SMOOTHING DEPTH MAPS
FOR IMPROVED STEREOSCOPIC IMAGE QUALITY

W. J. Tam, G. Alain, L. Zhang, T. Martin, R. Renaud, D. Wang

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

ABSTRACT

A technique to improve stereoscopic image quality of pictures generated from depth maps is proposed. Generally, there are two key problems with methods that render new images from an original image and its corresponding depth map: the depth map could contain artifacts (e.g., blockiness or contouring) and there is no information on newly exposed (disoccluded) regions. For the former, the solution is to acquire depth maps with better "resolution." For the latter, a common solution is to fill the newly exposed regions with the weighted average of luminance and chrominance values of neighbouring pixels. However, this latter solution often leads to visible and annoying artifacts at edges of objects. We propose to smoothen depth maps before the rendering of new views so as to minimize the effects of blockiness and contouring as well as to generate gentle depth transitions at object boundaries to alleviate problems with disoccluded areas. Experimental results from subjective assessments demonstrate the effectiveness of smoothing depth maps in improving the perceived quality of stereoscopic images.

1. INTRODUCTION

In recent years there has been a great impetus to enhance the viewing experience of visual displays and communication systems by incorporating stereoscopic information into video signals [1]. In particular, with the maturation of technologies and infrastructure for digital video, as epitomized by standards for digital television and the advent of digital cinema, there has been a strong momentum to find efficient and commercially viable methods of creating, recording, transmitting, and displaying stereoscopic images and sequences.

The fundamental problem of working with stereoscopic video images is that there are two streams of images to deal with, as compared to a single stream when dealing with standard two-dimensional images. One stream of images is needed for the left eye and a second stream for the right eye. Given this requirement to double the number of images, the problem of storage and transmission, in addition to creation and display, is an outstanding research issue.

One method of transmitting stereoscopic video sequences (such as for 3D-TV) is to transmit depth maps of visual scenes together with a standard stream of two-dimensional (2D) images that have been captured from a single camera viewpoint [2]. At the receiving end, the depth information and corresponding 2D images are used to synthesize new images as if they were captured from a slightly different camera viewpoint and, thus, form corresponding image pairs for stereoscopic viewing. A major problem with this method is that there are difficulties involved in obtaining high quality depth maps. Although there are methods that utilize semi-automatic processing of 2D images to generate pseudo-depth maps, such as [3], which are relatively good for venues in mass entertainment, they still require an appreciable amount of human intervention to ensure accuracy of the depth maps. There are also hardware that can be used to acquire depth maps based on range-finding techniques, such as the Z-camera (3DV Systems, Israel), but the lighting conditions and range of shooting distances for these techniques are often restricted. As well there are software methods to generate depth maps (which are really horizontal disparity maps[1]) from pairs of stereoscopic images, see [4], but the resulting depth maps may contain blocky artifacts, depth instabilities, and inaccuracies. Another major problem with the method of depth-image based rendering (DIBR) concerns the rendering of newly exposed regions that occur at the edges of objects where the background was previously hidden from view. This is a problem because the depth maps do not provide any information on how to properly fill in these regions. Although not ideal, a common method is to fill these regions with the weighted average of luminance and chrominance values of neighbouring pixels. However, this solution often leads to visible and annoying artifacts at edges of objects.

In the present study we report the technique of smoothing depth maps so as to remove blocky artifacts and generate gentle transitions of depth at object boundaries before the rendering of new views. We present results showing the effectiveness of smoothing the depth maps in improving the perceived quality of stereoscopic images. In particular, the stereoscopic images are of asymmetrical quality, with one member of each stereoscopic pair being the source and the

---

[1] In this article we will not make a distinction between depth maps and disparity maps. Strictly speaking, the former contains information about the absolute distances of objects in a visual scene from the camera. The latter contains information of the relative distances among objects in the visual scene and requires a scaling factor in order to obtain absolute distances of objects from the camera.

other member being a rendered image. In previous studies we have determined that asymmetrical coding can be useful in preserving picture and depth quality while reducing bandwidth requirements [5][6][7][8].

2. EXPERIMENTAL METHOD

Twenty three observers from a university, screened for normal visual acuity and stereoscopic vision, assessed perceived image quality and depth of a set of stereoscopic sequences. The stereoscopic sequences consisted of images for the left eye that were from an original image sequence (see Figure 1) and images for the right eye that were created by "warping" the original images based on their corresponding depth maps.

