Disparity Refinement near the Object Boundaries for Virtual-View Quality Enhancement

Gyu-cheol Lee* and Jisang Yoo[†]

Abstract – Stereo matching algorithm is usually used to obtain a disparity map from a pair of images. However, the disparity map obtained by using stereo matching contains lots of noise and error regions. In this paper, we propose a virtual-view synthesis algorithm using disparity refinement in order to improve the quality of the synthesized image. First, the error region is detected by examining the consistency of the disparity maps. Then, motion information is acquired by applying optical flow to texture component of the image in order to improve the performance. Then, the occlusion region is found using optical flow on the texture component of the image in order to improve the performance of the optical flow. The refined disparity map is finally used for the synthesis of the virtual view image. The experimental results show that the proposed algorithm improves the quality of the generated virtual-view.

Keywords: Stereo matching, Optical flow, Occlusion regions, Texture component

1. Introduction

In contrast to the box office success of the 3D movies, 3D broadcasting, which has been labeled as the next generation broadcasting, has not yet been able to find its place in the market. According to the data from Retrevo, a market analysis agency, 55% of the consumers who plan on buying a HDTV feel no need for the 3D function because of the cumbersome task of wearing 3D glasses and the lack of content [1]. Also, the current 3D display method (stereo 3D: S3D) uses only one viewpoint to synthesize the 3D images so the realistic feeling and the vividness of the object lessens when seen from another viewpoint. The alternative to this stereo 3D image is the multi-view display technique that does not require 3D glasses. This alternative offers a more realistic viewing because it provides more viewpoints than the stereo display. Therefore, the viewer can enjoy the 3D image from any perspective [2].

There are many ways to obtain the multi view image but the simplest way is to install as many cameras as the viewpoints needed in order to obtain the images from each view [3]. However, this method lacks practicality because of the difficulty of calibrating the cameras and the high expenses of the cameras themselves.

Therefore, other methods of generating the multi-view image have been researched. The first alternative uses stereo matching algorithm in order to obtain the depth image from the stereo images [4]. The other alternative involves the use of the depth camera which allows the

color and depth image to be obtained simultaneously [5]. Because stereo matching is relatively robust to environment, it is easy to obtain the disparity map. However, it takes a long execution time and lacks in accuracy of depth image. Depth camera enables for the obtainment of depth images with high accuracy but has low resolution and high equipment costs.

In this paper, we propose the method of disparity refinement near the object boundaries for virtual-view quality enhancement. The disparity map is obtained using stereo matching. However, this results in noise near the object boundaries because of discontinuity. In order to improve the quality of the virtual view image, the disparity map refinement is necessary. First, the error region is detected by investigating the consistency between left and right disparity maps [6]. In the occlusion region, the region in the left image does not exist in the right image [7]. Therefore, error is prone to happen because of the difficulty in getting the disparity information.

This phenomenon creates a larger range of error as the resolution of the image becomes higher because the arithmetic operation gets more complicated [8]. The optical flow algorithm is applied to the textural component of the image, which has structural characteristics of the image, in order to extract occlusion region [9]. The texture component has the structural characteristics of the image. It shows good performance because light, noise, and shadow have been almost removed when extracting motion information.

The motion information of all the pixels is extracted by Lucas-Kanade method, which is a dense optical flow algorithm [10]. Using the extracted motion information, the consistency of the left and right image is investigated. Then, for the pixels that do not consistent, they are considered

[†] Corresponding Author: Dept. of Electronic Engineering, Kwang-woon University, Korea. (jsyoo@kw.ac.kr)

* Dept. of Electronic Engineering, Kwangwoon University, Korea. (gyucheol0116@gmail.com)

Received: August 11, 2014; Accepted: June 23, 2015

non-existent and defined as occlusion region. The error and occlusion region of the extracted left and right disparity maps are fused to create a new region. The new labeled regions are filled with the appropriate disparity values by using the joint bilateral filter that conserves object boundaries in the reference image [11]. Finally, refined disparity map is used to synthesize the virtual-view image by using bidirectional linear interpolation [12].

