Efficient Disparity Map Generation Using Stereo and Time-of-Flight Depth Cameras

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
{jws,hoyo}@gist.ac.kr

Abstract. Three-dimensional content (3D) creation has received a lot of attention due to numerous successes of 3D entertainment. Accurate estimation of depth information is necessary for efficient 3D content creation. In this paper, we propose a disparity map estimation method based on stereo correspondence. The proposed system utilizes depth and stereo camera sets. While the stereo set carries out disparity estimation, depth camera information is projected to left and right camera positions using 3D warping and upsampling is processed in accordance with the image size. The upsampled depth is used for obtaining disparity data of left and right positions. Finally, disparity data from each depth sensor are combined. The experimental results demonstrate that our method produces more accurate disparity maps compared to the conventional approaches which use stereo cameras and a single depth sensor.

Keywords: Depth estimation · Stereo matching · Tof depth camera

1 Introduction

With the huge success of three-dimensional (3D) movies, interest in 3D entertainment systems is increasing recently. 3D multimedia applications give audiences the opportunity to experience 3D. The 3D experience comes from the left and right eyes seeing different views. Thus, audiences can perceive 3D using two separate views for the left and right eyes. For this, several data formats have been proposed. The texture plus depth approach is one of these formats. This format uses an ordinary 2D image accompanied by a depth map. The non-existing view can be synthesized by depth image based rendering (DIBR) [1]. Benefits of this format are the flexibility to render views with variable baseline and the increased compressibility of depth data due to its characteristics [2]. Thus, this format is practical for many 3D multimedia applications.

Depth information can mainly be acquired by two approaches: active and passive sensor based depth estimation methods. The former employs physical sensors, such as infrared ray (IR), laser and light pattern, to measure depth data directly. Depth cameras, structured light sensors, and 3D scanners are employed in this approach [3]. Usually, active sensors are more effective than passive sensors in terms of the quality of produced depth images. However, they produce only low resolution images and generally require expensive devices. The latter, on the other hand, indirectly estimates depth information from 2D images captured by at least two cameras [4]. Stereo matching is

Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 623–631, 2015.
DOI: 10.1007/978-3-319-24078-7_64

the most widely used passive sensor based method [5]; its advantages are low cost and flexible resolution. However, passive sensors also obtain miscalculated depth information in several types of regions.

Disparity data can be converted into depth information by using a stereo image pair in combination with triangulation [6]. Disparity data can be acquired by finding the corresponding points in other images for pixels in one image. The correspondence problem is to compute the disparity map which is a set of the displacement vectors between the corresponding pixels. For this problem, two images of the same scene taken from different viewpoints are given and it is assumed that these images are rectified for simplicity and accuracy of the problem. From this assumption, corresponding points are found in same horizontal line of two images. A disparity map acquired by stereo matching can be represented by a gray scale image. Depth of each pixel is perceived from the disparity map. The object is close to viewpoint as intensity value of a pixel in the disparity map is high.

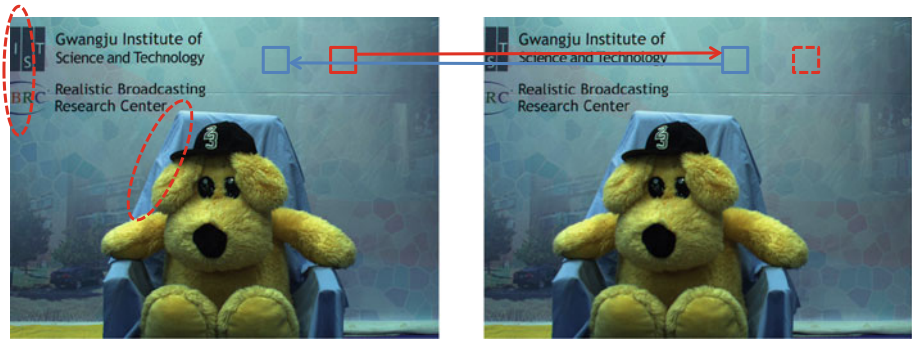
The objective of this paper is to obtain accurate depth information using depth and stereo images. Thus, we present an accurate disparity map acquisition method through stereo correspondence. We design a disparity estimation system to strengthen the merits and make up for the weaknesses for active and passive depth sensors. The proposed method deals with fully unsolved problems by fusing and refining the depth data.