Fig. 1. Samples of single frames from the four video sequences are shown. In clockwise order starting from the top left, “Interview,” “Puppy,” “Tulips,” and “Soccer.”

The important aspect of the experiment was that the level of smoothing of the depth maps was manipulated. This was achieved by filtering the depth maps with a Gaussian blur filter in which the standard deviation and kernel size were adjusted. A computer program that was developed in-house was used to create new views by "warping" the original images based on the information in the processed depth maps (see Fig. 2). The depth maps for the sequences were obtained from the institutions that generously provided the source sequences (HHI, Germany and ETRI, Korea) or through using our own in-house developed software for disparity estimation [4]. Of the depth maps, "Interview" and "Tulips" had pixel depth resolution and were relatively stable. Those for "Soccer" and "Puppy" had 8x8-block depth resolution and were not as stable in that there were blocky artifacts that appeared and disappeared over the duration of the sequence.

Viewers rated the sequences using the double-stimulus continuous-quality scale method that is a standard procedure as described in ITU-R Recommendation 500 [9]. Specifically, viewers were presented with two versions of the same sequence, announced as "A" and "B." Either "A" or "B" was a reference sequence (both eyes see the same source image, thus, monoscopic source sequence). The other sequence was a test sequence, which consisted of either a source image for the left eye and a rendered image for the right eye (stereoscopic), or rendered images for both eyes (non-stereoscopic).

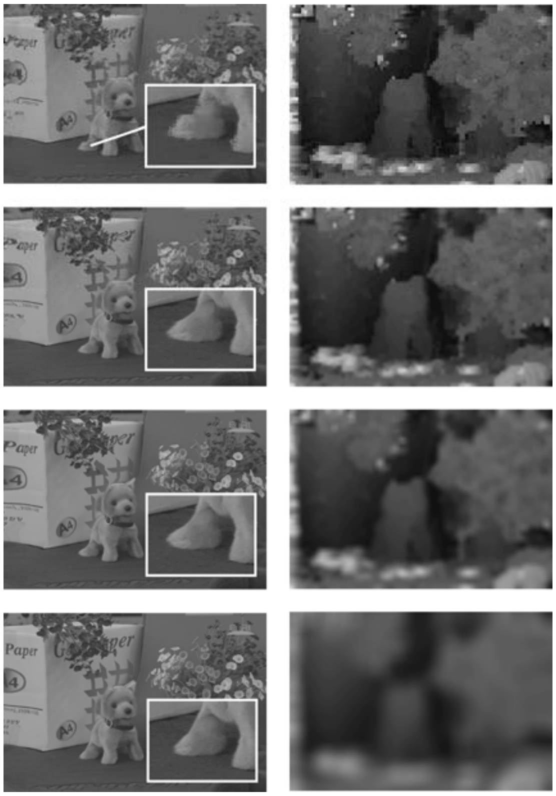

Fig. 2. Sample set of rendered images for a single frame from the sequence “Puppy,” and their corresponding depth maps that have been processed using different levels of Gaussian blur. Notice in the first row the distortions in the text as well as in the rear leg of the puppy (magnified in inserts) and how the distortions are reduced, moving downward to the bottom row, as the level of smoothing (Gaussian blur) is increased. The original image for "Puppy" can be found in Fig. 1.

The order in which the Reference and the Test appeared was randomized, and viewers were not informed whether “A” or "B” was the Reference or the Test. This pair of “AB” presentations was repeated to form a single trial. Each sequence was 8 seconds in duration and there was a 3-second interval between the “A” and “B” sequences. Each trial was separated by a 6-second response interval. During the blank intervals a gray field was displayed. At the end of the trial viewers rated both the "A" and the "B" sequences separately using a rating scale that ranged from "BAD" to "EXCELLENT," as shown in Figure 3.

Each stereoscopic pair of images, consisting of a left-eye image (720 x 480) and a right-eye image (720 x 480), was stored in one HDTV video frame (1920 x 1035 interlaced). The image sequences were then played back on a DVS HDStationPro digital disk recorder through a sync doubler, Crystal Eyes GDC-3, at 120 Hz on an Electrohome 9500LC rear-projection system. This display system has a fast P41 green phosphor that helps minimize possible ghosting effects, arising from crosstalk due to phosphor persistence. The active picture area of the display was 48” x 36”and peak monitor luminance was 70 cd/m^2. The liquid-crystal shutters, placed in front of viewers’ eyes, had a neutral light transmittance of

30%, and response times of 0.2 ms and 2.8 ms for closing and opening, respectively. Viewing distance was 4H (48 inches).