This paper is organized as follows. Section 2 explains the error regions detection of the disparity map. Section 3 describes the extraction of the occlusion region using the optical flow. The performance of the proposed algorithm is given through experiments in section 4. Finally, section 5 contains the conclusion.

2. The Error Regions Detection of the Disparity Map

Stereo matching algorithm can only work on the premise that the randomly picked pixel values of the left image exist on the right image. However, depending on the

viewpoint of the two cameras, the lighting and the amount of reflected light changes. That allows for the same points on the left and right image to have different pixel values. Additionally, if a certain region has identical pixel values, finding the pixel that corresponds with that region becomes difficult. Therefore, the possibility of extracting incorrect information is high. The same error occurs in the occlusion region which only exists in one image and does not exist in other regions. Thus, in order to improve the quality of the virtual-view image, the disparity map information has to be accurate. This section will explain how to detect the error region by investigating the consistency of the disparity map extracted using the stereo matching.

Fig. 1(c) and Fig. 1(d) shows the disparity map obtained using the stereo matching. There is a possibility of detecting the wrong disparity information because the fluctuation range of the disparity values near the object boundaries is high. As the resolution of the image gets higher, the arithmetic operations get more complicated and the range of error gets wider. The wrong disparity value can be detected by calculating the consistency of the left and right disparity map as shown in Eq. (1) and Eq. (2).

Fig. 1. The disparity map extracted using stereo matching: (a) Left color image; (b) Right color image; (c) Left disparity map; (d) Right disparity map

Fig. 2. The error region extraction: (a) The left disparity map; (b) The right disparity map

$$x^l + d^r(x^l + d^l(x^l)) + d^l(x^l) = x^r \quad (1)$$

$$c^l(x^l) = d^r(x^l + d^l(x^l)) + d^l(x^l) \quad (2)$$

where x^l and x^r are the coordinates of the left and right image. d^l and d^r represent the left and right disparity maps respectively. If $c(x)=0$, the disparity value of the corresponding coordinate is consistent. If $c(x) \neq 0$, the corresponding coordinate has the wrong disparity value.

Fig. 2 represents the image extracted from the error region of the disparity map. It shows that the error is mostly detected near the object boundaries.

3. Occlusion Extraction using Optical Flow

Optical flow is the method of tracing the motion within two frames. There is a sparse type which only traces the region with the noticeable properties as the object's boundary and a dense type which obtains the movement information of every pixel in the image.

In this paper, Lucas-Kanade method, which is a dense optical flow, is used to trace motion. After investigating the consistency of the extracted motion information from the left and right images, the pixels that do not match are regarded as the occlusion region which exists in the current image but does not exist in the other. Also optical flow is applied to the texture component of the image in order to improve the quality of the optical flow. Fig. 3 represents the flow chart of the proposed occlusion extraction.

Fig. 3. Block diagram of the occlusion extraction

3.1 Extraction of texture component

Generally, an image can be separated into the structure and texture components. The structure component represents the object's appearance, color and etc. Therefore, it also contains pixels and shadow that violate the brightness constancy. However, the texture component represents a measure for characteristics such as smoothness, roughness and regularity. Thus, the performance of the optical flow can be improved by using the image with the texture component.

Fig. 4. Extraction of texture component

The separation structure component from intensity image is accomplished using the method of Rudin, Osher and Fatemi [22] that removes noise exploiting total variation. For the intensity image $I(x)$, structure-texture separation is done by Eq. (3) and Eq. (4).

$$\min_{I_s} \int_{\Omega} \left\{ |\nabla I_s| + \frac{1}{2\theta} (I_s - I)^2 \right\} dx \quad (3)$$

$$I_T(x) = I(x) - I_s(x) \quad (4)$$

where $I(x)$ is the intensity image, $I_s(x)$ is the structure component and θ is a constant. ∇I_s represents the gradient variation of the structure component. The component that minimizes Eq. (3) is the solution of the structure component $I_s(x)$. $I_T(x)$ is the texture component and is calculated by finding the difference between the intensity image and its structure component as the expression of Eq. (4). Fig. 4 shows the image after it has been separated into the structure and texture components. In the texture image, it can be seen that neither the image's characteristic nor its shadow component can be found.

