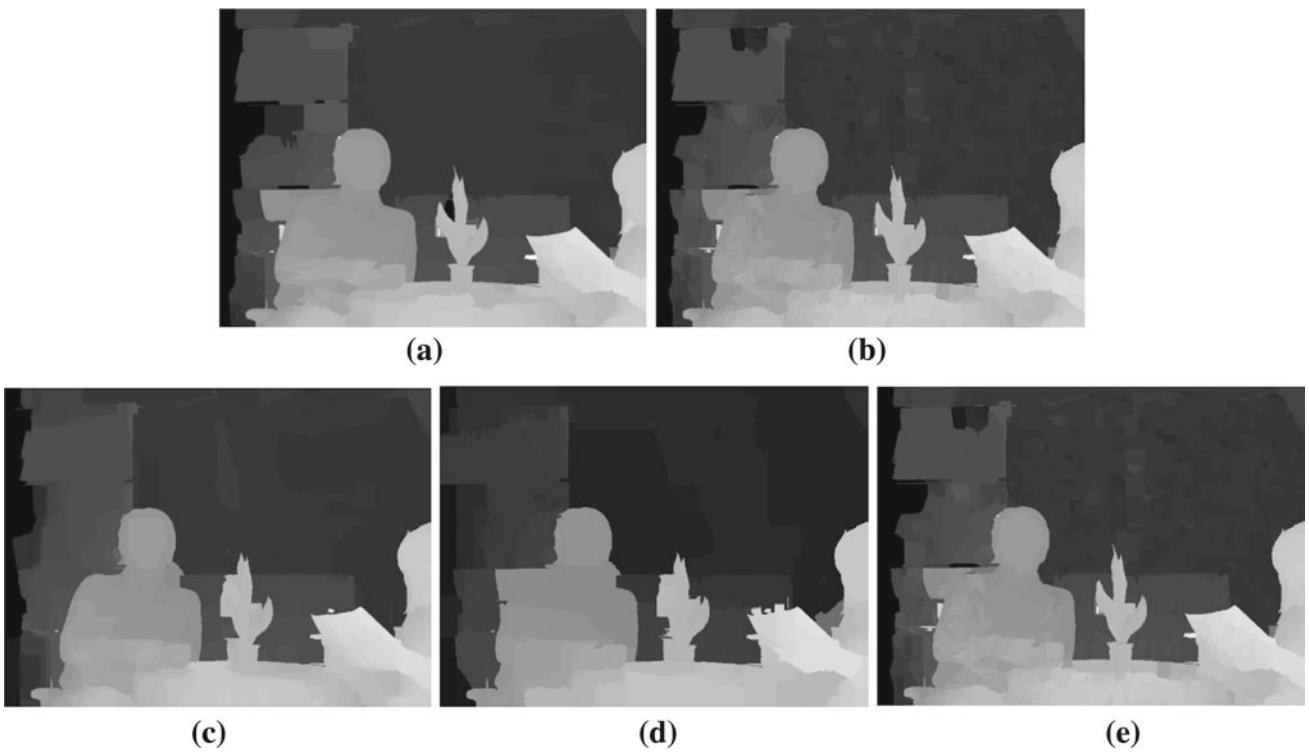

Fig. 9 Depth map comparison of *Newspaper*. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's method. **e** Proposed method

