SEGMENT-BASED MULTI-VIEW DEPTH MAP ESTIMATION USING BELIEF PROPAGATION FROM DENSE MULTI-VIEW VIDEO

Sang-Beom Lee, Kwan-Jung Oh, and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST), Korea

ABSTRACT

In this paper, we propose a new depth map estimation algorithm based on image segments. We assume that the three-dimensional scene is composed of several non-overlapping planes in a depth space. After homogeneous color segments are determined by the image segmentation, we assign one depth value to one segment using 3D warping and segment-based matching technique. In the refinement process, we apply a segment-based belief propagation method to refine the initial depth map. Experimental results demonstrate that the refined depth map maintains object boundaries and contains proper depth values.

Index Terms—depth map estimation, multi-view video, image segmentation, belief propagation

1. INTRODUCTION

Owing to significant advancements in computing power, interactive computer graphics, immersive displays, and digital transmission, we experience and reproduce simulations of reality. Especially, technological advances in displays have been aimed at improving the range of vision, such as high-definition television (HDTV) or immersive displays.

A three-dimensional television (3DTV) using multi-view images is in the spotlight as one of next-generation broadcasting systems to provide the feeling of presence [1]. In order to obtain multi-view images, we use several cameras to capture wide-viewing angle scene and we are able to be immersed in the content as displaying captured multi-view images.

In general, there are two major problems in multi-view camera system. The first problem is the reliability of distance between cameras. The other problem happens when view is changed. When users change their views while watching contents in the 3D display, the flickering will occur on the display if the distance between cameras is large and the scene is changed suddenly. It causes a visual discomfort to viewers' eyes. To solve these problems, we reconstruct intermediate views. An intermediate view is an image captured from a virtual camera between real multi-view cameras. We can provide high quality 3D contents by inserting intermediate views so that the visual discomfort of the viewer is reduced.

In order to reconstruct intermediate images at virtual viewpoints, we need depth information. Many works have been carried out for the acquisition of 3D depth information. As one of the passive 3D depth sensing methods, stereo matching is well-known [2]. The task of stereo matching is the computation of the disparity for two input images. It is exactly what the human visual system is doing when we perceive depth. Since two images captured by our eyes are obtained from slightly different perspectives, the position of a scene point in one view is horizontally displaced in the other view. The amount of displacement allows reasoning about the depth of the scene point.

Recently, segment-based stereo matching algorithms are attracting issue because of their good performance. These methods assume that a scene has a set of non-overlapping planes in a disparity space and that these planes correspond to at least one homogeneous color segment obtained by image segmentation for the reference image. However, there still exist several problems such as wide-baseline and occlusion area which is visible in the reference view but invisible in the right (or left) view.

In order to solve the occlusion problem, we can exploit multi-view images as inputs of stereo matching algorithm to obtain more accurate depth map. In this method, we regard the center view as a reference view. Since the occlusion is occurred only in either left or right view, we can easily solve this problem by comparing the reference view with one view that doesn't have the occlusion. There are several algorithms using multi-view images, such as pixel-based PDE methods [3], graph cuts, [4] and volumetric methods [5].

In this paper, we propose a new depth map estimation method using multi-view video. After we perform image segmentation for the center image, we find an initial depth map using 3D warping technique to estimate the depth value directly. During this process, we apply segment-based matching using homogeneous color segment and assign one depth to one segment. In the final step we refine the initial depth map to remove erroneous areas. In addition, we adopt segment-based belief propagation.

978-1-4244-1755-1/08/$25.00 ©2008 IEEE 3DTV-CON'08, May 28-30, 2008, Istanbul, Turkey

2. STEREO MATCHING ALGORITHMS

2.1. Segment-based Stereo Matching

In the conventional stereo matching algorithms, it is observed that most errors are distributed near the disparity discontinuity, that is, object boundary. These errors make the ambiguous object boundary when we reconstruct the intermediate view. Furthermore, since most of viewers are sensitive to the edges of the image, the mismatch of the object boundary and disparity boundary causes a visual discomfort to viewers' eyes. However, the segment-based stereo matching can reduce the boundary mismatch. This approach assumes that the scene is composed of several non-overlapping planes in a disparity space and that each plane does not contain any object boundary.

