STEREOSCOPIC IMAGE GENERATION BASED ON DEPTH IMAGES

Liang Zhang, *Wa James Tam, Demin Wang*

Communications Research Centre Canada
3701 Carling Avenue, Ottawa, Ontario, K2H 8S2, Canada
email: liang.zhang@crc.ca

ABSTRACT

A depth-image-based rendering system for generating new views is proposed. One important aspect of the proposed system is that the depth maps are pre-processed using an asymmetric filter to smoothen the sharp changes in depth at object boundaries. In addition to ameliorating the effects of blocky artifacts and other distortions contained in the depth maps, the smoothing reduces or completely removes disocclusion areas where potential artifacts can arise from image warping which is needed to generate images from new viewpoints. The asymmetric nature of the filter reduces the amount of geometric distortion that might be perceived otherwise. We present some results to show that the proposed system provides an improvement in image quality of stereoscopic virtual views while maintaining reasonably good depth quality.

1. INTRODUCTION

Depth-image-based rendering (DIBR) techniques have recently received much attention in the broadcast research community as a promising technology for three-dimensional television (3D TV) systems [1][2]. Whereas, the classical approach requires the transmission of two streams of video images, one for each eye, 3D TV systems based on DIBR will provide a single stream of monoscopic images and a second stream of associated images that convey per-pixel depth information. One advantage of using depth images is that they can be coded more efficiently than two streams of natural images, thereby reducing the bandwidth required for transmission. More importantly, DIBR provide the added flexibility of being able to create new images, using information from the depth images or depth maps, as if they were captured with a camera from a different viewpoint. The drawback is that with this type of data representation, one or more "virtual" images of the 3D scene have to be generated at the receiver side in real time. As well, it is not an easy task to create the new images with decent image quality.

A depth map is essentially a two-dimensional (2D) function that gives the depth (with respect to the camera position) of a point on an object in a visual scene as a function of the image coordinates. Since the depth of every point in an original image is known, a virtual image of any nearby point can be rendered by projecting the pixels of the original image to their proper 3D locations and re-projecting them onto the virtual image plane. Thus, this rendering of the original images according to their depth maps can create virtual images of the scene.

The most significant problem in DIBR is how to deal with holes appearing in the virtual images. Holes are due to the accretion (disocclusion) of features of objects or background that would have been visible only from the new viewpoint but not from the original location that was used in capturing the original image. There is no information in the original image for these disoccluded regions and, therefore, they are empty, like holes. To fill in the holes, the most commonly used method is to map a pixel in the original image to several pixels in the virtual image by simple interpolation of pixel information in the foreground or background. A somewhat more complex extrapolation technique can also be used. However, depending on the scene layout, all these approaches are known to produce more or less annoying visible artifacts in the virtual images.

To deal with the disocclusion artifacts in the virtual images several approaches have been suggested. In one approach, layered-depth-image (LDI) [3] provides a set of original images of a scene and their associated depth maps. The images and depth maps store not only what is visible in the original image, but also what is behind the visible surface. In another approach, depth maps are pre-processed. In a technique that we have proposed depth maps are processed using a 2D Gaussian filter, so that the disocclusion artifacts are incrementally removed as the smoothing of depth maps becomes stronger [4]. Experimental results using formal subjective evaluation techniques indicate that this technique (s*ymmetric smoothing*) can be used to significantly improve the image quality of novel stereoscopic views especially when there are blocky artifacts in the depth maps and potential distortions in the newly generated images as a result of disocclusion [4]. Recently, Fehn also proposed a similar idea to pre-process the depth map using a 2D Gaussian smoothing such that "no disocclusions" appear in the "virtual" stereoscopic images [6].

