A TRIANGULAR-WARPING BASED VIEW SYNTHESIS SCHEME WITH ENHANCED ARTIFACT REDUCTION FOR FTV

Chao-Hsuan Li, Jang-Jer Tsai, and Hsueh-Ming Hang
Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan.
chli647@gmail.com, jjtsai@aa.nctu.edu.tw, hmhang@mail.nctu.edu.tw

ABSTRACT

View synthesis is one key enabling technology for free-viewpoint television (FTV). It uses multiple image frames and depth maps to generate the intermediate view at nearly any arbitrary viewpoint. In our proposed 2-stage algorithm, we adopt triangular warping with a new feature point extraction scheme for the first texture mapping stage. The new feature point extraction scheme uses the correlation between the image luminance gradients and the depth map. In the second stage, we use the median filtering and the multi-band blending technique to reduce the artifacts on the image object boundaries caused by the discontinuity and the imperfection of depth maps. Experimental results show that our proposed method produces better visual quality in comparison with a latest view synthesis algorithm.

Index Terms—Free-viewpoint TV (FTV), Triangular warping, Correlation of the gradients, Median Filtering, Multi-band blending.

1. INTRODUCTION

Free-viewpoint television (FTV) [1], which can provide users with arbitrary views of 3D scene, recently attracts the attention of both the academia and the industry. A complete 3DTV system consists of function units of depth estimation, multiview video codec, and view synthesis. The Depth Estimation (DE) unit generates the depth maps of 3D scene from the images captured by multiple cameras. Among the known DE methods [4], the stereo correspondence algorithms using energy minimization [5] are the most successful but the output depth maps still contain artifacts. The multiview video codec encodes and decodes all the data to ease the transmission and the storage of 3D scene. The View Synthesis (VS) unit generates an intermediate viewpoint image by using the depth maps and the captured images [6]. However, because of the discontinuity and the imperfection of depth map, annoying noises, particular on the object boundaries, appear in the synthesized images. Our goal in this study is to eliminate these artifacts and to improve the subjective visual quality of VS.

Most VS algorithms consist of two stages. The first stage is texture mapping, and the second stage is artifact reduction. 3D warping based methods [7][8][9] are the most popular texture mapping methods, which use depth map to assist the pixel-by-pixel warping. Recently the triangular warping method [3], which performs region-by-region warping on triangular meshes, was proposed. When the image segmentation tool accurately constructs the triangle mesh model of image objects, the depth value of a triangle mesh (assuming it is constant across a mesh) is more reliable than the depth of individual pixel. In this case, the triangular warping method leads to a better VS quality. However, the key is to select the feature points properly, because the feature points determine the triangular meshes and thus the quality of triangular warping. This paper proposes an effective scheme to acquire the good feature points.

Moreover, neither 3D warping nor triangular warping can completely eliminate the artifacts due to the depth discontinues and the incorrect depth map. Many researchers adopt some post-processing algorithms to improve its subjective quality. For example, [7] treats all boundaries as unreliable regions; thus, it proposes a two-layer method to alleviate the artifacts. It uses the matting algorithm to improve the visual quality. And, [9] classifies the boundaries of occlusion regions into three types, the foreground boundary, the hole region and the background boundary. It views the background boundary as unreliable region and uses post-filtering to reduce the boundary noises. In this study, we employ the median filtering and the multi-band blending [2] to reduce the visual artifacts in VS.

In brief, the two main contributions of this work are 1) a feature point selection technique in triangular warping, and 2) artifact reduction techniques for quality enhancement. The rest of paper is organized as follows. In section 2, we give a short review of a latest VS algorithm [8]. Section 3 presents our proposed VS algorithm. Experimental results are given in Section 4. And section 5 concludes this paper.

2. LATEST VIEW SYNTHESIS ALGORITHM

Fig. 1 (a) and (b): image and depth map of "Champagne Tower", and (c) and (d), those of "Kendo".

Among the studies on view synthesis, we focus on the fundamental case with two dense cameras. One camera is in the left and the other in the right. And the depth maps for the left and right image are given. As shown in Fig. 1, we

synthesize a virtual view at a specific viewpoint by using a pair of images and its corresponding depth map.