Figure 1 shows the general block diagram of the segment-based stereo matching algorithms. After image segmentation is performed for the reference image, local window-based matching is used to estimate initial disparities. Squared intensity differences, absolute intensity differences, and cross-correlation are well-known matching functions. Next, the disparity plane is extracted by initial disparities and the homogeneous color segments. Finally, the disparity plane is approximated by using optimization methods such as graph cut or belief propagation.

2.2. Belief Propagation

The disparity map obtained by the initial disparity estimation has many erroneous areas. One of the problems is that the background of the image has monotonous texture and the segment in the background estimates the wrong disparity value. In order to refine the initial disparity map, many algorithms such as graph cut, dynamic programming, and belief propagation have been proposed.

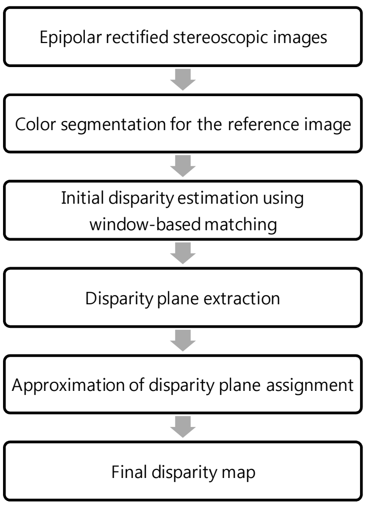


Fig. 1. Block diagram of the segment-based stereo matching

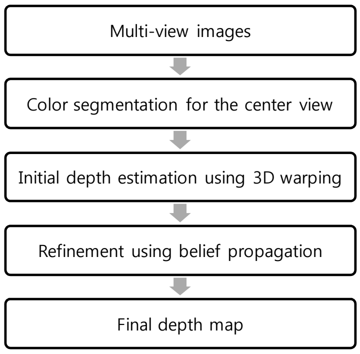


Fig. 2. Block diagram of the proposed scheme

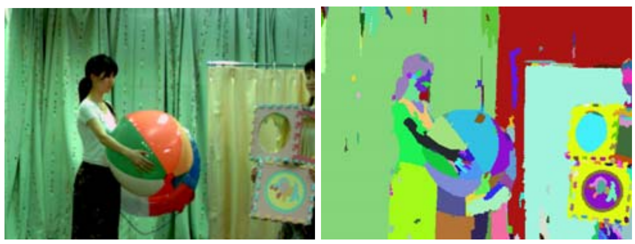
Recently, many stereo matching algorithms adopt belief propagation as an optimization process of disparity plane assignment. Belief propagation is an iterative inference algorithm that propagates messages in the network [6]. The basic idea is that we refine the initial disparity map by considering the neighbors' matching score iteratively.

3. PROPOSED MULTI-VIEW DEPTH ESTIMATION

In this section, we describe the proposed segment-based depth map estimation scheme. Figure 2 shows the block diagram of the proposed scheme. It is noticed that the whole procedure is similar to segment-based stereo matching. However, the remarkable difference is that the proposed scheme uses 3D warping and segment matching technique. In addition, we use the segment-based belief propagation.

3.1. Image Segmentation

We assume that the whole pixels in one segment have the same depth value. Since most depth discontinuity occurs near the object boundary, we also assume that each segment does not contain the object boundary. From these assumptions, we deduce that the better segmentation results guarantee the higher performance of depth map. In this paper, we employ a 'mean shift' image segmentation scheme [7] to segment images. Figure 3 shows segmented center view image for 'Akko&Kayo' (view 27). As shown in the figure, we confirm that the mean shift algorithm fix the accurate object boundary.