In this paper, we present a new algorithm for DIBR that provides an improvement in results over our previous technique. The first part pre-processes the depth maps,

0-7803-8554-3/04/$20.00 ©2004 IEEE.

which includes setting of the zero parallax planes and smoothing of the depth maps. The second part builds the 3D images by warping the original center images to create virtual left-eye and right-eye images, based on information from the depth maps. The third part deals with disocclusion by filling the holes, as necessary. Based on this system, different types of artifacts and distortions that could appear in the virtual images are experimentally investigated for different system parameters. Following this investigation, we propose the concept of *asymmetric smoothing* of depth maps to improve the image quality of the rendered stereoscopic images. This is in contrast to our previous technique where *symmetric smoothing* was used in [4].

The present paper is organized as follows. After the introduction, the proposed rendering system is described in Section 2. Section 3 is devoted to the comparison of experimental results with different system setups. Section 4 proposes the concept of *asymmetric smoothing* of depth maps. Section 5 provides a discussion of the results.

2. DEPTH-IMAGE BASED RENDERING SYSTEM

A flowchart describing the proposed depth-image-based rendering system is illustrated in Fig. 1. This system consists of three steps: (i) pre-processing of depth maps, (ii) 3D image warping and (iii) hole-filling. Note that (iii) may not be necessary if there are no holes to fill as a result of optimal pre-processing. In the following, these three steps will be addressed in detail.

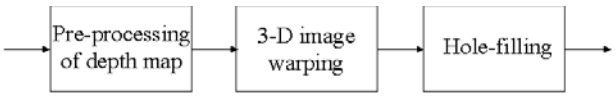


Fig. 1. Flowchart of the proposed depth-image-based rendering system.

A. *Pre-processing of depth maps*

The pre-processing of depth maps includes two issues: one is in choosing the *convergence distance* Z_c (so-called *zero-parallax setting* (*ZPS*)) and the other is in smoothing the depth maps.

To establish a *ZPS*, two methods have been proposed. In the so-called "toed-in" approach, the *ZPS* is chosen by a joint inward rotation of the left-eye and right-eye cameras. In the so-called *shift-sensor* approach, a plane of convergence is established by a small shift h of the CCD sensors in the pair of parallel cameras. Different from these two methods, in the present rendering system, the *ZPS* is chosen by "shifting" the depth map. Without loss of generality, we choose

$$Z_c = (Z_{near} - Z_{far})/2$$

as the *ZPS* plane, where Z_{near} and Z_{far} are the nearest clipping plane and the farthest clipping plane of the depth map. In an 8-bit depth map, Z_{near}= 255 and Z_{far}=0 (Fig. 2). After that, the depth map is further normalized with the factor of 255, so that the values of the depth map lie in the interval of [–0.5, 0.5], values that are required by the image warping algorithm.

Fig. 2. The test image: "Interview". The original center image is on the left and its associated unprocessed depth map is on the right. The luminance value in the depth map reflects the depth value.

The second issue in the pre-processing step is to smooth the depth maps. To this end, a Gaussian filter $g(x, \sigma)$

$$g(x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2}{\sigma^2}\right\}, \quad for \quad -w \le x \le w \ , \qquad (1)$$

is used, where w is the filter's window size and σ the standard deviation. The value of σ determines the severity of smoothing of the depth maps. Let $s(x,y)$ be a depth value in the depth map at the pixel (x,y). Then, depth value $\hat{s}(x,y)$ after smoothing is equal to

$$\frac{\sum_{\upsilon=-w}^{w}\left\{\sum_{\mu=-w}^{w}\left(s(x-\mu,y-\upsilon)g(\mu,\sigma_\mu)\right)g(\upsilon,\sigma_\upsilon)\right\}}{\sum_{\upsilon=-w}^{w}\left\{\sum_{\mu=-w}^{w}\left(g(\mu,\sigma_\mu)\right)g(\upsilon,\sigma_\upsilon)\right\}} \qquad (2)$$

It is expected that different values of w and σ have different impact on the image quality of the virtual images generated from the original center image. We will discuss this issue in the next section. In the following experiment, we let w be equal to 1.5σ empirically.