Fig. 2 shows the flow chart of one recently proposed VS method [8], which has 4 steps. In its texture mapping stage, Step T1 executes the forward warping to project the two views at different viewpoints to the desired intermediate viewpoint. In Step T2, the two warped views are overlapped into one by linear blending, which takes into account the distances between the virtual viewpoint and the positions of real cameras. In its artifact reduction stage, a reliability check in Step T3 calculates the differences between the pixels in the warped image and the original images. A pixel is treated as unreliable if the absolute difference between the corresponding pixels in the original image and the warped image is larger than a given threshold. Thus, two reliability tables (maps) are generated. In Step T4, when a pixel is located in the unreliable region of one reliability table and the reliable region of the other reliability table, the synthesized image uses the pixel in the reliable image. When a pixel is located in the unreliable regions of both reliability tables, we set its value to 0. Readers may refer to [8] for more details. Because of its good performance, it is selected as the benchmark in this study.

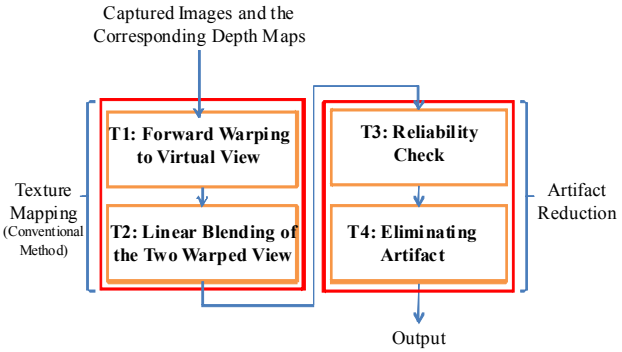

Fig. 2 Flow chart of Yang's VS. [8]

3. PROPOSED VIEW SYNTHESIS ALGORITHM

Fig. 3 shows the flow chart of our proposed algorithm. There are 3 steps in its texture map stage, and 2 steps in its artifact reduction stage.

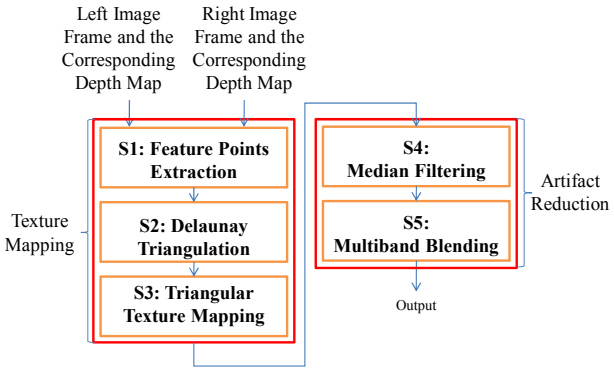

Fig. 3 Flow chart of our proposed VS.

3.1. Texture Mapping

We use triangular warping for the texture mapping in our proposed algorithm because the region-based synthesis may reduce the incorrect depth map artifacts. First, we extract feature points. Then, we apply Delaunay triangulation on the feature point set. At last, texture mapping of the triangles are conducted.

Feature Point Extraction (S1)

We calculate the gradient of the image luminance intensity (the image gray level) and the gradient of the depth map at each pixel position. For a specific pixel position, when the correlation between these two gradients is high, it is likely to be a feature point. Yet, because there always are noises in an image, we first use Gaussian smoothing filter to suppress its noises.

Specifically, we calculate the absolute value of the inner product of the gradients - $E(D, f)$ in (1), wherein ∇ is the gradient operator, D represents the depth map, and f denotes the Gaussian smoothed gray level image. That is, f is produced by (2), where g_{mask} is the Gaussian smoothing filter, and F is the gray level image. When the $E(D, f)$ value of a point is larger than those of all its neighboring points and a given threshold, it is judged as a feature point.

$$E(D, f) = |\nabla D \cdot \nabla f| \quad (1)$$

$$f = g_{mask} * F \quad (2)$$

Delaunay Triangulation (S2)

As shown in Fig. 4, we use the Delaunay triangulation method [3] to determine a unique set of triangles for a given feature points set.