3.2 Occlusion detection

We use the Lucas-Kanade's optical flow method [10] which can estimate motion information between two images in order to determine the occlusion region. After investigating consistency of the motion information between left and right images, inconsistent pixels are defined as the occlusion region.

Optical flow assumes brightness constancy, but in the actual image, the brightness value of left and right image changes because of the camera sensor noise, the object's respective reflectance, and the shadow. For these reasons, the performance of the optical flow is not good. To determine more accurate motion information, we apply the optical flow to the textural part of the image. As a result, the performance of the optical flow could be improved.

Fig. 5 compares the extraction results of the occlusion region depending on whether the texture component is applied or not. The occlusion region is usually located in one side of object. In Fig. 5(a) and 5(b), however, the detected occlusion region that did not use texture

Fig. 5. Result depending on the use of texture image left image, texture nonuse; (b) right image, texture nonuse; (c) left image, texture use (d) right image, texture use

separation process exists anywhere indiscriminately. Fig. 5(c) and 5(d) represents the result based on texture separation. It is superior to that of the result that did not use texture separation process.

The addition of the occlusion region found using the above method and the detected error region of the disparity map is defined as the new error region.

3.3 Disparity map refinement

In this paper, the error regions are rectified by using a joint bilateral filter which fills the holes while preserving the boundaries of the reference image. The joint bilateral filter is defined as in Eq. (5) and Eq. (6).

Fig. 6. Disparity map refinement: (a) before processing and (b) after processing

$$D'_p = \frac{1}{W_p} \sum_{q \in s} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(I_p - I_q) D_q \quad (5)$$

$$W_p = \sum_{q \in s} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(I_p - I_q) \quad (6)$$

where D is the depth image and I is the intensity image and D'_p represents the pixel value generated by applying a joint bilateral filter to D and I . G is the Gaussian function and $\|p - q\|$ is the Euclidean distance between p and q . s is a set of neighboring pixels of p . σ_s and σ_r are parameters defining the size of neighborhood and W_p is the normalization constant. Fig. 6(a) is the disparity map obtained by using the stereo matching. Because of the disparity value error from the occlusion area and the boundary region, the image has blurring phenomenon or unclear shape. Fig. 6(b) is the rectified disparity map using the proposed method. This image shows that noise and the error region near the object boundaries are rectified

4. Experimental Results

To evaluate the performance of the proposed algorithm, we used “samgye” and “gyebeck”(MBC Drama) sequences with a size of 1920x1080 as test sequences. In order to detect the occlusion region, we used Lucas-Kanade method,

which is a dense optical flow algorithm, and window size is 5x5. A virtual-view image is simply synthesized by applying the bidirectional linear interpolation. The θ value when separating the texture component was set to 0.125 based on the experimental results.

Fig. 7 shows the 1st, 3rd, 5th, 7th, 9th, and 11th view images out of the eleven virtual-views synthesized by using the bidirectional linear interpolation with “Samgye” as the test sequence.

In Fig. 8, the generated virtual views by four different algorithms are shown. We compared the performance of the proposed algorithm with before processing [12], error region + JBF algorithm [6] and occlusion region + JBF algorithm.

Fig. 7. Virtual view-point image synthesized by using bidirectional linear interpolation: (a) (b) (c) (d) (e) (f) Results of 1st, 3rd, 5th, 7th, 9th, 11th view-points

Fig. 8. Performance comparison(image quality) : (a) before processing; (b) error region + JBF; (c) occlusion region + JBF; (d) the proposed method

As shown in Fig. 8(a), 8(b) and 8(c), especially near the cockscomb area, the quality of the virtual-view images synthesized by using the other algorithms are poor. In Fig. 8(d), it shows that the cockscomb looks more natural after the proposed algorithm has been applied. It can also show that the quality of the virtual view image improves when using the proposed algorithm by comparing Fig. 9(d) and others.