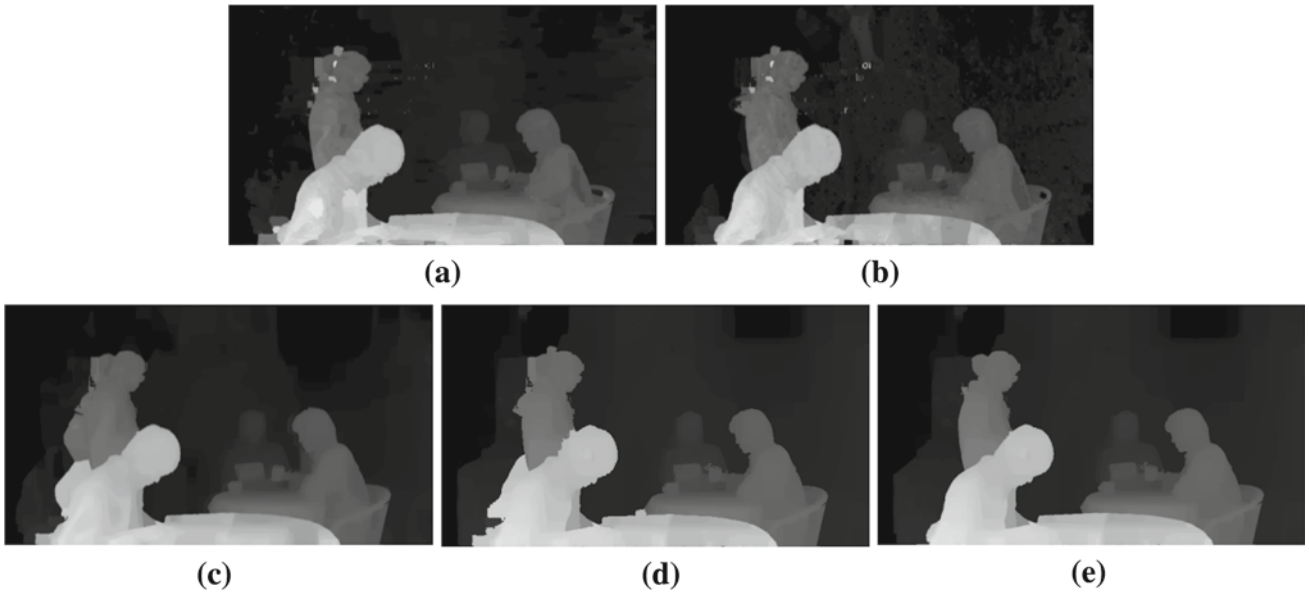

Fig. 10 Depth map comparison of *Café*. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's method. **e** Proposed method

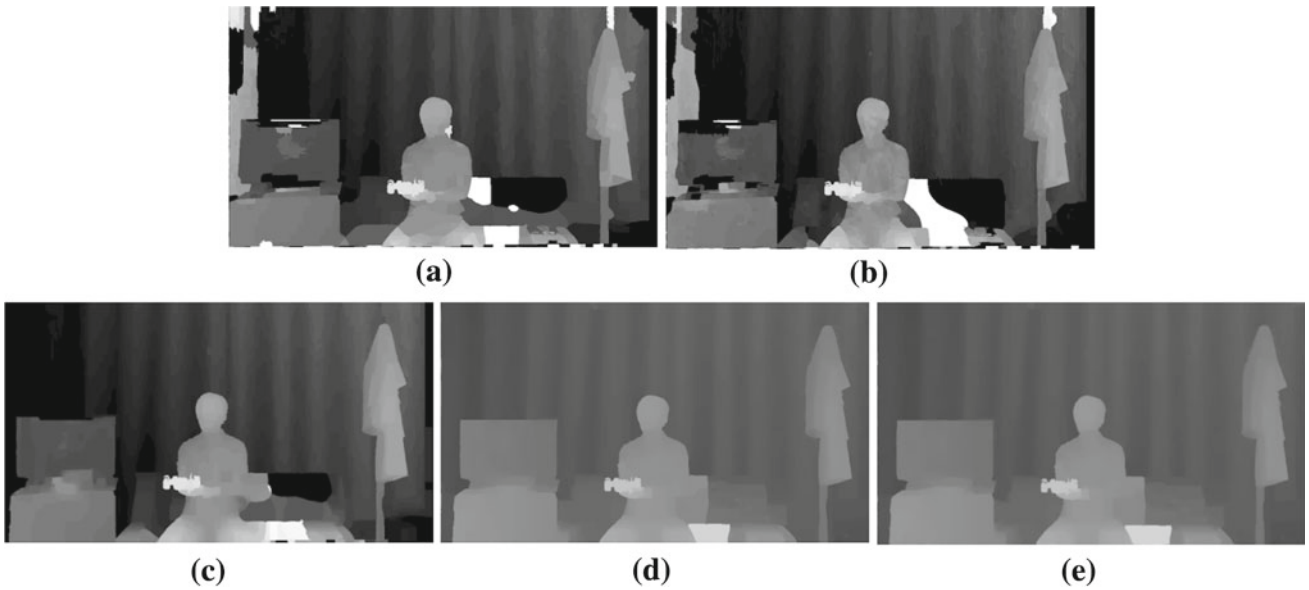

Fig. 11 Depth map comparison of *Fitness1*. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's method. **e** Proposed method

of (d) in the figures are obtained using the theory based on Robert's approach [19]. The approach does not apply image rectification. Since the conventional methods were adapted to fit our framework including the depth refinement process, their results are improved compared to their original results.