(a) Original image (b) Segmented image

Fig. 3. Segmented image for 'Akko&Kayo' (View 27)

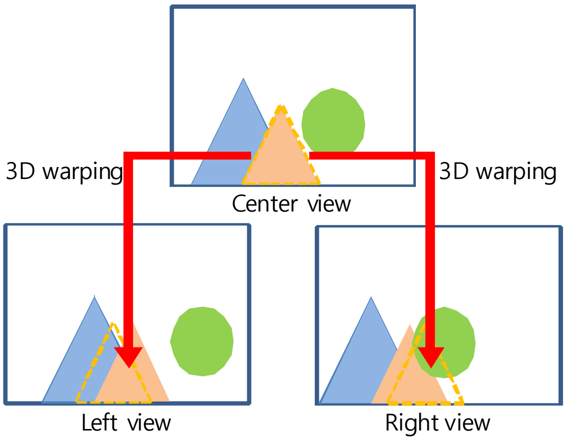


Fig. 4. Initial depth estimation using 3D warping

3.2. Initial Depth Map Estimation

After the center view image is segmented, we conduct initial depth map estimation for each segment. In order to directly determine the depth value, not the disparity value, we use the 3D warping technique. This approach does not need image rectification and disparity-to-depth conversion. By changing the depth value and calculating the matching score, we make up the initial depths when the score has minimum value. Moreover, we consider both left and right views simultaneously to resolve the occlusion problem.

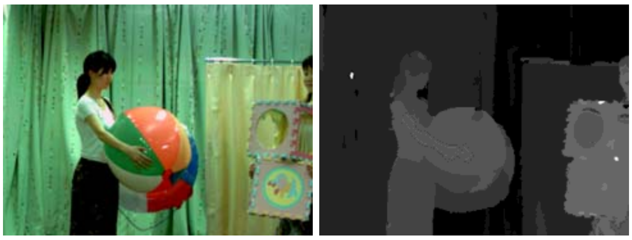
Figure 4 demonstrates the example for the initial depth estimation using 3D warping. The small triangle in the center view is warped to both left and right view to find the initial depth value. As shown in the Fig. 4, although the triangle is occluded by the circle in the right view, it is not occluded in the left view. In this case, we compare the center view with the left view. Therefore, we can easily solve the occlusion problem by using multi-view images.

As a matching function for the depth estimation, SD (squared intensity differences) and AD (absolute intensity differences) are well-known. However, these functions are not robust to illumination changes between cameras. In the proposed scheme, we use a self-adaptation dissimilarity measure [8] as a matching function. This function consists of the mean absolute gradient difference and the MAD. Since the gradient map represents the luminance change, not the absolute value, it is robust to absolute luminance mismatch between views. It is defined by

$$C(x,y,d) = (1-\omega)\times C_{MAD}(x,y,d) + \omega \times C_{MGRAD}(x,y,d) \quad (1)$$

where ω represents the weighting factor. The first and second terms of Eq. (1) are represented by

$$C_{MAD}(x,y,d) = \frac{1}{M}\sum_{(x,y)\in S_k}\left|I_1(x,y) - I_2(x',y')\right| \quad (2)$$



(a) Original image (b) Initial depth map

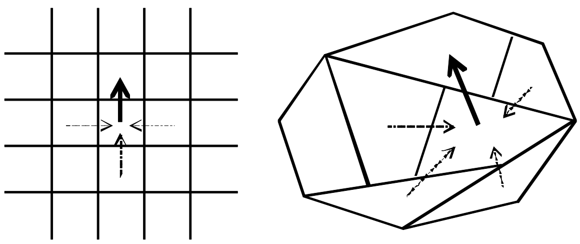
Fig. 5. Initial depth map for 'Akko&Kayo' (View 27)

$$\begin{aligned} C_{MGRAD}(x,y,d) = \frac{1}{M}\sum_{(x,y)\in S_k}\{ & \left|\nabla_x I_1(x,y) - \nabla_x I_2(x',y')\right| + \\ & \left|\nabla_y I_1(x,y) - \nabla_y I_2(x',y')\right| + \\ & \left|\nabla_{-x} I_1(x,y) - \nabla_{-x} I_2(x',y')\right| + \\ & \left|\nabla_{-y} I_1(x,y) - \nabla_{-y} I_2(x',y')\right|\} \end{aligned} \quad (3)$$

where M represents the number of pixels in the k-th segment S_k and (x',y') represents the position of the warped left or right view. ∇_x, ∇_y, ∇_{-x}, ∇_{-y} represents the gradient map in the $+x$, $+y$, $-x$, $-y$ direction, respectively. The matching score is calculated for each segment and this process assigns one depth value to one segment. Figure 5 shows the result of the initial depth estimation for 'Akko&Kayo' (View 27). As shown in Fig. 5(b), we notice that the initial depth map preserves the object boundary accurately but some erroneous areas exist in the background.