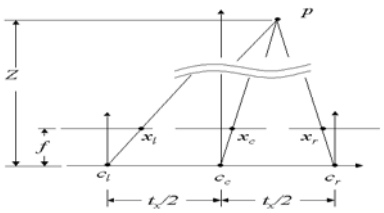


Fig. 3. Camera configuration used for generation of virtual stereoscopic images.

B. *3-D image warping*

For simplicity, we only consider the commonly used camera configuration for generating virtual stereoscopic images from one center image associated with one depth map for 3D TV (Fig.3). In this case, the vertical coordinate of the projection of any 3D point on each image plane of three cameras is the same. Let c_c be the viewpoint of the original center image, c_l and c_r the viewpoint of the virtual left-eye and right-eye images to be generated. t_x is the distance between these two virtual cameras. Under this camera configuration, one point p with the depth Z in the world is projected onto the image plane of three cameras at pixel (x_l, y), (x_c, y) and (x_r, y), respectively. From the geometry shown in Fig. 3, we have

$$x_l = x_c + \frac{t_x}{2}\frac{f}{Z} \quad , \quad x_r = x_c - \frac{t_x}{2}\frac{f}{Z} \quad , \tag{3}$$

where information about x_c and f/Z is given in the center image and the associated depth map, respectively. Therefore, with formulation (3) for 3D image warping, the virtual left-eye and right-eye images can be generated from the original center image and its depth map by providing the value of t_x.

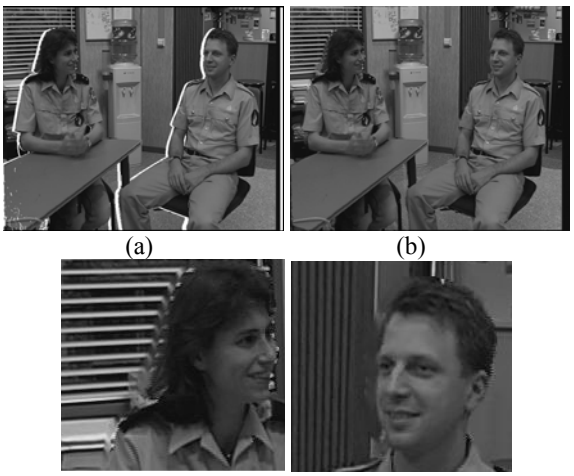
C. *Disocclusion and hole-filling*

Due to a difference in viewpoints, some areas that are occluded in the original image might become visible in the virtual left-eye or the right-eye image. These areas, referred to as "disocclusion" in the Computer Graphics literature, have no texture after 3D image warping because information about the disocclusion is not available either in the center video or in the accompanying depth map. We fill in the disocclusion areas by averaging textures from neighborhood pixels, and this process is called *hole-filling*.

3. EXPERIMENTAL COMPARISON

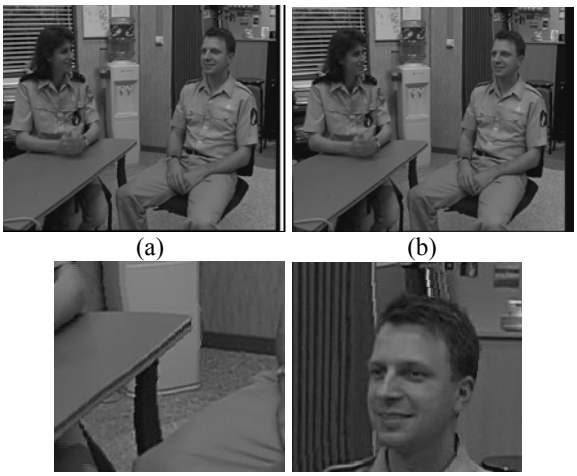
This section is devoted to evaluating the performance of the proposed rendering system using natural images. As an example, only results with the test image "Interview" (Fig. 2) are shown. In the following investigation, the distance of two virtual left-eye and right-eye cameras is fixed at 48 pixels.