2 Problem Statement

Over the past several decades, a variety of stereo-image-based depth estimation methods have been developed to obtain accurate depth information. However, accurate measurement of stereo correspondence from natural scene still remains problematic due to difficult correspondence matching in several regions: textureless, periodic texture, discontinuous depth, and occluded areas [7]. First, since color data of the textureless and periodic texture region in left and right images are so similar each other in a wide range, correspondence matching often fails because of its ambiguity. Second, in case of the depth discontinuous region, such as the edge region, smeared color values exist, which leads to ineffective correspondence matching. Lastly, in the occluded region, some pixels may appear in one image but not in the other image; So there is no corresponding pixel. These problems can be solved for accurate depth information. Figure 1 illustrates correspondence matching problem in stereo matching.

Usually, depth cameras are more effective in producing high quality depth information than the stereo-based estimation methods. However, depth camera sensors also suffer from inherent problems. Especially, they produce low resolution depth images due to challenging real-time distance measuring systems. Such a problem makes depth cameras not practical for various applications. Figure 2 represents resolution difference between the regular color camera and depth camera.

Recent approaches of fusing active and passive sensors have shown improvements on depth quality by making up for weakness of each sensor method [8, 9]. In our work, we propose a disparity fusion method to make up for weakness of stereo-based estimation and carry out better correspondence matching by adding a depth camera to the stereo system.



(a) Left image (b) Right image

Fig. 1. Correspondence problem in stereo matching

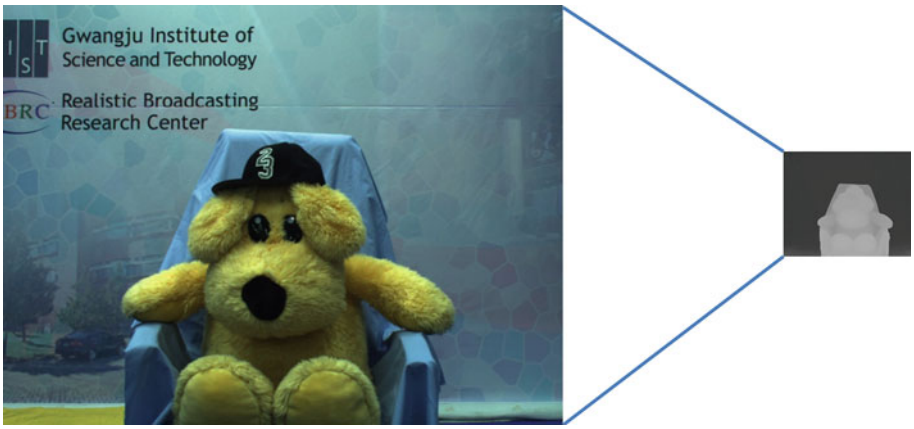


Fig. 2. Principle of Harris corner detector

3 Depth Fusion System

In general, stereo matching can be categorized into two approaches: local and global methods. Local methods are generally efficient for complexity [10]. However, these make blurred object borders and the removal of small detail at the depth discontinuity depending on the size of correlation window. In order to solve this problem, global methods have been proposed [11]. Global methods define an energy function by Markov Random Field (MRF) and optimize this function using several optimization algorithms such as belief propagation [12] and graph cut [13].

The proposed method is initially motivated by global stereo matching [14]. The information of the depth camera is included as a component of the global energy function to acquire more accurate and precise depth information. Depth camera processing is implemented as follows: (1) Depth data is warped to its corresponding position of the stereo views by 3D warping. 3D warping is composed of two processes.

