Superpixel-Based Approach for View Synthesis Robust to Noise

Joo-Hyeon Kim[1], Tae-Yeon Kim[2], Hong-Chang Shin[3], Jong-Il Park[1]
[1]Department of Computer and Software, Hanyang University
[2]Division of Computer Science and Engineering, Hanyang University,
[1,2] 222, Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea
[3] Electronics and Telecommunications Research Institute
[3] 218, Gajeong-ro, Yuseong-gu, Daejeon 34129, Korea
E-mail : [1]jhkim@mr.hanyang.ac.kr, [2]teayounty@hanyang.ac.kr, [3]hcshin@etri.re.kr, [1]jipark@hanyang.ac.kr

Abstract: Depth image based rendering (DIBR) technique allows to create a virtual view image by using images and disparity maps captured from different viewpoints. In general, the quality of the synthesized image highly depends on the accuracy of the disparity maps; therefore, the synthesized image has poor quality if noise exists in some region of the disparity map. For solving this problem, we propose a robust method to synthesize a high quality image of virtual view despite of disparity noise. The reference images are segmented into superpixels, and we use the disparity value which is not regarded as noise among disparity values in each superpixel for DIBR in order to suppress the effect of noise. The experimental results shows the effectiveness of the proposed method.

1. Introduction

Despite the 3D imaging techniques make user feel immersive realism, users have been inconvenienced by wearing glasses in order to experience the video. Because of this inconvenience, a variety of studies about the auto-stereoscopic 3D displays have been conducted. Auto-stereoscopic 3D display allows users to view 3D image without glasses by attaching lenses or filters to the display devices. Since images from different viewpoints should be projected onto a single frame in the auto-stereoscopic display, the data size increases in proportion to the number of viewpoints; this yields the problems of processing and transmitting huge amount of data. These problems can be effectively solved by synthesizing images for most viewpoints virtually using images and depth maps from only a couple of viewpoints [1][2].

Synthesizing multiple viewpoint images from stereo images requires two techniques. The one is extracting 3D information, for example, disparity map, and the other is generating virtual viewpoint images from the stereo images and the extracted 3D information using view synthesis method such as depth image based rendering (DIBR) [3]. Theoretically, positions of the viewpoints to be synthesized can be accurately calculated when the accurate disparity maps are acquired from every reference viewpoint images, and this allows generating high-quality virtual viewpoint images. Unfortunately, virtual view synthesis method based on highly accurate disparity map may not be suitable for real-time purpose since in general, it takes too much time to create the highly accurate disparity map.

In order to generate high-quality virtual viewpoint images, we adopted superpixel-based approach. Our method is robust to disparity noise because only the disparity value which is not regarded as noise among disparity values in each superpixel is used for view synthesis, and superpixel-wise warping is performed. Our method is not the first superpixel-based approach; Chaurasia *et al.* Proposed view synthesis method based on superpixel for generating a virtual viewpoint using images taken by multiple cameras [4]. However, in contrast to their method, our method uses disparity values rather than camera parameters for warping superpixel to virtual viewpoint. Our method allows to synthesize virtual viewpoint images even if camera parameters are unknown and disparity maps are corrupted by noise.

corresponding author: Jong-Il Park

2. Proposed Algorithm

In this paper, aligned right and left view images and disparity map were used to synthesis the virtual view. The implementation consists of three steps. First, the image was split into the super pixels. Second, to synthesize virtual view image from reference view images, warping coordinate was calculated from pixel and disparity value in superpixel. Finally the virtual view was created between right and left view by blending warped view image.

2.1 Simple linear iterative clustering (SLIC)

Superpixel is a region that does not include the edge by clustering pixels with similar features in the image. Because superpixel has feature which classifies pixel with similar properties, it is used to preprocess for various vision algorithms for image processing and analysis such as image classification, image recognition, visual tracking, and so on. To date, many image segmentation methods using intensity and color space have been proposed [5]. SLIC algorithm can divide image rapidly and uniformly [6].