First, we evaluate the performance of this system *without* smoothing the depth maps. Fig. 4 shows an example of the results of the virtual left-eye image. The image after 3D image warping is illustrated in Fig. 4 (a). In the figure, white areas are the disocclusion areas. We can see the disocclusion areas mainly along the object boundaries and



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

Fig. 4. Virtual left-eye image generated *without* smoothing of the depth map. (a) Image after 3D image warping. White areas represent disocclusion areas; (b) Image after *hole-filling*; (c) and (d) Artifacts more clearly seen in enlarged segments of the image from (b).



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

Fig. 5. Virtual left-eye image generated using *symmetric smoothing* of the depth map with σ_μ=σ_v=30. (a) Image after 3D image warping. White areas are disocclusion areas; (b) Image after *hole-filling*; (c) and (d) Enlarged segments of the image shown in (a). Notice the curved table leg in (c) and the curved vertical lines in (d).

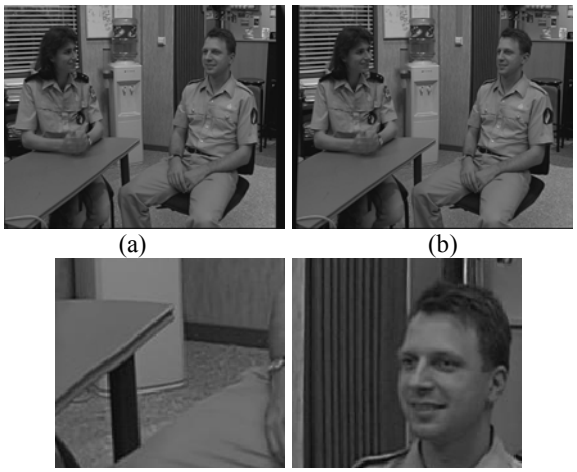
the right margin of the image. After *hole-filling*, as shown in Fig. 4 (b), significant artifacts appear in the virtual image. Fig. 4 (c) and (d) show clearer illustrations of the artifacts by enlarging segments of Fig. 4 (b). The texture artifact stem from hole-filling using neighborhood textures because information about the previously occluded areas is available neither in the monoscopic video nor in the accompanying depth maps.

We then evaluate the performance of the proposed system *with* smoothing of the depth maps. Similar to [4], we let σ_μ and σ_v of two Gaussian filters have the same value, so-called *symmetric smoothing*. Fig. 5 shows the results of the virtual left-eye image with σ_μ=σ_v=30. The image after 3D image warping is illustrated in Fig. 5 (a). White areas represent the disocclusion areas. Compared to Fig. 4 (a), we can see from Fig. 5 (a) that the disocclusion areas within the image almost disappear except for the area in the right margin of the image. This can be explained as follows. Due to the smoothing of the depth map, there are no more sharp depth discontinuities. In other words, the disocclusion areas become sparse through smoothing and even disappear, as the smoothing becomes stronger. After *hole-filling*, it can be seen from Fig. 5 (b) that no *texture artifacts* appear in the virtual image within the image, compared to Fig. 4 (b). However, some vertically straight object boundaries now become curved. This can be more clearly seen in Fig. 5 (c) and (d), which show enlarged segments of the image in Fig. 5 (b). We call this type of distortion, *geometric distortion*. The origin of this type of distortion can be explained as follows. Let us examine the table leg in Fig. 5 (c). In the unprocessed depth map, this table leg has the same depth value along its vertical length and, at the bottom of the leg, the legs of the man and the woman with relatively large depth are in its neighborhood

(as can be seen in Fig. 2). After processing, due to smoothing of the depth map in the horizontal direction and at a level that is as strong as that in the vertical direction, the bottom of the table leg has a slightly larger value than that of its top. This creates a curved table leg after 3D image warping.

4. ASYMMETRIC SMOOTHING OF DEPTH MAP

Based on the analysis above, the strength of the smoothing of depth map in the horizontal should be much less than that of the smoothing in the vertical, so that the object, e.g. the table leg, has similar depth value throughout after the smoothing. We call this *asymmetric smoothing*.