In the parallel camera array, image rectification considerably enhances depth accuracy. Nevertheless, the proposed method without image rectification generates high-quality results as well.

In the arc camera array, some parts of the conventional results lose information, implying inaccuracy. On the other hand, depth values of the whole region are estimated in the proposed results.

In order to evaluate the performance of the proposed method objectively, we applied view synthesis and calculated PSNR values of the synthesized views for our method and the conventional methods. Tables 2 and 3 represent comparisons of their performances in the parallel and arc arrays.

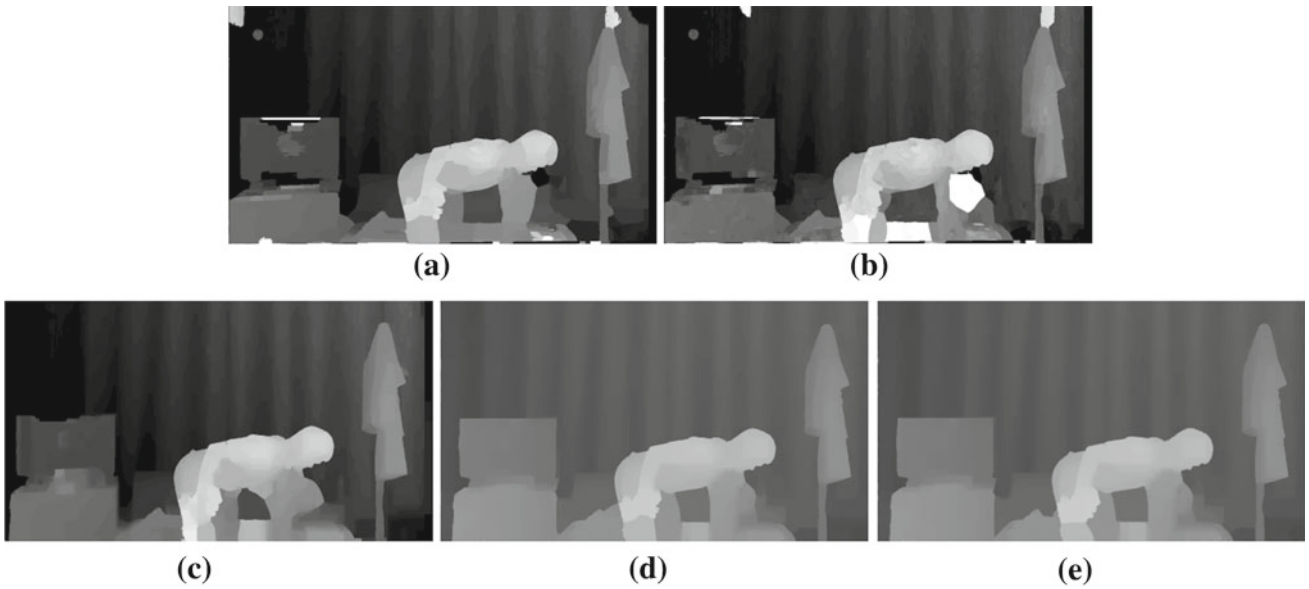

Fig. 12 Depth map comparison of *Fitness2*. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's method. **e** Proposed method