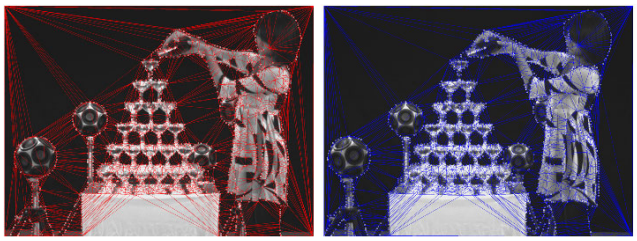

Fig. 4 Connecting feature points by Delaunay triangulation. Left: 'Champagne Tower' [10] viewpoint 39 (number of feature points: 1496), Right: viewpoint 41 (number of feature points: 1658).

Triangular Texture Mapping (S3)

For each triangle, we use 'its mode of depth' as 'the depth of the triangle' and calculate its disparity value accordingly. We perform texture warping based on the disparity. We first project the triangle with the smallest disparity to the warped image, and then the triangles with larger disparities. When two triangles overlap on the warped image, the triangle with the largest disparity is displayed. Thus, we ensure that the foreground triangles have higher priority than the background triangles in the displayed image. On a virtual image, the regions with no corresponding warped triangles are declared occlusion regions. Two examples of the virtual images and their occlusion regions are shown in Fig. 5.

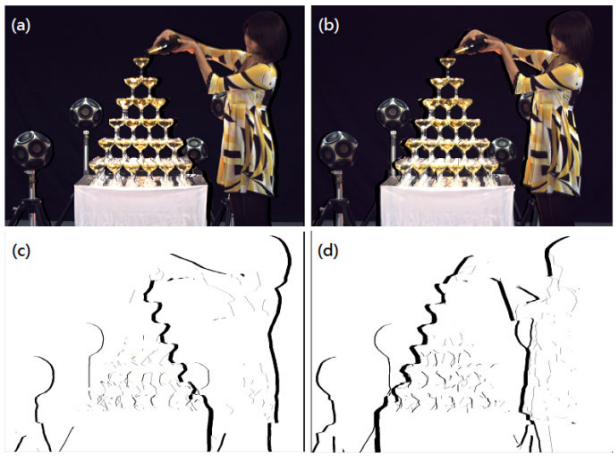

Fig. 5 (a) Virtual image generated using viewpoint 39, and (b) that using viewpoint 41, and (c) occlusion region map of case (a), and (d) occlusion region of case (b).

3.2. Artifact Reduction

Using the triangular warping, we obtain two virtual views for a specific viewpoint based on the given right and left images. Both views have occlusion regions and artifacts. In the artifact reduction stage, we use the higher confidence regions of one view to fix the artifacts on the other view. We employ the median filter and the multi-band blending technique in reducing artifacts.

Median Filtering (S4)

On the virtual image, the artifacts often appear in the form of small gaps. We use the median filter to remove them. In Fig. 6, (a) and (c) are the images before processing and (b) and (d) are the images after median filtering. Clearly, small gaps are significantly decreased.

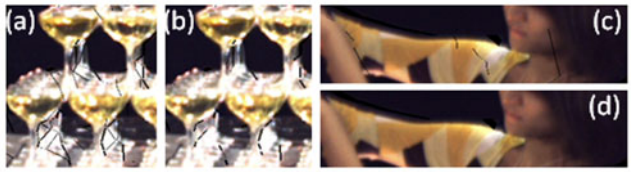

Fig. 6 Small gap removed by median filter.

Multi-band Blending (S5)

Furthermore, we adopt the multi-band blending [2] to further eliminate artifacts in VS. Multi-band blending blends low frequencies over a large spatial range and high frequencies over a smaller range. Generally, it provides good subjective visual quality.

