INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2008/M15597
July 2008, Hannover, Germany

Source: GIST (Gwangju Institute of Science and Technology)
Status: Proposal
Title: Boundary Filtering on Synthesized Images for 3D Video
Author: Cheon Lee and Yo-Sung Ho

1. Introduction

This document describes the extracting method of boundary noise on view synthesis caused by inaccurate depth values. Since the depth maps for every viewpoint are provided as input, we employ 3D warping technique to reconstruct a scene at a desired viewpoint. Hence, the accuracy of depth map directly affects the quality of the synthesized image. It can be obtained by using stereo matching algorithm as a passive sensing method. Although there are many algorithms of stereo matching, the performance is still unstable especially at the disparity discontinuity region. This instability will create boundary noise around the boundary of objects. In this document, we propose a method which extracts boundary noise efficiently.

2. View Synthesis Method

2.1. Scene Reconstruction Method for Arbitrary Viewpoint

Techniques for generating an arbitrary viewpoint image has been researched with wide approaches starting with the plenoptic function. This function can give us the perfect description of a 3D scene, but it has high complexity. To reduce the complexity, many approaches were proposed. Among them, 3D warping is a convenient method since it needs only depth information. In 3DV, view synthesis method can exploit 3D warping because the depth maps are provided as the input data [1][2].

We explain the 3D warping technique briefly. If the multi-view cameras are calibrated, we can use camera parameters to map an arbitrary point $\tilde{\mathbf{M}}$ which is located in the world coordinates into the position of the camera coordinates using Eq. (1).

$$\tilde{\mathbf{m}} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]\tilde{\mathbf{M}} \tag{1}$$

where $\mathbf{A}$, $\mathbf{R}$, and $\mathbf{t}$ denote the intrinsic matrix, rotation matrix, and translation vector, respectively. $\tilde{\mathbf{M}} = [X \quad Y \quad Z \quad 1]^T$, $\tilde{\mathbf{m}} = [x \quad y \quad 1]^T$ are the homogeneous vectors of the world and camera coordinates. When we try to find the corresponding position $\mathbf{m_r}$ in the reference image, we have to put it back into the world coordinates using Eq. (2). After this backward projection, we project again into the target virtual camera coordinates using Eq. (2).

$$\mathbf{M_r} = \mathbf{R_r}^{-1} \cdot \mathbf{A_r}^{-1} \cdot \mathbf{m_r} \cdot \mathbf{d(m_r)} - \mathbf{R_r}^{-1} \cdot \mathbf{t_r} \tag{2}$$

We may have some problems when we use the depth map in view synthesis. One of them is the depth error. It is obvious that the accuracy of depth map is directly related to the quality of the synthesized image. To get a good depth map, we can use any method among the stereo matching algorithms. Although there are many good methods, the performance is still unstable around the disparity discontinuity region. This creates undesired fraction noise which we want to deal with it in this proposal.

Another problem on view synthesis is the hole area. Predicting texture information of the disoccluded area is impossible when we use only one reference image. Fortunately, 3DV allows us to use multi-view reference images and multi-view depth maps. Hence we can find the corresponding information for the disoccluded area from the other view images. In the following subsection, we explain about the simple hole filling method and boundary fraction noise filtering.

2.2. Hole Filling

In this subsection, we explain our proposal which is the general procedure of hole filling method in multi-view video system.

The hole detection method determines the newly exposed area caused by 3D warping as shown in Fig. 1. The block area in Figure 1(a) is the hole region exposed by referring to the left view image. Since the hole region is created by changing viewpoint, there is no information at the reference view. The visible area is the commonly exposed area among reference views. Therefore, we can divide the synthesized image into hole area and common visible area. Later, we will deal with this distinguished hole area to determine the background boundary.

The hole region is placed at the right side of the objects when we warp the left reference image into right virtual viewpoint as shown in Fig. 1(a). The corresponding texture information of the hole region can be found at the right reference view. We can fill in the hole of Fig 1(a) by copying the textures from Fig. 1(b).