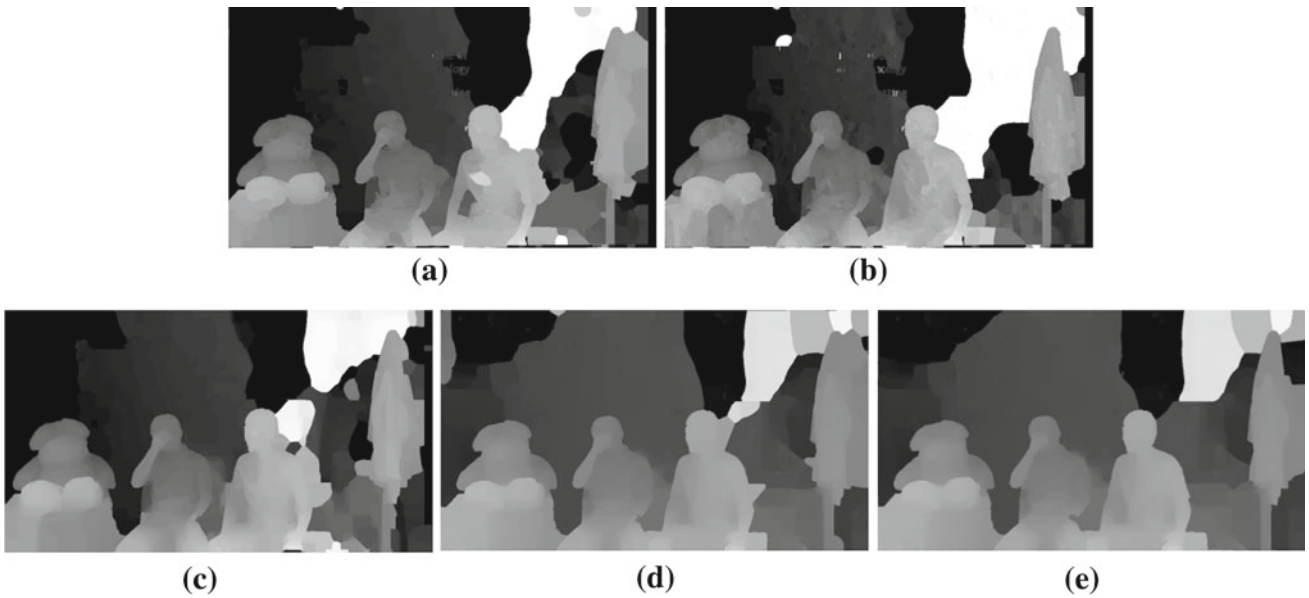

Fig. 13 Depth map comparison of *Friends*. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's method. **e** Proposed method

Table 2 PSNR comparison of view synthesis result in the parallel array (dB)

	BPA	BPS	CSBP	Robert's method	Proposed method
Newspaper	28.59	29.14	28.37	27.98	29.64
Cafe	32.25	31.23	32.21	32.45	33.13
Average PSNR	30.42	30.19	30.29	30.22	31.39

Table 3 PSNR comparison of view synthesis result in the arc array (dB)

	BPA	BPS	CSBP	Robert's method	Proposed method
Fitness1	20.51	20.55	20.23	31.43	32.73
Fitness2	20.17	20.08	19.83	31.33	32.44
Friends	18.55	18.54	18.39	30.68	31.11
Average PSNR	19.74	19.72	19.48	31.15	32.09