First, the depth data is backprojected to the 3D space based on the camera parameters. Then, the backprojected data in the 3D space is projected to the target stereo views. 3D warping is performed as follows.

$$(x,\ y,\ z)^T = RsrcA_{src}^{-1}(u,\ v,\ 1)^T du, v + tsrc, \tag{1}$$

$$(l,\ m,\ n)^T = AdstR_{dst}^{-1}(x,\ y,\ z)^T - tdst, \tag{2}$$

$$(u',\ v') = (1/n,\ m/n) \tag{3}$$

where A_{src}, R_{src}, and t_{src} are intrinsic, rotation, and translation parameters in the depth camera data, respectively. Similarly A_{dst}, R_{dst}, and t_{dst} are those in the target color view. $d_{u,v}$ is depth value at (u,v) coordinate in the depth camera. The depth data is sent to the 3D space by Eq. (1) and projected to the target view by Eq. (2). (u', v') in (3) represents the projected coordinate to the target view. Figure 3 shows the 3D warping of depth data. (2) depth-disparity mapping is processed [15]. Due to the different representation of the actual range of the scene, correction of the depth information is required. (3) We perform joint bilateral upsampling (JBU) to interpolate the low resolution depth data. This method used high resolution color and low resolution depth image to increase the resolution of the depth image. Figure 4 illustrates JBU process. The processed information of the depth camera is applied as the additional evidence for data term of disparity fusion energy function.

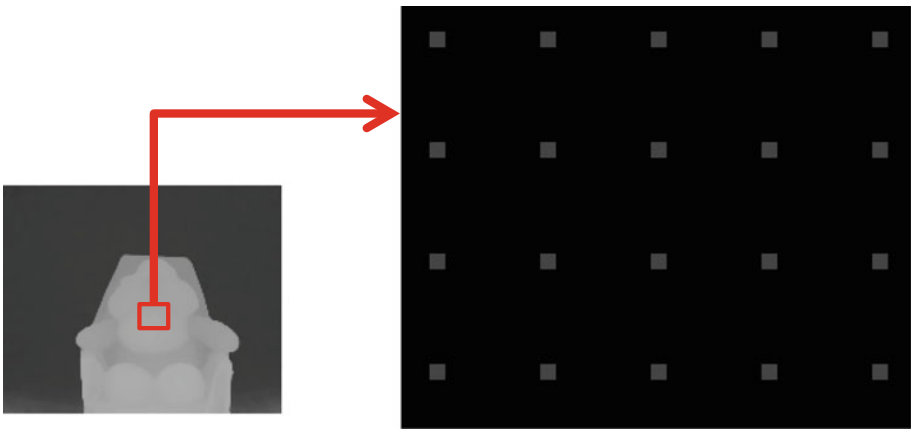


Fig. 3. 3D warping of depth camera data

Figure 5 illustrates error detection in stereo matching. To fuse both depth information, we consider error regions from both sensing. The error regions are obtained from the results of stereo matching. We applied the occlusion detection method using several constraints [7] to find the error regions.

Fig. 4. Joint bilateral upsampling process

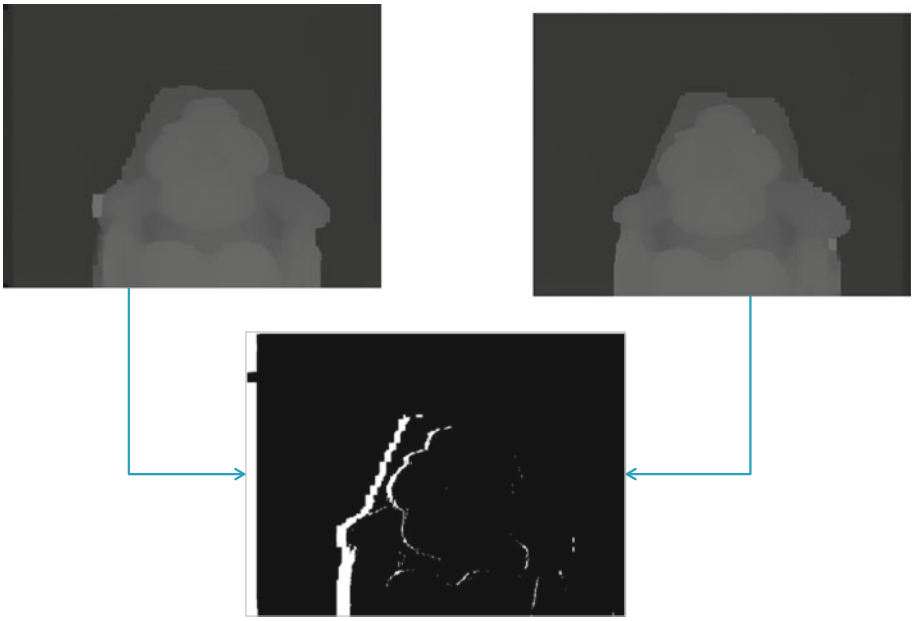


Fig. 5. Detection of error region in stereo matching

The disparity fusion energy function for accurate disparity estimation is defined as