Specifically, in (3), we acquire the Gaussian blurred image I_σ . We filter the original image I by a Gaussian blur filter $G_\sigma$with variance σ^2. In (4), $I_{(k+1)\sigma}$ is the output of Gaussian blur filter $G_{\sigma'}$ with input $I_{k\sigma}$, wherein σ' is defined in (5). In (6), the first band - B_σ is the difference between the original image - I and the Gaussian blurred image - I_σ. Similarly, in (7), the (k+1)-th band $B_{(k+1)\sigma}$ is the difference between $I_{k\sigma}$ and $I_{(k+1)\sigma}$. We obtain M_σ by (8), wherein M represents the occlusion regions after edge filtering, and M_σ denotes M filtered by the Gaussian filter G_σ. In (9), $M_{(k+1)\sigma}$ denotes $M_{k\sigma}$ filtered by $G_{\sigma'}$. In (10), we generate the multi-band blending image in the K-th band $I_k^{multiband}$ by summing $M_{k\sigma}^i B_{k\sigma}^i$ for all i, and n denotes the number of the virtual images. At last, in (11), we get the multi-band blending image by adding up $I_k^{multiband}$ for all k and m denotes the number of bands.

$$I_\sigma = G_\sigma * I \tag{3}$$

$$I_{(k+1)\sigma} = G_{\sigma'} * I_{k\sigma} \tag{4}$$

$$\sigma' = \sqrt{2k+1}\sigma \tag{5}$$

$$B_\sigma = I - I_\sigma \tag{6}$$

$$B_{(k+1)\sigma} = I_{k\sigma} - I_{(k+1)\sigma} \tag{7}$$

$$M_\sigma = G_\sigma * M \tag{8}$$

$$M_{(k+1)\sigma} = G_{\sigma'} * M_{k\sigma} \tag{9}$$

$$I_k^{multiband} = \sum_{i=1}^{n} M_{k\sigma}^i B_{k\sigma}^i \tag{10}$$

$$I^{multiband} = \sum_{k=1}^{m} I_k^{multiband} \tag{11}$$

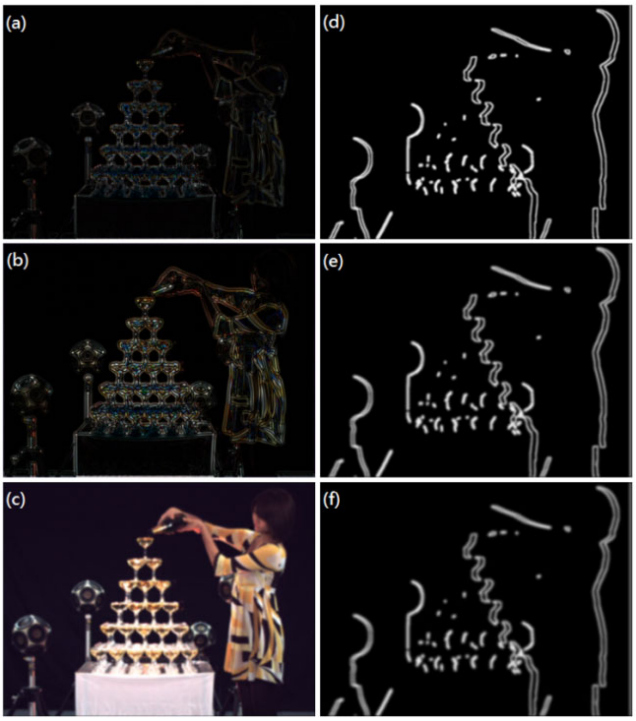

Fig. 7 A multiband blending example with 3 bands. The band $B_{k\sigma}$ for k = 1, 2, 3 are shown on the left, and the corresponding occlusion weights $M_{k\sigma}$ on the right.

In Fig. 7, we show a multi-band blending example with m=3 bands. In this study, we use the Gaussian blur filter with kernel size=15. For the first band, σ=5.

4. EXPERIMENTAL EVALUATIONS

In our experiments, we test two multiview image sequences "Champagne Tower"(1280×960) and "Kendo" (1024×768) generated by [10]. The associated depth maps are provided by [11]. Furthermore, we use Yang's VS [8] as the benchmark for comparison.