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

Fig. 6. Virtual images generated using *asymmetric smoothing* of the depth map with σ_μ=10 and σ_υ=90. (a) Left-eye image; (b) Right-eye image; (c) and (d) Enlarged segments from the image shown in (a).

Fig. 6 shows the results of rendering using *asymmetric smoothing* of the depth map with σ_μ=10 and σ_υ=90. The virtual left-eye and right-eye images generated from the original center image (Fig. 2) are shown in Fig. 6 (a) and (b). Two enlarged segments from the left-eye image are shown in Fig. 6 (c) and (d) for better comparison. It can be seen from Fig. 6 (c) and (d) that *geometric distortions* are strongly reduced compared to Fig. 5 (c) and (d). Also, no *texture artifacts* appear. Generally speaking, using *asymmetric smoothing,* the virtual images have very sharp texture and high image quality. When viewed in a stereoscopic display, they also create reasonably good and stable depth. Our next step will be to conduct formal subjective assessments on stereoscopic images that have been generated using this *asymmetric smoothing* technique on depth maps.

5. DISCUSSIONS AND CONCLUSIONS

DIBR has the inherent problem of having to deal with disocclusion areas, and filling in these "holes" so as to create new images with decent image quality is not easy. In this paper we propose pre-processing of the depth maps to smooth the sharp changes in depth at object boundaries. In addition to ameliorating the effects of blocky artifacts and other distortions that might be contained in the depth maps, the smoothing reduces or completely removes disocclusion areas where potential artifacts can arise from image warping. In a previous experimental study we found that subjective ratings of image quality in the stereoscopic virtual views can be improved with *symmetric smoothing* [4] and [5]. Results presented in this article demonstrate that the proposed system using *asymmetric smoothing* provides a significant improvement in image quality over symmetric smoothing by reducing the amount of geometric distortion that might be present otherwise.

In addition to improving overall image quality of the virtual views, smoothing depth maps can potentially lead to other benefits. Firstly, smoothing reduces the contrast of depth maps and, thus, narrows the range of disparities contained in the rendered images. This will lead to increased visual comfort for viewing virtual stereoscopic images that are rendered from depth maps with large disparities. Secondly, smoothing reduces the sharp transitions at the edges and borders of objects that are in front of a background and, therefore, smoothes out the depth at and near the outlines of objects. Informal observations indicate that this removal of "crispness" at the borders of objects removes chances of perceiving the "cardboard effect" in which objects appear in depth but seam to be flat like a sheet of cardboard.

Nevertheless, smoothing of depth maps, while attenuating some artifacts, will lead to a diminution of the depth resolution contained in the rendered stereoscopic views. Future studies will be required to examine this tradeoff more closely, although based on the previous and current studies the benefits appear to outweigh this disadvantage.

6. REFERENCES

[1] A. Redert, M. Op de Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, P. Surman, "ATTEST –advanced three-dimensional television system techniques", Proc. of 3DPVT' 02, pp. 313-319, Padova, Italy, Jun. 2002.

[2] J. Flack, P. Harman, S. Fox, "Low bandwidth stereoscopic image encoding and transmission", Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems X, Vol. 5006, pp. 206-214, CA, U.S.A., Jan. 2003.

[3] J. Shade, S. Gortler, L. He, R. Szeliski, "Layered depth image", Proc. of SIGGRAPH'98, pp. 231-242, Jul. 1998.

[4] G. Alain, W. J. Tam and L. Zhang, "Improving stereoscopic image quality of pictures generated from depth maps", Internal CRC report, Apr. 2003.

[5] W. J. Tam, G. Alain, L. Zhang, T. Martin, R. Renaud, "Smoothing depth maps for improved stereoscopic image quality", Proc. of SPIE (ITCom) Three-dimensional TV, Video, and Display III, Philadelphia, U.S.A., Oct. 2004.

[6] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)", Proc. of VIIP 03, Benalmadena, Spain, Sept. 2003.