(a) Holes by Referring Left View (b) Holes from Referring Right View

Fig. 1. Hole Region Created by Warping

2.3. Boundary Noise

View synthesis using 3D warping is highly dependent on accuracy of depth map. As we mentioned in subsection 2.1, unstable discontinuities around boundary creates boundary fraction noise. As shown in Fig. 2, fractions of the thumb are located on the background. It is because of the inaccurate depth value around the object boundary region. This boundary fraction noise is the problem that we want to deal with.

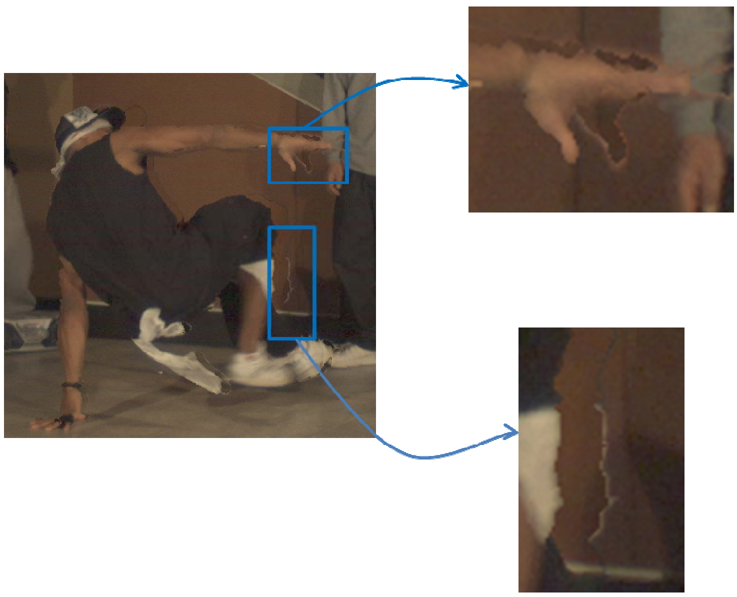


Fig. 2. Boundary Fraction Noise Cause by Depth Error

3. Boundary Noise Filtering Method

From here, we explain about the method of boundary noise filtering. The noise area is located near the boundary of objects having hole. After detecting the hole area, we choose the boundaries of background region. Then, we find the alternative texture information at the other reference view. Finally, we replace the texture information with the corresponding texture data.

Fig. 3. Process of Boundary Noise Extraction

3.1. Detecting the Boundary Contour

To extract the background boundary noise, we use the hole area in Fig. 1. White contours in Fig. 4 indicate the boundary of hole area. We determine this contour by detecting the hole area. In Figure 4, the left side of the hole is the boundary of foreground. In the same manner, the right side is the background. Of course, the boundary noise can be existed at the foreground, but we do not think that it lowers the quality of the synthesized image. In the following process, we divide the contour into background and foreground contours.

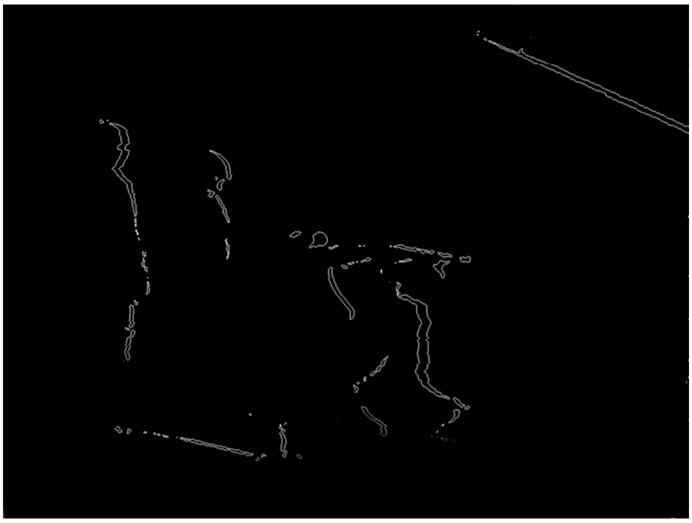
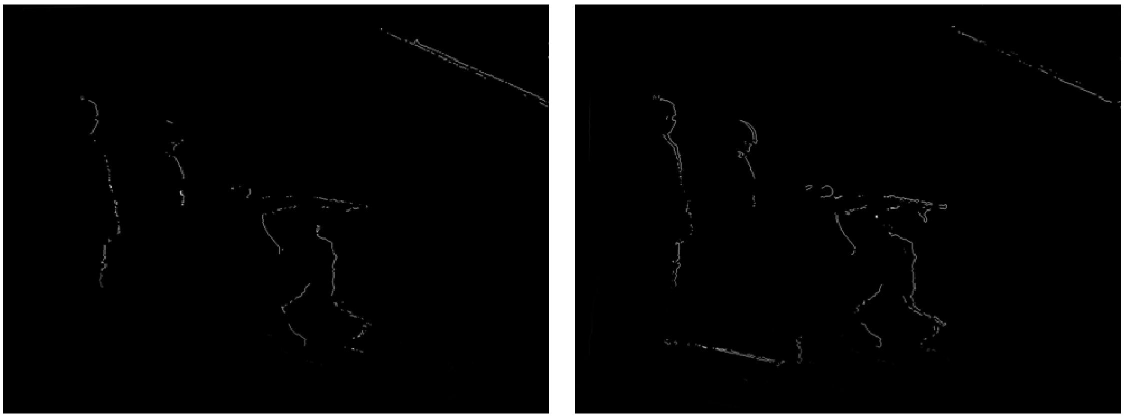