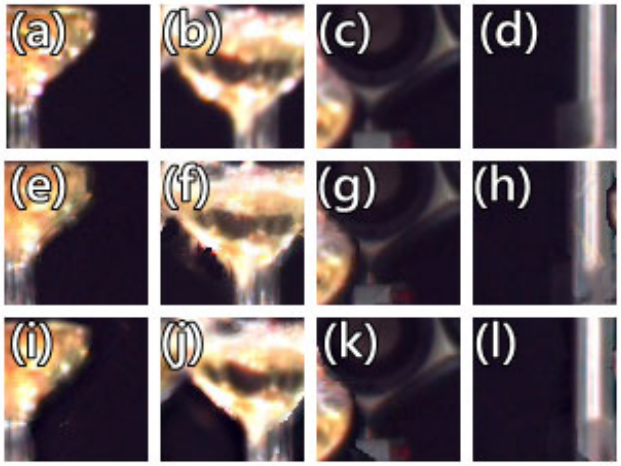

Fig. 8. "Champagne Tower". (a) to (d): ground truth images, (e) to (h): by Ynag's VS [8], (i) to (l): by our proposed method.

Fig. 9 "Kendo". (a) to (d): ground truth images, (e) to (h): by Yang's VS [8], (i) to (l): by our proposed method.

Fig. 8 and Fig. 9 show the (partial) ground truth images and the corresponding synthesized image for the test sequences "Champagne Tower" and "Kendo", respectively. In both figures, the images in the first (upmost) row show the ground-truth images; those in the second row are synthesized by [8], and those in the third (lowest) row, by our proposed algorithm. For "Champagne Tower", the artifacts near the cups and bar are lessened by our proposed algorithm. For "Kendo", our proposed algorithm effectively reduces some high frequency artifacts. Clearly our proposed VS method provides better subjective visual quality.

For "Champagne Tower", the average PSNR of the first 100 frames synthesized by our proposed method is 29.71dB, which is comparable to that by [8] (29.74dB). For "Kendo", the average PSNR by our proposed method is 34.85dB, which is slightly inferior to that by [8] (35.54dB).

5. CONCLUSIONS

In this study, we improve the feature point extraction scheme for triangular warping and propose an artifact reduction mechanism. The proposed scheme uses the correlation between the luminance information of the image and the depth map to decide adequate feature points. Due to the discontinuity and imperfection of depth map, artifacts appear in the synthesized virtual images. In the artifact reduction stage, we use the median filter and the multi-band blending technique to improve image subjective quality. The multi-band blending scheme produces smooth transitions when mixing two synthesized images. Overall, the resulting synthesized image quality is visually pleasant.

6. ACKNOWLEDGMENT

This work was supported in part by the NSC, Taiwan under Grants 98-2221-E-009 -087 and 98-2219-E-009 -015.

7. REFERENCES

[1] M. Tanimoto, "Overview of free viewpoint television," *Signal Processing: Image Communication*, vol.21, no.6, pp.454-461, July, 2006.

[2] M. Brown, and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision* (IJCV),vol. 74(1), pp.59-73, 2007.

[3] B. Choi, et al., "Intermediate view synthesis algorithm using mesh clustering for rectangular multiview camera system," *Proceeding of SPIE*, vol. 49, no.2, Feb., 2010.

[4] D. Scharstein, and R.Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," I*nternational Journal of Computer Vision* (IJCV), vol.47, no. 1-3, pp.7-42, 2002.

[5] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1222–1239, Nov. 2001.

[6] C. Fehn, "3D-TV using depth-image-based rendering (DIBR)," in *Proc. Picture Coding Symp.*, pp.307-312, Dec. 2004.

[7] C. L. Zitnick, et al., "High-quality video view interpolation using a layered representation," *ACM SIGGRAPH and ACM Trans. Graphics*, vol. 23, pp. 600–608, 2004.

[8] L. Yang, et al., "Artifact reduction using reliability reasoning for image generation of FTV," *Journal of Visual Communication and Image Proc.*, vol.21, no.5-6, pp.141-151, 2010.

[9] C. Lee and Y. S. Ho, "Boundary filtering on synthesized views of 3D video," in *Proc. International Conference on Future Generation Communication and Networking Symposia*, pp. 15–18, 2008.

[10] M. Tanimoto et al., Coding of moving pictures and audio, ISO/IEC JTC/SC29/WG11, N11095, Jan. 2010.

[11] M. Tanimoto et al., "Reference softwares for depth estimation and view synthesis", ISO/IEC JTC/SC29/WG11, M15377, Apr. 2008.