H-030

奥行きマップ安定化のための前景検出方法
Foreground Detection Method for Depth Map Stabilization

パナヒプル テヘラニ メヒルダド、石川 彰夫、河北 真宏、井ノ上 直己† 藤井 俊彰‡
Mehrdad Panahpour Tehrani, Akio Ishikawa, Masashiro Kawakita, Naomi Inoue Toshiaki Fujii

1. Introduction

In the recent years, 3D video attracted many attentions and applications such as 3DTV [1] and FTV (Free-viewpoint TV) [2]. In such a system, if virtual views are needed, DIBR (Depth Image Based Rendering) method [3] is commonly used. This method requires multiview video plus depth maps as input data. For multiview video, depth maps are generally estimated by stereo matching [4] and optimization, such as graph cuts [5]. Looking at the depth map in different frames indicates that even for the objects, where there is no movement, the depth value varies frame to frame, due to luminance change and noises in the images. In order to address this problem several methods have been proposed that can be categorized in two approaches.

The first approach tries to stabilize the depth map at each frame individually, by eliminating noise before stereo matching [6], or in-cooperating the estimated depth map in previous frame as initial value for estimation of depth map of a current frame [7], as it is provided in DERS by Nagoya University. Our method falls in the second approach, where the multiview video contains the static areas images, so that depth map for the static areas can be estimated once, in advance. Given the mask and depth map of the foreground objects in the following frames, the depth map of the current frame is generated by merging the foreground and background depths. In this case, the processing time is reduced and the depth map in the occluded area is more accurate. This type of work has been recently proposed in [8]. However, this method did not investigate a suitable detection method, and only perform the experiments on computer generated sequences, where the detection of the foreground object is not challenging.

In this paper, we propose a foreground detection algorithm that deals with actually captured scene. The generated masks by this method are suitable to reconstruct a stable depth map in time domain. By stabilizing the depth maps, virtual images synthesized in time domain are also stabilized and their quality improved, in comparison with the quality of virtual view synthesized by the depth maps that are generated for each frame individually. Furthermore, we report the improvement in compression efficiency of the stabilized depth maps in comparison with the depth maps that are generated originally for each frame individually.

2. Foreground Detection Method and Generating the Stabilized Depth Map

In this section we aim to detect foreground objects that satisfy the stability of the depth map in time domain.

† 情報通信研究機構、超臨場感映像研究室
‡ 名古屋大学、工学研究科

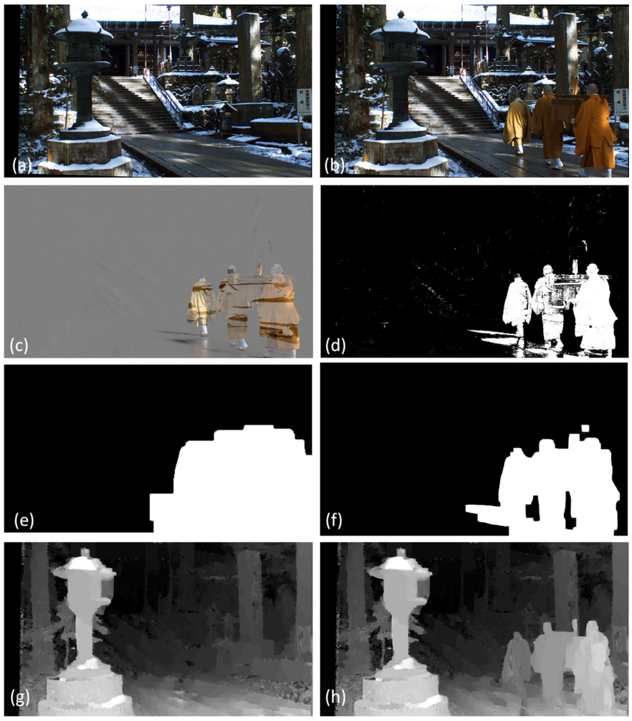

Fig. 1 (a) Background frame, (b) Foreground frame, (c) Subtraction of foreground and background, frames (d) Binarization of subtraction, (e) Coarse mask detection, (f) Fine mask detection, (g) Background frame depth map, and (h) Merged background and foreground depth maps.

The detected mask, at least must cover all pixels of the foreground object, and exclude all other areas, as much as possible. In other word, the detected foreground mask shall be equal or bigger than the area were the foreground object is located. Therefore the mask does not need to be very precisely matched to the foreground object shape.

The detection contains two steps. First step detects a large mask that covers the foreground object, and excludes the wrongly detected areas. In the second step, the large mask of the object is further brush up and more accurate mask of the foreground object is reproduced.

Figure 1 shows example results of the proposed method in each step. Details of each step are explained in the following.