Fig. 4. Contour of Hole Region

3.2. Extracting the Boundary Contour

Since our objective is to process the boundary area, we delete the contours indicating the boundary of foreground. We can detect the background boundary by checking the depth values around the contour. Figure 5 shows the divided contours describing the foreground and background. Figure 5(b) indicates the background area that we want to process.



(a) Foreground Contour (b) Background Contour

Fig. 5. Background/Foreground Division

3.3. Boundary Noise Filtering

We find the corresponding information at the other reference image. Since we are using multi-view video, the alternative texture information is located at the other reference view. We can find the texture data along the background contour as an alternative. Figure 6(a) is the synthesized image referring to the right image. From the contours of Fig. 5(b), we copy the textures that does not have any noise. For the final step, we replace the boundary noise with the copied information. The filtered image is described in Fig. 7. The boundary fraction noise of Fig. 7(a) has been extracted clearly.

(a) Synthesized Image using Right View (b) Copied Image using Background Contour

Fig. 6. Alternative Image for Background Noise Region

(a) Boundary Noise (b) After Filtering

Fig. 7. Result of Background Noise Extraction

4. Experimental Results

We have experimented on ‘breakdancers’ sequences since its multi-view depth maps are available. We choose two reference images and synthesized the intermediate view image as illustrated in Figure 8. Total number of views is 6 and the result images are demonstrated in Fig. 9. Left images are the synthesized results, and right images are the original images. We can see the viewpoints are the same with the original images.

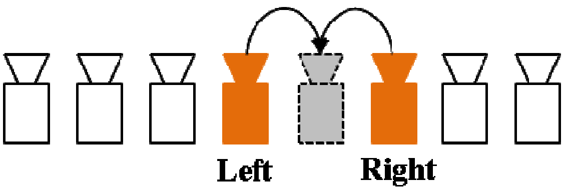


Fig. 8. Intermediate View Synthesis using Neighboring View Images

(a) Synthesized image (left) and original image (right) of ‘breakdancers_1’

(b) Synthesized image(left) and original image(right) of ‘breakdancers_2’

(c) Synthesized image(left) and original image(right) of ‘breakdancers_3’

(d) Synthesized image(left) and original image(right) of ‘breakdancers_4’

(e) Synthesized image(left) and original image(right) of ‘breakdancers_5’

(f) Synthesized image(left) and original image(right) of ‘breakdancers_5’

Fig. 9. Alternative Image for Background Noise Region

5. Conclusion

We described boundary filtering method on view synthesis for 3DV. The boundary fraction noise is caused by the erroneous value of depth map around boundary. Assuming that the depth map is untouchable, we extract boundary noise existing around the object boundary. We determined a contour map indicating the background region, and replace it with the corresponding texture information referring to the other viewpoint image. By experiments, we obtained synthesized images that does not have any boundary noise.

6. Acknowledgements

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

7. References

[1] ISO/IEC JTC1/SC29/WG11 “Preliminary FTV Model and Requirements,” N9168, July 2007.
[2] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N9783, May 2008.
[3] http://vision.middlebury.edu/stereo/eval/