Fig. 9. Performance comparison(image quality): (a) before processing; (b) error region + JBF; (c) occlusion region + JBF (d) the proposed method

Fig. 10. Virtual view-point image synthesized by using bidirectional linear interpolation. (a) (b) (c) (d) (e) (f) Results of 1st, 3rd, 5th, 7th, 9th, 11th view-points

Fig. 10 shows the 1st, 3rd, 5th, 7th, 9th, and 11th viewpoint images out of the eleven synthesized virtual views synthesized by using bidirectional linear interpolation with MBC drama “Gyebeck” as the test sequence.

Fig. 11 shows the comparison of results of specific regions in order to test the performance of the proposed algorithm. When the existing disparity map is used without refinement, it is shown from the area near the people’s ears in Fig. 11(a) that distortion exists because of the error from

Fig. 11. Performance comparison(image quality): (a) before processing; (b) error region + JBF; (c) occlusion region + JBF (d) the proposed method

Fig. 12. Performance comparison(image quality): (a) before processing; (b) error region + JBF; (c) occlusion region + JBF; (d) the proposed method

the disparity map. When the proposed algorithm is used, it can be shown from Fig. 11(d) and 12(d) that the quality of the virtual-view image improves.

Table 1 shows the average PSNR of each method on Middlebury sequences (Tsukuba, Venus, Teddy and Cones) [23]. The PSNR of the proposed method is greater than other methods. It is shown that the disparity map is greatly refined by the proposed algorithm.

Table1. PSNR of each method on Middlebury sequence

Methods	PSNR
Before processing	30.23
Error region + JBF	31.25
Occlusion region + JBF	31.29
Proposed method	32.54

5. Conclusion

In this paper, we proposed a virtual-view synthesis algorithm using disparity refinement in order to improve the quality of the synthesized image. The disparity map obtained by using the stereo matching contains lots of noise and error regions. Those regions usually exist near the object boundaries and cause the quality of image to deteriorate when the virtual-view image is generated.

In the proposed algorithm, the error region is detected by investigating the consistency between the left and right disparity maps. Also, the texture component of the image that represents the structural characteristics is separated from image. Then, the optical flow algorithm is applied to the obtained texture component in order to extract motion information with high accuracy. After investigating consistency of the motion information between left and right images, inconsistent pixels are defined as the occlusion region. The error region is combined with the occlusion region to define a new region. Then, the joint bilateral filter is applied to this new region in order to acquire the appropriate disparity value. Finally, the virtual-view image is generated by applying bidirectional linear interpolation to the refined disparity map. Experimental results show that the quality of the virtual-view images using the proposed algorithm enhanced.

Acknowledgements

This work was supported by the ICT R&D program of MSIP/IITP. [B0101-15-1360, Loudness Based Broadcasting Loudness and stress Assessment of Indoor Environment Noises]

