Direct Depth Value Extraction Method for Various Stereo Camera Arrangements

Woo-Seok Jang and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Republic of Korea
Telephone: +82-62-715-2258, Fax: +82-62-715-3164
E-mail: {jws, hoyo}@gist.ac.kr

***Abstract*—In this paper, we propose a direct depth map acquisition method for a convergent camera array as well as a parallel camera array. Since image rectification is necessary for conventional stereo matching methods, disparity values are found in the same horizontal line of stereo images. Consecutively, the acquired disparity values are transformed to depth values. However, image rectification may lead to unwanted consequences for a convergent camera array. Thus, the proposed method excludes image rectification and directly extracts depth values via the epipolar constraint. The MRF energy for depth estimation is optimized by modified constant-space belief propagation (CSBP) which has low complexity. In order to acquire a more accurate depth map, we apply post-processing finally. Experimental results show that the proposed method has fewer limitations than conventional methods and generates depth map stably.**

Keywords—disparity map, epipolar constraint, stereo matching

I. INTRODUCTION

Recently, interest in 3D entertainment system is rising with the financial success of several 3D films. In order to acquire 3D image, multiple cameras capture multiview images with wide viewing angle. However, acquiring depth image which represents depth information of 3D scene is necessary for a wider range of applications [1], [2]. The depth information makes it possible to synthesize non-existing views by employing depth image based rendering (DIBR) [3]. The performance of DIBR primarily relies on the quality of depth data.

Depth information can be acquired by several ways: active depth sensors, passive depth sensors, and hybrid depth sensors. Active depth sensors [4] directly measure depth data using a physical sensor, while passive depth sensors extract depth data using correlation of images captured from more than or equal to two cameras [5][6]. Hybrid depth sensors combine both methods to make up for weakness of each method [7]. Active depth sensors and sensors have low resolution images because of hardware limitation and require expensive physical sensors such as depth cameras. Therefore, we acquire depth data using passive depth sensors with low price, although the process time is long and the accuracy is relatively low.

Stereo matching [8], widely researched topic in computer vision, is one of the passive depth sensors. In stereo matching, two images of the same scene taken from different viewpoints are given and 3D data is acquired by finding the corresponding points in the other image. It is the application of human visual system that guesses depth information with disparity obtained from both eyes. Algorithms of stereo matching to date have mainly developed on the assumption that a camera array is parallel. However, a convergent array is used to obtain similar field of view from both images when 3D image acquisition such as filming is processed. In this situation, the previous algorithms used in parallel arrays generate problems.

Most algorithms of stereo matching rectify both images for the sake of simplicity and accuracy. Corresponding points are found in the same horizontal line of two images. However, image rectification produces problems such as image distortion in a convergent array although it may be useful in a parallel array. The reason is as follows. Two images are originally on similar plane in a parallel array, while both images that have a convergent angle are compulsorily transmitted to the same plane in a convergent array.

In this paper, we have a discussion about solving the problem of the conventional methods in a convergent array. We point out the problem of image rectification which is a prerequisite in conventional stereo matching methods. We propose the direct depth extraction method using epipolar constraint to overcome the conventional problems. The proposed method has less limitation and contributes flexible depth map generation.

II. RECTIFICATION PROPLEM IN CONVERGENT ARRAY

The assumption that corresponding points lie in the same horizontal line is used to obtain 3D information in stereo matching. Image rectification is necessary for this. Image rectification is the transform which makes epipolar lines of the stereo images parallel. The rectified images have the parallel epipolar lines on the coplanar image planes [9]. Thus, the corresponding points of two images have same vertical coordinate. As a result, it has only horizontal displacement between two images and we can obtain 3D information by transforming disparity to depth information. Fig. 1 illustrates image rectification in a convergent camera array. Two views located on different planes are transformed to coplanar image plane and epipolar lines of corresponding points become identical.

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

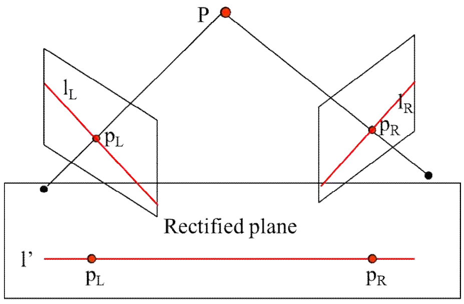


Figure 1. Image rectification in convergent camera array

Image rectification offers advantages that obtain accurate 3D information in a parallel camera array while it causes problems in an arc camera array. The reason is that both image planes transform to significantly different image plane from original images. This cause to distort the original images much. Fig. 2 shows the images after application of image rectification.

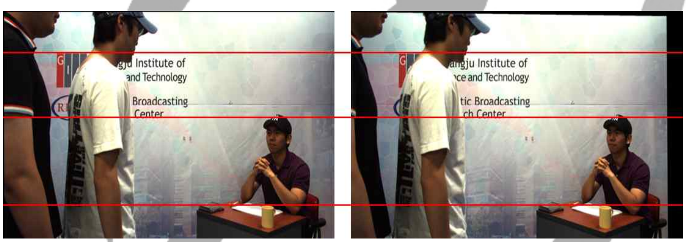


Figure 2. Rectified result of both images

Fig. 2 shows that visual distortion is generated by image rectification. In this situation, to acquire disparity information for whole pixels of an original image is difficult. In order fit the same size of an original image, the focal distance in camera parameter is controlled. Fig. 3 represents this result.