3.3. Refinement

The final step is a refinement of the initial depth map. The proposed scheme adopts the belief propagation as a refinement method. While the previous works using belief propagation are the pixel-based method, we utilize the segment-based approach in the refinement process. Figure 6 depicts the examples of belief propagation. As shown in Fig. 6(a), the pixel-based method used in the previous works refers to messages of 4-connected neighbor pixels. On the other hand, we refer to those of neighbor segments.



(a) Pixel-based method (b) Segment-based method

Fig. 6. Belief propagation

Fig. 7. Refined depth map for ‘Akko&Kayo’ (View 27)

4. EXPERIMENTAL RESULTS

4.1. Refined Depth Map

In order to evaluate the proposed scheme, we used the rectified ‘Akko&Kayo’ test sequence (View 26, 27, 28) provided by Nagoya University [9]. Figure 7 shows the final result of depth map estimation for ‘Akko&Kayo’ (View 27). From the refined depth map as shown in the Fig. 7, we noticed that many erroneous areas are removed in comparison with the initial depth map. In addition, the object boundary is still preserved after refinement.

4.2. View Synthesis Results

We tested the depth map by warping from the center view to left or right view and hole-filling technique. Figure 8 and Figure 9 show the view synthesis results for ‘Akko&Kayo’ from View 27 to View 26 and from View 27 to View 28, respectively. As shown in Fig. 8(b) and Fig. 9(b), we noticed that the refined depth map contains accurate depth values enough to synthesize the images.

(a) Original image (b) Synthesized image

Fig. 8. Synthesized image (view 27→26)

(a) Original image (b) Synthesized image

Fig. 9. Synthesized image (view 27→28)

5. CONCLUSIONS

In this paper, we described a segment-based multi-view depth map estimation scheme. The proposed scheme employed the multi-view-extended matching method to solve the occlusion problem. The 3D warping technique was used to find the depth value directly and the depth value for each segment was obtained by comparing the matching score obtained by the segment-based matching function. Finally, we efficiently removed the erroneous areas in the initial depth map by using segment-based belief propagation. From experimental results, we showed that the depth map generated by the proposed scheme efficiently preserves the object boundary and the depth values are correct.

ACKNOWLEDGMENT

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

REFERENCES

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, “3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards,” *IEEE International Conference on Multimedia and Expo*, pp. 2161-2164, July 2006.

[2] D. Sharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms,” Proc. of IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, Dec. 2001.

[3] C. Strecha, T. Tuytelaars and L. V. Gool, “Dense Matching of Multiple Wide-baseline Views,” *Proc. of International Conference on Computer Vision*, vol. 2, pp. 1194-1201, Oct. 2003.

[4] V. Kolmogorov and R. Zabih, “Multi-camera Scene Reconstruction via Graph Cuts,” *Lecture Notes in Computer Science*, vol. 2352, pp. 82-96, 2002.

[5] K. N. Kutulakos and S. M. Seitz, “A Theory of Shape by Space Carving,” *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199-218, 2000.

[6] P. F. Felzenszwalb and D. P. Hutenlocher, “Efficient Belief Propagation for Early Vision,” *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-261-I-268, July 2004.

[7] D. Comaniciu and P. Meer, “Mean shift: A Robust Approach toward Feature Space Analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603-619, May 2002.

[8] A. Klaus, M. Sormann, K. Karner, “Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure,” *Proc. of International Conference on Pattern Recognition*, vol. 3, pp. 15-18, 2006.

[9] M. Tanimoto, T. Fujii, T. Senoh, T. Aoki and Y.Sugihara, “Test Sequences with Different Camera Arrangements for Call for Proposals on Multiview Video Coding,” ISO/IEC JTC1/SC29/WG11, M12338, July 2005.