Since SLIC has weakness in splitting image incorrectly in similar color space, we additionaly used the disparity map to divide boundary of object more accurately. That is, we applied SLIC algorithm in six-dimension space for Lab color space, pixel coordinates, and disparity as follows:

$$D_{labd}=\sqrt{(1-\alpha)((l_k-l_i)^2+(a_k-a_i)^2+(b_k-b_i)^2)+\alpha(d_k-d_i)^2}\quad , \tag{1}$$

where l, a, b, and d represent the Lab color and the disparity value respectively, subscript k and i denote center of kth cluster and ith pixel respectively. The α represents the ratio between the color and the disparity; If this value is set as 0.5, the algorithm will be performed using same portion of color and disparity. In this paper, the α value was set as 0.5.

2.2 Mesh warping

We utilize a mesh structure which is used to calculate the warp of the coordinates, because all of the super-pixels are not created in the same size. Mesh is structured independently in each superpixel and generated to cover a region in superpixel (Figure 1). One mesh consists of eight triangles and each of triangle vertex is used to calculate warping coordinate (Figure 2). Calculating position of warped superpixel is equivalent to

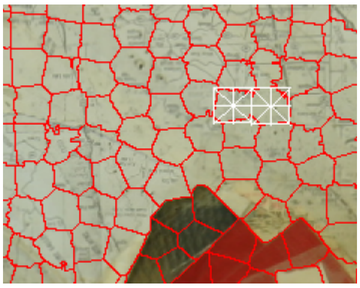


Figure 1. The mesh which is used as warp the superpixel.

figure out position of warped triangle in mesh. Thus we propose warping algorithm that calculates points of warped mesh using triangle vertices of mesh, barycentric coordinate of pixels in triangle and relations of pixel position between images of reference view and virtual view. The virtual view will be created with homography. First, we express the relationship for triangle and inner pixel position by the barycentric coordinate. Second, equation of coordinate relation is calculated for virtual view using inner pixel of triangle and the disparity value. The mesh is created as the structure of Figure 2 to include superpixel area while pixels in both superpixel area and triangle in mesh are expressed by

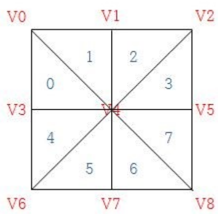


Figure 2. Structure of mesh

$$x = \alpha V_1 + \beta V_2 + \gamma V_3. \tag{2}$$

x is a pixel in both triangle and superpixel area. V_1, V_2, and V_3 are vertices of triangle and α, β, and γ are ratio of triangle area for pixel x. It could represented as matrix equation :

$$\mathbf{Ax} = \mathbf{b}. \tag{3}$$

A matrix has α, β, and γ value, **x** matrix is triangle vertex (V_1, V_2, and V_3) in mesh and **b** matrix is pixels in triangle. Pixel-based DIBR creates virtual view image using pixel and disparity [1]. The position of pixel in virtual view on the basis of left view image be expressed by

$$X_S = X_L + (1.0 - V_L) \cdot D_L. \tag{4}$$

X_S is virtual view and X_L is left image. V_L is virtual view position from left view. For example, the value will be 0.5 for creating the central view between right and left view. D_L is disparity value of X_L. The form of triangles of mesh is determined by disparity values. We could find the relationship for calculating the coordinates of the mesh that is warped by the Equation (3) and (5). Because we know where the pixels in the mesh are warped by a disparity value.

$$\mathbf{Ax}' = \mathbf{b}' \tag{5}$$

where **A** matrix has α, β, and γ value and x′is warped vertex of triangle. **b**′is position at virtual view of pixels in superpixel are warped by Equation (4). At this point, because unknown variable **x**′could calculated by using the singular value decomposition (SVD), it is possible to determine the warping mesh coordinates. If every disparity value in noise mixed region of superpixel are used, warped vertex position will be inaccuracy. To remove the noise added in the disparity map, we used not the every disparity value in the superpixel region but only used values which satisfied the threshold compared the disparity values between right and left view. In this case, disparity sample defines as the values that satisfy the conditions.