References

- [1] Retrevo Corporation, Could low interest in 3DTV

- hurt the TV business?, Retrieved Nov. 2011, from <http://www.retrevo.com/content/node/1915>
- [2] G. M. Um, G. H. Cheong, W. S. Cheong and N. H. Hur, "Technical development and standardization trends of multi-view 3D and free-viewpoint video," *The Magazine of the IEEK*, vol. 38, no. 2, pp. 18-23, Feb. 2011.
- [3] F. Yasutaka, J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision*, vol. 84, pp. 257-268, Sep. 2009.
- [4] T.J. Kim and J. S. Yoo, "Hierarchical stereo matching with color information," *The Journal of Korea Institute of Communications and Information Sciences*, vol. 34, no. 3, pp. 279-287, Mar. 2009.
- [5] J. M. Lim, G. M. Um, H. C. Shin, G. S. Lee, N. H. Hur and J. S. Yoo, "Multi-view image generation using grid-mesh based image domain warping and occlusion region information," *The Journal of Korean Society of Broadcast Engineers*, vol. 18, no. 6, pp. 859-871, Nov. 2013.
- [6] S. Y. Cho, I. S. Sun, J. M. Ha and H. Jeong, "Occlusion detection and filling in disparity map for multiple view synthesis," *8th International Conference on Computing and Networking Technology (ICCNT)*, pp. 425-432, Aug. 2012.
- [7] C. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 675-684, Jul. 2000.
- [8] Y. S. Kang and S. Y. Ho, "Generation of high-resolution disparity map using multiple cameras and low-resolution depth camera," *The Conference of Korea Institute of Communications and Information Sciences*, pp. 287-288, Nov. 2012.
- [9] J. Aujol, G. Gilboa, T. Chan and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *International Journal of Computer Vision*, vol. 67, no. 1, pp. 111-136, Apr. 2006.
- [10] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision (IJCAI)," *7th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 674-679, Apr. 1981.
- [11] L. Zhao and H. Wang, "Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial Domain," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2364-2369, Oct. 2009.
- [12] C. J. Park, J. H. Ko and E. S. Kim, "A new intermediate view reconstruction scheme based-on stereo image rectification algorithm," *The Journal of Korea Institute of Communications and Information Sciences*, vol. 29, no. 5C, pp. 632-641, May. 2004.
- [13] G. C. Lee, Y. H. Seo and J. S. Yoo, "GPGPU-based multiview synthesis using kinect depth image," *The Conference of Korea Institute of Communications and Information Sciences*, Yongpyong, Korea, Jan. 2012.
- [14] K. H. Oh, S. Y. Lim and H. I. Hahn, "Estimating the regularizing parameters for belief propagation based stereo matching algorithm," *The Journal of the Institute of Electronics Engineers of Korea*, vol. 47, no. 1, pp. 112-119, Jan. 2011.
- [15] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L1 optical flow," *In Statistical and Geometrical Approaches to Visual Motion Analysis*, vol. 5604, pp. 23-45, Jul. 2008.
- [16] M. S. Ko and J. S. Yoo, "Boundary noises removal and hole filling algorithm for virtual viewpoint image generation," *J. KICS*, vol. 37, no. 8, pp. 679-688, Aug. 2012.
- [17] W. Sun, O. Au, L. Xu, Y. Li and W. Hu, "Novel temporal domain hole filling based on background modeling for view synthesis," *IEEE International Conference on Image Processing 2012*, Florida, USA, Oct. 2012.
- [18] K. J. Oh, S. Yea, and Y.S. Ho, "Hole filling method using depth based inpainting for view synthesis in free viewpoint television and 3-d video," *Proc. of the 27th conference on Picture Coding Symposium (PCS'09)*, pp. 233-236, May. 2009.
- [19] J. H. Park and C. G. Song, "Effective shadow removal from aerial image of golf course to extract components," *The Journal of Korean Institute of Information Scientists and Engineers*, vol. 39, no. 7, pp. 577-582, Jul. 2012.
- [20] G. C. Lee and J. S. Yoo, "Real-time virtual-view image synthesis algorithm using Kinect camera," *The Journal of Korea Institute of Communications and Information Sciences*, vol. 38, no. 5, pp. 409-419, May. 2013.
- [21] N. E. Yang, Y. G. Kim and R. H. Hong, "Depth hole filling using the depth distribution of neighboring regions of depth holes in the Kinect sensor," *2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, Aug. 2012.
- [22] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *physica D: Nonlinear Phenomena*, vol. 60, pp. 259-268, Nov. 1992.
- [23] "<http://vision.middlebury.edu/stereo/data/>"

Gyu-cheol Lee He received the B.S. and M.S. degrees in electronics engineering from Kwangwoon University, Seoul, Korea in 2013 and 2015, respectively. He is currently a PhD student at Kwangwoon University. His research interests include stereo matching and 3D image processing and

object recognition.

Jisang Yoo He was born in Seoul, Korea in 1962. He received the B.S. and M.S. degrees from Seoul national university, Seoul, Korea in 1985 and 1987, all in electronics engineering, and Ph.D. degree from Purdue University, West Lafayette, IN, USA, in electrical engineering in 1993, respectively.

From september 1993 to august 1994, he worked as a senior research engineer in industrial electronics R&D center at Hyundai Electronics Industries Co., Ltd, ichon, Korea, in the area of image compression and HDTV. He is currently a professor with the department of electronics engineering, Kwangwoon University, Seoul, Korea. His research interests are in signal and image