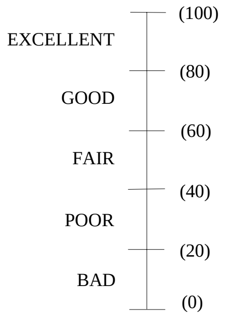

Fig. 3. Rating scale for image quality and perceived depth. The numbers in parentheses were not printed on the actual response sheet, but are indicated here to suggest how the ratings were digitized to range between 0 and 100 for analysis.

There were three independent variables: Video Sequence (4 sequences: "Interview," "Soccer," "Puppy," "Tulips"), Format (2 Display formats: Stereo, Non-stereo), Blur (5 blur levels: 000-00, 010-10, 020-20, 060-30, 120-80). For Blur, the first number in each pair refers to the standard deviation and the second number refers to the size of the kernel in pixels and they were used as parameters in the MatLab library function for Gaussian blur (The MathWorks, Inc., USA). The larger the value the greater the level and extent of smoothness. The combination of these three variables yielded 40 experimental conditions. Each condition was embedded in a trial consisting of the sequential presentation of the Test sequence and a Reference, which is the non-stereoscopic version of the test sequence as described previously. Experimental trials were presented in a fixed quasi-random order. Viewers rated the image quality of the sequences in one session and the overall perceived depth in a second session. The order of the sessions for image quality and depth were counter-balanced among viewers.

3. EXPERIMENTAL RESULTS

The results for subjective image quality are presented in Fig. 4. For simplicity, the ratings for all the Reference sequences were averaged and the mean is shown as a gray horizontal line in the graph. Standard deviation of the mean was smaller than 10 (rating) units. Recall that the Reference was a non-stereoscopic source sequence and did not carry any of the artifacts that would have resulted from image-based rendering. Thus, it is not surprising that the image quality is in general higher than the ratings for the other conditions. The most important aspect of the figure is that ratings of all four stereoscopic sequences increased with increasing levels of smoothing. Interestingly, this trend occurred for the "Tulips" sequence that had relatively high ratings throughout, as well as for the other sequences where ratings were much lower. This result suggests that smoothing of depth maps can be useful for improving perceived image quality of stereoscopic sequences independent of their inherent picture quality. Not shown in Fig. 4 are the results for the non-stereoscopic rendered sequences which also showed an upward trend with increasing levels of smoothing. This result indicates that smoothing of depth maps improved the image quality of the (monoscopic) rendered images.

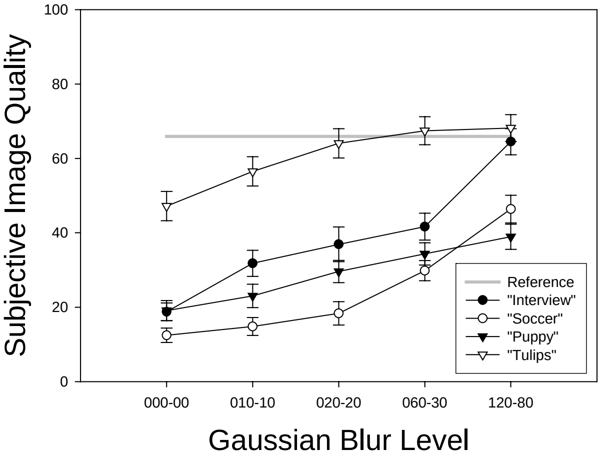

Fig. 4. Ratings of perceived image quality of four stereoscopic sequences, with one member of each stereoscopic pair being a source and the other member being a rendered image. The depth maps were processed by smoothing with a Gaussian blur filter at different levels, indicated by the pairs of numbers on the x-axis. The first number in each pair refers to the standard deviation and the second number refers to the size of the kernel of the filter. The larger the values the greater the level and extent of smoothness. The symbols represent the mean and the error bars are standard errors based on the ratings from 23 viewers. The Reference refers to the mean of all reference non-stereoscopic (source) sequences.