2.3 Blending

Blending is required for plausible synthesis from warped right and left image. The basic concept is as Figure 3. Blending step

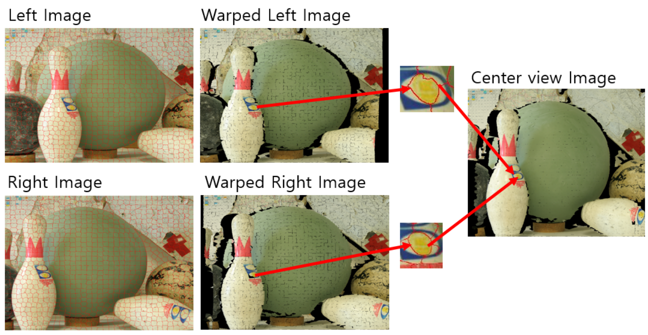


Figure 3. Concept of the proposed blending method

is composed of three steps. First we should calculate applied disparity value of pixel in warped superpixel. There may be some differences with disparity value of disparity map. Because warped position is calculated from non-noisy disparity values in area of super pixel. The position of corresponding pixel of the opposite view from reference view can be calculated by the epipolar geometry. We compare the difference of colors by the pixel referring disparity value which was found in the first step and corresponding position of the pixel in the opposite view derived from the positional relationship of virtual view image. Finally the virtual view is synthesized with the weight on the computed differences of colors after comparing right and left images. To warp view image we should find out homography relationship from the vertex of mesh and the vertex of warped mesh. It could calculated how much the pixel moves from the reference image by

$$P_R(x, y) = \mathbf{H}^{-1} \cdot P_N(x, y). \quad (6)$$

P_R (x, y) is position of the reference view image and P_N (x, y) is position of the virtual view of image. We can compute the relation of pixels by the homography matrix. we could calculate the value which is used to warp the virtual view to superpixel unit via Equation (6). Next, we should compare the pixel of opposite view image to the pixel of reference view image. It can be expressed as

$$D_R(x, y) = D_L(x, y) \cdot (1.0 - V_L)/V_L, \quad (7)$$

where V_L in Equation (7) is the position of virtual view generated from the left image, D_L(x, y) is the disparity value of the virtual view from left referring image calculated from Equation (6), and D_R(x, y) is a value computed by the virtual view positon and D_L(x, y) to refer to the right view. Once calculate the D_R at the opposite view, it is possible to calculate the position referring the right image as shown in

$$I_R(x, y) = P_R(x, y) + D_L(x, y) + D_R(x, y). \quad (8)$$

I_R indicates to the right view image (opposite view), and P_R means the reference view image (the left view image). we define the difference value of the left view as

$$Diff_L(x, y) = |I_R(x, y) - I_L(x, y)|, \quad (9)$$

where I_R(x, y) represents the right image and I_L(x, y) represents the left image. Thus, $Diff_L$ refers to the difference between the value of right and left view images.

This section was described on the basis of the left image, and the difference of the right image calculated by the same principle will be used in the synthesis stage. We used the weight value which was compared by the pixel of reference view and the pixel of opposite view. Execution condition is to run only if it is not in the hole, if the value is a case in the hole, we synthesized the opposite view of pixel. The weight was calculated as shown in

$$W_L = (E_R(x,y)+1.0) \cdot V_R / Diff_L(x,y) \cdot V_L + Diff_R(x,y) \cdot V_R + 1.0 \quad (10)$$

where 1.0 was added to handle if the difference of color was 0. The weight of opposite view can be expressed as

$$W_R = 1.0 - W_L. \quad (11)$$

The virtual view image is generated by combining the weight value. Warped right and left view images are synthesized through