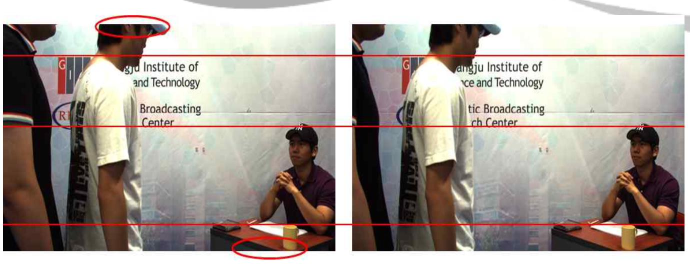


Figure 3. Rectified images with adjustment of focus distance

From fig. 3, some parts which exist in original images disappear in rectified images. The circled part in fig. 3 represents example of these parts. It is impossible to acquire 3D information by preserving original images of an arc array with image rectification. Thus, we propose a depth acquisition method without image rectification which distorts images.

III. Proposed Depth Estimation

A. *Epipolar Geometry for Stereo Images*

As we discussed before, the method that find corresponding points in same horizontal line through image rectification makes visual distortion and is unable to obtain disparity values in whole region. Even so, the work that scans not only horizontal direction but also vertical direction without any conditions is time consuming and reduces the accuracy. In order to reduce correspondence problem to one dimensional search, epipolar constraints can be used. Fig. 4 illustrates the method that finds corresponding points via epipolar constraint. For the stereo view located C_1 and C_2, we have 3D position P and two image points p_1 and p'_1, which are corresponding points on each image [10]. 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. Thus, the corresponding point of p_1 is located on the epipolar line in the right image plane [10].

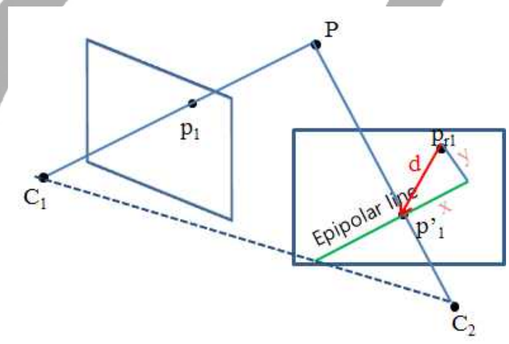


Figure 4. Epipolar constraint for stereo images

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

$$Z = \frac{f \times l}{d} \quad (1)$$

Z is the depth value, f is the focal length, l is the gap between cameras, and d represents the disparity value. However, to define disparity between corresponding points without rectification is difficult since the direction of disparity is not only horizontal but also vertical. Furthermore, disparity value may have negative number. Thus, in the proposed method, we present a direct depth acquisition method without depth transformation of disparity information since detection of corresponding points by epipolar constraints does not have one directional search because of omission of rectification.

B. Depth Map Acquirement Using Epipolar Constraints

We use a global method to extract depth values. For this, energy function about matching is defined by Markov Random Field (MRF), solved through an optimization technique. The processing speed of the global method, one of the stereo matching approaches, is relatively slow compared to that of the local method [11]. However, it provides high quality results. Since the accuracy is priorly considered, we adopt one of the global methods. We solve the problem of high complexity through the stereo matching method using hierarchical structure. The energy function for matching is defined as in

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

where x and y are horizontal and vertical coordinate of a reference image and d represents the depth value. Energy function, $E(x, y, d)$, is comprised of the date term and the smoothness term. The matching costs for data and smoothness terms are defined in

$$E_{data}(x,y,d) = U\{I_L(x,y), I_R(x,y,d)\} \qquad (3)$$

$$E_{smooth}(x,y,d) = \sum_{(p,q)\in N} W(p.q). \qquad (4)$$

$U\{\bullet\}$ is the pixel value difference between the left and right images, N is neighbors of pixels, and $W(p, q)$ represents disparity value difference of neighboring pixels. $I_L(x,y)$ is pixel value when coordinate is (x,y) in the left image. $I_R(x,y,d)$ is the matched pixel value in the right image when the coordinate is (x,y) and the depth value of the position is d in the left image. In order to find the matched pixel in $I_R(x,y,d)$, the 3D warping technique is used. The two processes of 3D warping are as follows. 3D warping backprojects the pixel of the left image to the 3D space based on the camera parameters and depth information. The projected pixel in the 3D space is then projected to the right image [12]. 3D warping is performed as follows.