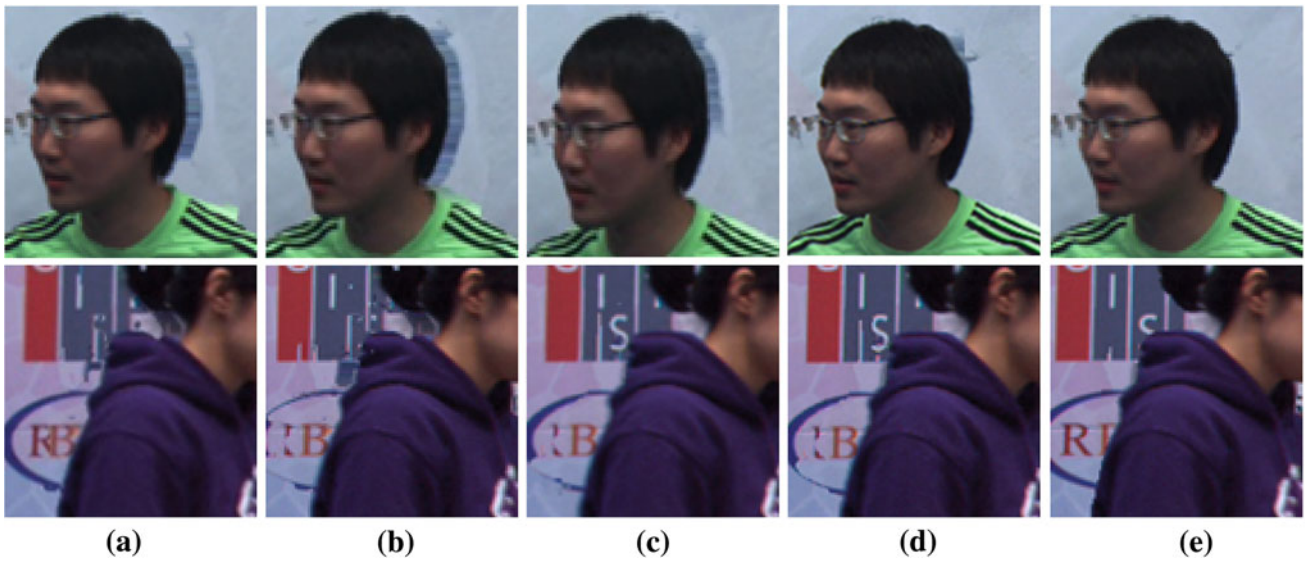

Fig. 14 Parts of view synthesis results near occlusion. **a** BPA. **b** BPS. **c** CSBP. **d** Robert's. **e** Proposed method

Table 4 Runtime comparison (sec.)

	BPA	BPS	CSBP	Proposed method
Newspaper (1,024 × 768)	9,357	21,965	6,622	14,200
Café (1,920 × 1,080)	183,319	611,021	71,580	37,303
Fitness1 (1,920 × 1,080)	52,588	170,914	17,457	37,156
Fitness2 (1,920 × 1,080)	108,589	339,006	21,035	37,448
Friends (1,920 × 1,080)	103,257	329,059	20,801	37,202
Average	91,422	29,4393	27,499	32,661

From the objective evaluation, the synthesis qualities of all methods are similar in the parallel camera array. However, the proposed method produces superior accurate results compared to the conventional methods with image rectification in the arc array. Furthermore, occlusion handling of the proposed method provided better performance compared to Robert's method.

Figure 14 exhibits parts of view synthesis results. The results of Fig. 14 demonstrate that the proposed method improves near occlusion regions regardless of camera arrangements.

Processing time comparison of the conventional and proposed methods is presented in Table 4. Experiments were conducted on a 2.40 GHz CPU processor with a 48 GB RAM. The runtime of the proposed method depends on only image resolution. Thus, runtimes of all images are similar except for Newspaper in the proposed method. On the other hand, runtimes of the conventional methods depend not only on image resolution but also disparity range. While the proposed method had the second fastest depth estimation on average, the performance was the highest.

Table 5 shows the processing time for the main steps of the proposed method. From Table 5, we recognize that initial depth extraction consumes most of the processing time. The other steps are for depth map quality improvement, while their processing time is relatively minimal.

The proposed method allows for generation of accurate depth values without image rectification regardless of any camera arrays. Therefore, our method effectively reduces restrictions in camera settings prior to image capturing.

5 Conclusion

In this paper, we proposed a depth map acquisition method which eliminates drawbacks of conventional algorithms for practicality. In general, stereo matching through image rectification generates efficient results in aspects of accuracy and complexity. However, the use of image rectification is problematic, producing image distortion in the arc array. Thus, the proposed method leaves out the transformation process of disparity information obtained through image rectification; rather, a direct depth acquirement method using an epipolar constraint is applied. Such process increases the probability of obtaining reliable depth information in the whole image region. Furthermore, we improved the depth map quality through occlusion handling. Consequently, our method generated more stable results than the conventional methods in several arrays except the parallel array. Therefore, our method can be beneficial in 3D image production.

Table 5 Stepwise computation time of the proposed method (sec.)

	Initial depth extraction	Occlusion detection	Occlusion filling	Post-processing	Total processing time
Newspaper (1,024 × 768)	13,863	76	32	228	14,200
Café (1,920 × 1,080)	36,740	204	58	302	37,303
Fitness1 (1,920 × 1,080)	36,692	202	44	217	37,156
Fitness2 (1,920 × 1,080)	36,883	203	41	320	37,448
Friends (1,920 × 1,080)	36,453	202	93	454	37,202
Average ratio (%)	98.36	0.54	0.16	0.93	100

Acknowledgments This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009228).