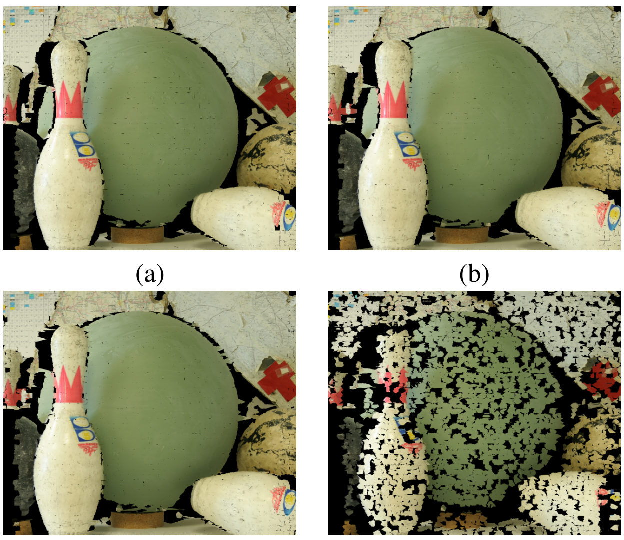
$$I_N(x, y) = W_L \cdot I_{WL}(x, y) + W_R \cdot I_{WR}(x, y). \quad (12)$$

I_N is the virtual view image, I_{WL} is warped left image, I_{WR} is warped right image, W_L is the weight value of the left image, and W_R is the weight value of the right image. Comparing two views that will be warped to the virtual view, we used the lower difference value, and the higher rate.

3. Experiments

Experiments in this paper use the Middlebury stereo data set[1]. Creation of noise is to have an incorrect disparity value by adding an arbitrary value (range: -100 $\sim$ +100) to the disparity value in a random position. Until the limited noise ratio (total pixels ratio number noise pixel), we produced a noise at non-overlapped position.

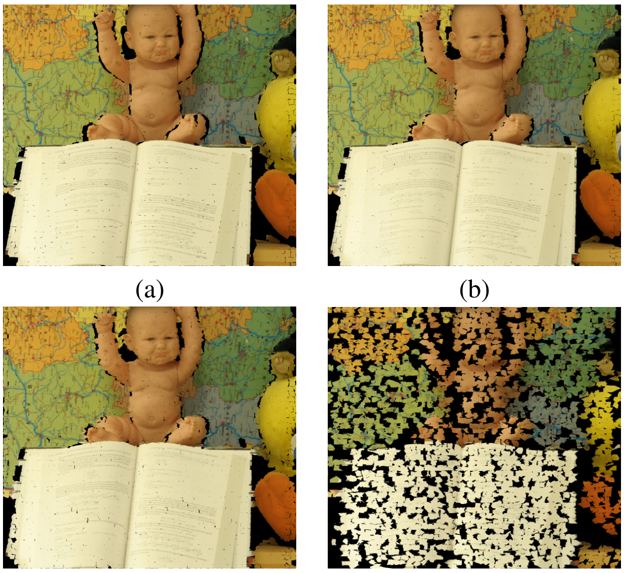
Image segmented into superpixel of approximate size in width 28 pixels and height 28 pixels. It requires less pixels for computing the warping coordinates of the super-pixel when the noise level is high. The virtual view image created by using noisy disparity map can be seen in (b), (c), and (d) of the Figure 4, and 5. The result of PSNR compared with a middle viewpoint using the proposed algorithm and a middle viewpoint of middlebury(Figure 6). The horizontal axis of graph represents the ratio of noise and the vertical axis of graph is the figure of PSNR. There was little difference in synthesized images that represents virtual view in noiseless disparity map and noisy (25%, 50%) disparity map. The PSNR represented the similar results. The synthesis result of a disparity map including 75% noise could not be produced virtual view image because amount of disparity values were not enough to calculate coordinate of warped mesh.