$$E = \sum_{x,y} (Edata + Esmooth), \tag{4}$$

where E_{data} is a data term which measure the pixel similarity and E_{smooth} is the smoothness term which penalizes depth variations. The following function is applied to data term for energy function of depth estimation.

```
For all pixels
   If(JBU result ≠ hole && non-error region in stereo)
        Edata(x, y, d) = α | IL(x, y) − IR(x′, y′, d) | + β | d − dup |)
   end if
   else if(JBU result == hole && non-error region in stereo)
        Edata(x, y, d) = | IL(x, y) − IR(x′, y′, d) |
   end else if
   else if(JBU result ≠ hole && error region in stereo)
        Edata(x, y, d) = | d − dup |
   end else if
end for
```

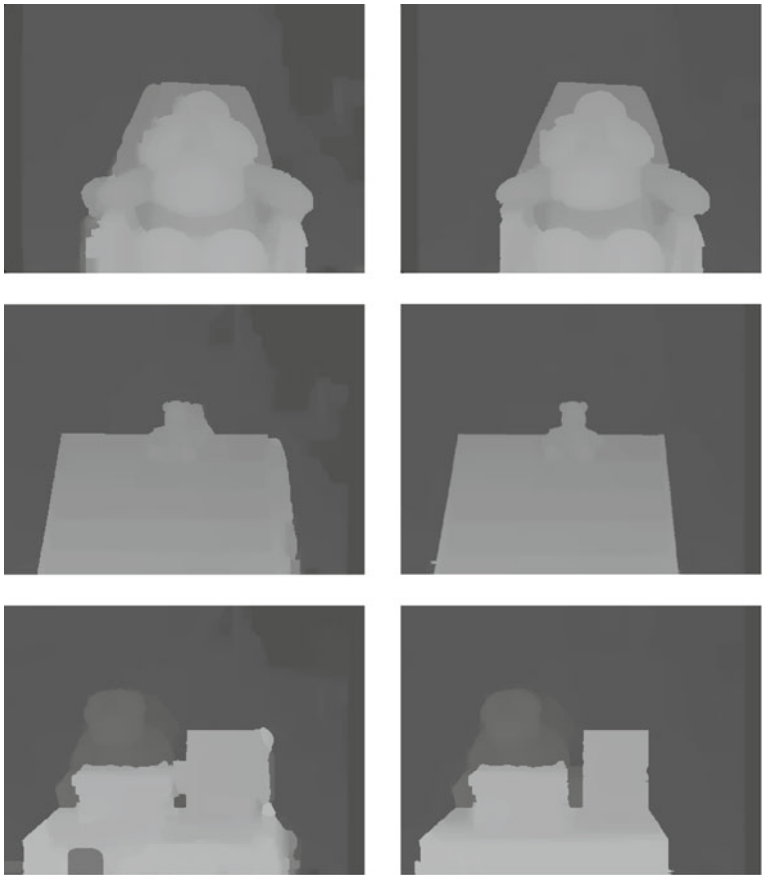
Fig. 6. Upsampling results of warped disparity

$I_L(x,y)$ is the pixel value in the left image given (x,y) coordinate. $I_R(x',y',d)$ is the matched pixel value in the right image given the disparity value at (x,y) in the left image, denoted by d. d_{up} is the upsampled disparity data obtained from the previous step. The upsampled disparity information enhances the precision and accuracy of the final depth values by allowing large depth variation.

The smoothness term is based on the degree of difference among the depth values of neighboring pixels.

$$Esmooth = \sum_{t \in N(x,y)} |d - d_t| \tag{5}$$

$N(x,y)$ represents the neighboring pixels of the current pixel. The algorithm is hierarchically processed to acquire more accurate depth values in the textureless region.