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

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

$$(u',v') = (1/n, m/n), \qquad (7)$$

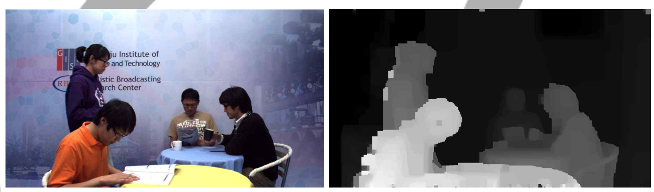
where A_{src}, R_{src}, and t_{src} are respectively the internal, rotation, and translation parameters in the left image. A_{dst}, R_{dst}, and t_{dst} are respectively internal, rotation, and translation parameters in the right image and $d_{u,v}$ is real depth value. The left image pixel is sent to the 3D space by (4) and is projected to the right image by (5). (u', v') in (6) represents the projected coordinate to the right image. Through above processes, $I_R(x, y, d)$ in (2) is determined and the energy function, $E(x, y, d)$, is optimized by Constance Space Belief Propagation (CSBP) [13] considering complexity. Finally, we apply post-processing to the acquired depth map to improve the quality of depth information [14].

IV. Experimental Results

We used stereo images captured from two kinds of arrays to evaluate the performance of our proposed method. Fig. 5 and fig. 6 show the original images and intermediate depth maps directly obtained by epipolar constraints in two kinds of camera array.

(a) Original image (b) Intermediate result

Figure 5. Result of direct depth acquisition using epipolar constraint in convergent array



(a) Original image (b) Intermediate result

Figure 6. Result of direct depth acquisition using epipolar constraint in parallel array

Figure 7 illustrates the visual comparison of the final depth map obtained from the proposed method with the depth map obtained from conventional method with image rectification. The conventional method generates the depth image which transforms disparity map obtained through rectification to depth map. The result in Fig. 7(b) is obtained by applying CSBP [13] and post-processing [14]. Since the upper part of the image in Fig. 7 (b) disappears by image rectification, the result in Fig. 7(b) is not accurate. On the other hand, the depth values of whole region are estimated in Fig. 7 (a).

Figure 8 compare the result of the proposed method with that of convention method with image rectification in the parallel camera array. Image rectification in a parallel camera array has considerable influence on the accuracy of depth map. However, the proposed method without image rectification generates the high quality result as well.

The proposed method enables us to generate accurate depth values without image rectification regardless of camera arrays. Therefore, our method has an effect that reduces restrictions in camera settings before capturing images.

(a) Depth map from proposed method

(b) Depth map using image rectification

Figure 7. Depth map comparison in convergent array

(a) Depth map from proposed method

(b) Depth map using image rectification

Figure 8. Depth map comparison in parallel array

V. CONSLUSION

In this paper, we propose the depth map acquisition method without any particular limitation in various stereo arrays. Stereo matching through image rectification generate the efficient results in aspects of accuracy and complexity. However, application of image rectification is not easy and produces image distortion in the convergent array. Thus, the proposed method leaves out transformation process of disparity information obtained through image rectification and presents the direct depth acquirement method via epipolar constraints. These processes increase the probability that obtain the depth information in whole region of the image. Furthermore, we improved the quality of the depth map through post-processing. Our method generated more stable results than the common conventional method in several arrays except parallel arrays. Therefore, our method can be usefully used for the production of 3D image.

ACKNOWLEDGMENT

This research is supported by MCST and KOCCA in the CT Research & Development Program 2012.

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] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3DTV," IEEE Trans. on Broadcasting, vol. 51, no. 2, pp. 191-199, June 2005.

[4] A. Frick, F. Kellner, B. Bartczak, and R. Koch, "Generation of 3DTV LDV-content with time of flight camera," IEEE International Conference on 3DTV, pp. 45-48, 2009.

[5] I. Y. Shin and Y.S. Ho, "Disparity estimation at virtual viewpoint for real-time intermediate view generation," International Conference on 3D Systems and Applications, paper S6-2, pp. 195-198, 2011.

[6] W.S. Jang and Y.S. Ho, "Efficient disparity map estimation using occlusion handling for various 3D multimedia applications," IEEE Transaction on Consumer Electronics, vol. 57, paper Issue 4, pp. 1937-1943, Nov. 2011.

[7] E.K. Lee and Y.S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-D video,"Journal of Visual Communication and Image Representation, Vol. 22, Issue 1, pp. 73-84, Jan. 2011.

[8] D. Sharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, 2001.

[9] Y.S. Kang and Y.S. Ho, "An efficient image rectification method for parallel multi-camera arrangement," IEEE Transactions on Consumer Electronics, vol. 57, paper no. 3, pp. 1041-1048, 2011.

[10] Y.S. Kang and Y.S. Ho, "Geometrical calibration of stereo images in convergent camera arrangement," IEEE-RIVF International Conference on Computing and Communication Technologies, pp. 44-47, 2012.

[11] H. Hirschmuller, P. R. Innocent, and J. M. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," International Journal of Computer Vision, vol. 47, no.1/2/3, pp. 229-246, Apr. 2002.

[12] Y.S. Kang and Y.S. Ho, "Disparity map generation for color image using a TOF depth camera," 3DTV Conference 2011, pp. 46(1-4), 2011.

[13] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.

[14] Q. Yang, C. Engels, and A. Akbarzadeh. "Near real-time stereo for weakly-textured scenes," British Machine Vision Conference, pp. 80–87, 2008.