(a) (b) (c) (d)

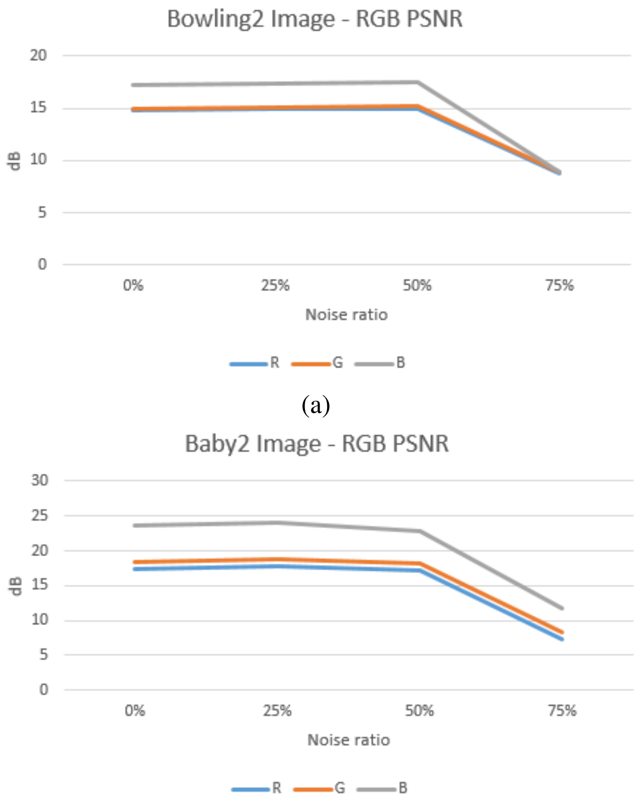
Figure 4. Synthesized virtual view images of Bowling data using disparity maps with different noise rates: (a) 0%, (b) 25%, (c) 50%, (d) 75%.

[1] http://vision.middlebury.edu/stereo/data/



(a) (b) (c) (d)

Figure 5. Synthesized virtual view images of Baby data using disparity maps with different noise rates: (a) 0%, (b) 25%, (c) 50%, (d) 75%.



(a) (b)

Figure 6. PSNR of the result images synthesized using disparity maps with different noise rates: (a) Bowling data, (b) Baby data.

4. Conclusion

In this paper, we proposed a method for virtual view synthesis, which is robust to disparity noise. We segmented aligned right and left view images using SLIC algorithm and synthesized virtual view image by superpixel-wise warping and blending methods. In this process, we could calculate the warping coordinates without affection of noise by using only the disparity samples. Consequently, we showed that the proposed method is robust to noise, via experimental results. As future work, we are studying about the view synthesis techniques using less disparity sample and superpixel-based hole filling technique.

Acknowledgement

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2016-H8601-16-1005) supervised by the IITP (Institute for Information & communications Technology Promotion)

References

[1] J. Park, J.Y. Choi, I. Ryu, J.I. Park. ”Universal View Synthesis Unit for Glassless 3DTV,” *IEEE Transactions on Consumer Electronics*, vol.58, no.2, pp. 706–711, May 2012.

[2] H.-C. Shin, Y.J. Kim, H. Park, J.I. Park. “Fast view synthesis using GPU for 3D display,” *IEEE Trans. Consumer Electronics*, vol.54, no.4, pp.2068–2076, Nov. 2008.

[3] H.-C. Shin, J. Lee, G. Lee, N. Hur. ”Stereo-To-Multiview Conversion System Using FPGA and GPU Device.” *Journal of Broadcast Engineering* vol.19 no.5 pp.616–626, 2014.

[4] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, G Drettakis. “Depth synthesis and local warps for plausible image-based navigation” *ACM Transactions on Graphics* vol.32 no.30 pp.30, Mar 2013.

[5] F. Ge, S. Wang, T. Liu. “New Benchmark Image Segmentation Evaluation,” *Journal of Electronic Imaging*, vol.16, no.3, 033011, July 2007.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, Süsstrunk. ”SLIC Superpixels” *EPFL Technical Report*, no.149300, 2010.