2.1 Detection of a coarse foreground mask

The main purpose of this step is to provide a coarse mask. The coarse mask must include all area of the moving object, and exclude the wrongly detected areas as foreground, as follows:

(i) Gaussian filtering of the background frame (BG), and foreground frame (FG), to eliminate the noises and reduce the detailed illumination variations. (ii) Subtracting of the filtered FG and BG, followed by thresholding to generate a mask, i.e.

binary image. (iii) Downsampling the mask in (ii) by order of N, using bilinear interpolation. This step further deemphasizes the area where wrongly detected as mask. Note that after this process, the mask is no longer a binary image. (iv) Detection Process: It contains several morphological and median filtering as follows: (Erosion+Median Filter) n times, (Dilation+Median Filter) m time, and (Erosion+Median Filter) p times. This process excludes the wrongly detected mask and generates a larger mask that covers all areas of the foreground objects. (v) Upsampling the output of process in (iv) by the order of N, followed by binarization.

2.2 Detection of a fine foreground mask

The main purpose of this step is to provide a fine mask from the coarse mask in 2.1. Our proposed detection algorithm in this step is exactly the same as "(iv) Detection Process" on the area that are detected as coarse mask. After the detection process, the binarization process is applied to generate the final fine mask.

2.3 Generating the stabilized depth map

Given the multiview images of the background scene, the depth map of the background is generated. Given the foreground and background frames, mask is generated as explained in section 2.1 and 2.2. Given the multiview images of the foreground scene, and the mask, the depth map is generated in the area of the mask. Depth map for the rest of areas in the foreground frame are copied from the background depth.

3. Experimental Results

In this section, we evaluate the effectiveness of the proposed foreground detection method for depth map stabilization on, quality of the synthesized virtual view, and compression efficiency of the coded depth maps.

Example of the individually estimated depth maps and synthesized virtual view are shown in figure 2. Figure 3 shows stabilized depth map and stabilized synthesized virtual view using our method. As it can be seen in figure 3, the depth map, and consequently the synthesized image are more stabilized than that of shown in figure 2 for individually estimated depth maps and synthesized images. Table 1 shows the compression performance of the coded 160 frames, 30fps, 2-view depth maps by MVC, GOP 16 and IPPP. The coding gain is larger for the compression of the stabilized depth map than compression of the individually detected depth maps.

4. Conclusions

In this paper, we proposed a foreground detection method that can be used for stabilizing the estimated depth maps in multiview video. The estimated depth map using our method improves the quality of the synthesized virtual image in time domain. Furthermore, it improves the coding efficiency of the coded depth map estimated by our method, in comparison with the depth map that is estimated individually for each frame. In future we will evaluate our method using more sequences.

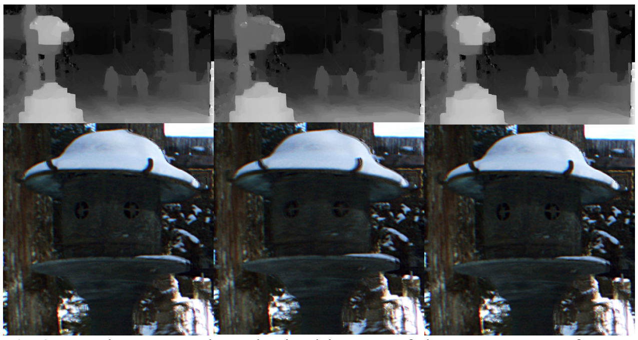

Fig 2 Depth maps, and synthesized images of three consequent frames, when the depth maps are estimated individually.

Fig 3 Depth maps, and synthesized images of three consequent frames, when the depth maps are estimated and stabilized by proposed method.

Table 1 Compression performance of codec depth maps, i.e. individually estimated, estimated using the foreground mask for stabilization.

QP (Quantization Parameter)		32	38	42	48
Non-stabilized	BR(Kbps)	3641.6	1359.9	688.42	257.1
Stabilized	BR(Kbps)	825.74	416.01	271.28	155.5

5. References

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, M. Lang, "Three-Dimensional video postproduction and processing", Proc. of IEEE 99(4), 607-625, 2011.

[2] M. Tanimoto, M. Panahpour Tehrani, T. Fujii, T. Yendo, "Free viewpoint TV," IEEE Signal Processing Magazine 28(1), 67–76, 2011.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compres-sion and transmission for a new approach on 3D-TV", Proc. of SPIE Stereoscopic Displays and Virtual Reality Sys-tems, XI, 93-104, Jan.2004.

[4] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmo-gorov, A. Agarwala, M.Tappen and C. Rother, "A comparative study of energy minimization methods for Markov ran-dom fields with smoothness-based priors," IEEE Trans. PAMI, 30(6), 1068-1080, 2008.

[5] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans PAMI, 1222–1239, Nov. 2001.

[6] K. Yamamoto, M. Kitahara, .H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, Y. Yashima, "Multiview Video Coding Using View Interpolation and Color Correction," IEEE Trans. CSVT, 17(11), pp. 1436 – 1449, Oct. 2007.

[7] "Description of Exploration Experiments in 3D Video Coding" MPEG 2008/W10173, Busan, Korea, October 2008.

[8] K. Suzuki, N. Fukushima, T. Yendo, M. Panahpour Tehrani, T. Fujii, and M. Tanimoto, "Free-viewpoint Image Generation using Moving Object Detection", Proc. of .IWAIT 2011, Jan 2011.