References

1. ISO/IEC JTC1/SC29/WG11: Applications and requirements on FTV. N9466 (2007)
2. ISO/IEC JTC1/SC29/WG11: Vision on 3D video. N10357 (2009)
3. Zhang, L., Tam, W.J.: Stereoscopic image generation based on depth images for 3DTV. IEEE Trans. Broadcast. **51**(2), 191–199 (2005)
4. Fieseler, M., Jiang, X.: Discontinuity-based registration of depth and video data in depth image based rendering. Signal Image Video Process. **5**(3), 353–361 (2011)
5. Frick, A., Kellner, F., Bartczak, B., Koch, R.: Generation of 3D-TV LDV-content with time of flight camera. In: Proceedings of IEEE International Conference on 3DTV, pp. 45–48. Potsdam, Germany, May 4–6 (2009)
6. Shin, I.Y., Ho, Y.S.: Disparity estimation at virtual viewpoint for real-time intermediate view generation. In: Proceedings of International Conference on 3D Systems and Applications, pp. 195–198. Seoul, Korea, June 20–22 (2011), paper S6–2
7. Jang, W.S., Ho, Y.S.: Efficient disparity map estimation using occlusion handling for various 3D multimedia applications. IEEE Trans. Consumer Electron. **57**(4), 1937–1943 (2011)
8. Lee, E.K., Ho, Y.S.: Generation of high-quality depth maps using hybrid camera system for 3-D video. J. Visual Commun. Image Represent. **22**(1), 73–84 (2011)
9. Sharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, Hawaii, Dec 9–10, 131–140 (2001)
10. Raghavendra, U., Makkithaya, K., Karunakar, A. K.: Anchor-diagonal-based shape adaptive local support region for efficient stereo matching. Signal Image Video Process. 1–9 (2013). doi:10.1007/s11760-013-0524-4
11. Kang, Y.S., Ho, Y.S.: An efficient image rectification method for parallel multi-camera arrangement. IEEE Trans. Consumer Electron. **57**(3), 1041–1048 (2011)
12. Robert, L., Deriche, R.: Dense depth map reconstruction: a minimization and regularization approach which preserves discontinuities. In: Proceedings of European Conference on Computer Vision, Cambridge, pp. 439–451, April 15–18 (1996)
13. Lopez, A., Pla, F.: A minimization approach for 3D recovery in region-based stereo vision. In: Proceedings of Image Processing and its Applications, Bangalore, India, pp. 47–51, July 13–15 (1999)
14. Kang, Y.S., Ho, Y.S.: Geometrical calibration of stereo images in convergent camera arrangement. In: Proceedings of IEEE-RIVF International Conference on Computing and Communication Technologies, Ho Chi Minh, Vietnam, pp. 44–47, Feb 27–Mar 1 (2012)
15. Wang, L., Jin, H., Yang, R.: Search space reduction for MRF stereo. In: Proceedings of 10th European Conference on Computer Vision, Marseille, pp. 576–588, Oct 12–18 (2008)
16. Hirschmuller, H., Innocent, P.R., Garibaldi, J.M.: Real-time correlation-based stereo vision with reduced border errors. Int. J. Comput. Vision **47**(1/2/3), 229–246 (2002)
17. Kang, Y.S., Ho, Y.S.: Disparity Map Generation for color image using a TOF depth camera. In: Proceedings of 3DTV Conference, Antalya, Turkey, pp. 46(1–4), May 16–18 (2011)
18. Yang, Q., Wang, L., Ahuja, N.: A constant-space belief propagation algorithm for stereo matching. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 1458–1465, June 13–18 (2010).
19. Yang, Q., Engels, C., Akbarzadeh, A.: Near real-time stereo for weakly-textured scenes. In: Proceedings of British Machine Vision Conference, Leeds, England, pp. 80–87, Sept 1–4 (2008)
20. Jung, J.-I., Ho, Y.-S.: Color correction algorithm based on camera characteristics for multi-view video coding. Signal Image Video Process. (2012). doi:10.1007/s11760-012-0341-1
21. Zhang, Z.: A Flexible New Technique for Camera Calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
22. Tappen, M.F., Freeman, W.T.: Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In: Proceedings of IEEE International Conference on Computer Vision, Nice, pp. 900–906, Oct 13–16 (2003)