(a) Stereo matching (b) Proposed fusion method

Fig. 7. Comparison of Depth results

Furthermore, we apply post-processing to the acquired disparity map to improve the quality [16].

4 Experimental Results

In order to evaluate the performance of the proposed method, we compare our proposed method with the stereo-image-based and depth upsampling methods. They are captured at the resolution of 1280 × 960 pixels. The depth camera employs the resolution 176 × 144 pixels. Camera calibration was applied to these data sets for the accuracy of the processes. Figure 6 shows the original image and disparity results of JBU.

Figure 7 shows the results of stereo matching and proposed fusion method. The JBU results represent that boundary is not good, but we can obtain the precision disparity data, especially in regards to homogeneous region. On the other hand, the stereo matching method cannot produce disparity detail in homogeneous region. The results indicate that the proposed method outperforms other comparative methods. The visual comparison of the experimental results demonstrates that the proposed method can represent the disparity detail and improve the quality in the vulnerable areas of stereo matching.

5 Conclusions

This paper presents a novel stereo disparity estimation method exploiting a depth camera. The depth camera is used to supplement crude disparity results of stereo matching. The camera array is determined to reduce the inherent problems of each depth sensor. We fuse the both depth data considering confidence regions of each depth sensor. The fusion depth estimation which includes refinement process overcomes the weakness in each depth sensing. This increases the precision and accuracy of the final disparity values by allowing large disparity variation. Experimental results show that our method produce more accurate disparity maps compared to conventional stereo matching especially in occlusion and homogeneous regions.

Acknowledgment. This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea(ROK). [GK15C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

References

1. Zhang, L., Tam, W.J.: Stereoscopic image generation based on depth images for 3DTV. IEEE Trans. Broadcast. **51**(2), 191–199 (2005)
2. Tech, G., Muller, K., Wiegand, T.: Evaluation of view synthesis algorithms for mobile 3DTV. In: 3DTV Conference, pp. 132(1–4) (2011)
3. Lee, E.K., Ho, Y.S.: Generation of multi-view video using a fusion camera system for 3D displays. IEEE Trans. Consum. Electron. **56**(4), 2797–2805 (2010)

4. Jang, W.S., Ho, Y.S.: Efficient disparity map estimation using occlusion handling for various 3D multimedia applications. IEEE Trans. Consum. Electron. **57**(4), 1937–1943 (2011)
5. Scharstein, D., Szeliski, R., Zabih, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**, 7–42 (2002)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn, pp. 262–278. Cambridge University Press, Cambridge (2003)
7. Jang, W.S., Ho, Y.S.: Discontinuity preserving disparity estimation with occlusion handling. J. Vis. Commun. Image Represent. **25**(7), 1595–1603 (2014)
8. Lee, E.K., Ho, Y.S.: Generation of high-quality depth maps using hybrid camera system for 3-D video. J. Vis. Commun. Image Represent. **22**(1), 73–84 (2011)
9. Kang, Y.S., Ho, Y.S.: Generation of multi-view images using stereo and time-of-flight depth cameras. In: International Conference on Embedded Systems and Intelligent Technology, pp. 104–107 (2013)
10. Hirschmuller, H., Innocent, P.R., Garibaldi, J.M.: Real-time correlation-based stereo vision with reduced border errors. Int. J. Comput. Vis. **47**(1/2/3), 229–246 (2002)
11. Veksler, O.: Fast variable window for stereo correspondence using integral images. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. **1**, 556–561 (2003)
12. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. IEEE Trans. Pattern Anal. Mach. Intell. **25**(7), 787–800 (2003)
13. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)
14. Yang, Q., Wang, L., Ahuja, N.: A constant-space belief propagation algorithm for stereo matching. In: Computer Vision and Pattern Recognition, pp. 1458–1465 (2010)
15. Kang, Y.S., Ho, Y.S.: High-quality multi-view depth generation using multiple color and depth cameras. In: International Workshop on Hot Topics in 3D, pp. 1405–1408 (2010)
16. Yang, Q., Engels, C., Akbarzadeh, A.: Near real-time stereo for weakly-textured scenes. In: British Machine Vision Conference, pp. 80–87 (2008)