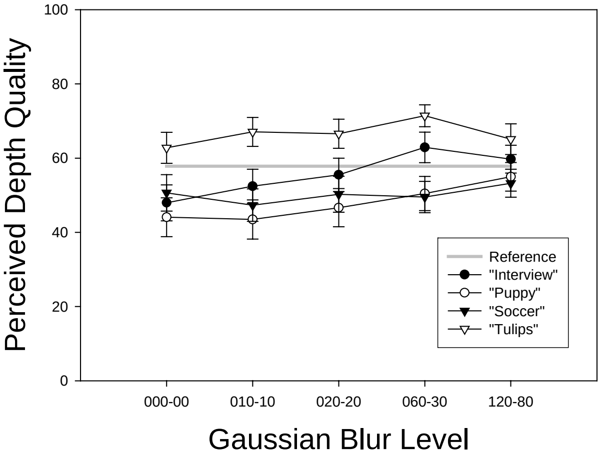

Fig. 5. Ratings of perceived depth quality of the same four stereoscopic sequences as described in the caption of Fig. 4. Data are from the same 23 viewers.

Fig. 5 shows the results of depth quality for the stereoscopic test sequences in which the images presented to one eye are from the source and the images presented to the other eye are rendered images. Again, for simplicity, the ratings for the Reference sequences were averaged and the mean is presented as a gray horizontal line in the graph. Standard deviation of the mean was less than 10 rating units. The results show that the stereoscopic depth effect from disparity information was most effective for the "Tulips" sequence, as indicated by the ratings being higher than the horizontal gray line. For the other sequences, the depth quality is less convincing. Nevertheless, it should be pointed out that for the sequence with the best depth quality ("Tulips"), ratings remained stable even though the depth maps have been smoothened to extreme extents. Notice also that the ratings of depth quality increased with increasing levels of smoothing even for the sequences ("Interview" and "Soccer") that were rated lower or the same as the Reference monoscopic sequence. In other words, smoothing of depth maps did provide improvements in perceived depth quality for those that started out with low ratings and, more importantly, did not marr ratings of depth quality in those that had high ratings of depth quality already. As for ratings of depth quality of the monoscopic rendered sequences, ratings were lower than for the stereoscopic test sequences and the results, being intuitive and as expected, are not illustrated here.

4. DISCUSSION AND CONCLUSION

The present study examined the effect of smoothing of depth maps on image quality and perceived depth of stereoscopic video sequences thus created. Experimental results obtained from 23 viewers indicated that smoothing of depth maps before image-based rendering can help improve the image quality of stereoscopic sequences because of elimination of blocky artifacts, redistribution of inaccuracies in the depth maps, as well as smoothing transitions at object boundaries in the rendered images. As a matter of fact, the smoothing at boundaries of objects in the depth maps reduced reports of the "cardboard effect" (objects appearing flat but in depth), when viewers were questioned informally after the experiment. This general improvement in image quality from smoothing of depth maps was achieved without any negative impact on the perceived depth quality.

The present findings have significant implications for 3D-TV and other stereoscopic display systems that are based on depth-image based rendering (DIBR). Although it is often thought that the spatial resolution of depth maps has to be high in order to obtain rendered views that are of the highest quality, the present results suggest that this need not be the case. We have shown that smoothing depth maps before the rendering of new views, i.e., the process effectively reduces the spatial resolution of the depth maps, actually helped improve the perceived image quality of the rendered images. We believe that this occurred because smoothing decreased the problem of visible artifacts in the stereoscopic sequences as a result of the removal of blocky artifacts and instabilities in the depth maps and the reduction of sharp depth transitions that usually lead to large disoccluded areas in the new views. Thus, the present findings suggest that depth maps for DIBR do not have to be of high spatial resolution, as was once thought, to produce stereoscopic images that are of relatively good subjective image quality.

The most important conclusion of the present study is that smoothing of depth maps can help improve the image quality of rendered images and, thus, is useful for depth-image based rendering of stereoscopic views, such as for 3D-TV. As a final note, in the current study uniform application of smoothing over the entire image frame was implemented. For future studies, it might be worthwhile to investigate whether any further benefits might be obtained with selective smoothing at localized regions, such as at edges of objects, and with extent of smoothing based on the disparity (depth) magnitude.

5. ACKNOWLEDGEMENTS

We thank Dr. Xianbin Wang and Dr. Wei Li for their constructive comments on an earlier version of the manuscript. W would also like to thank Dipl. C. Fehn of HHI and Dr. S. Cho of ETRI for providing us with a number of the stereoscopic test sequences and depth maps for the study.

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

[3] J. Flack, P. Harman, S. Fox, “Low bandwidth stereoscopic image encoding and transmission”, Proceedings of the SPIE: Stereoscopic Displays and Virtual Reality Systems X, Vol. 5006, pp. 206-214, Santa Clara, CA, USA, 21-24 January 2003.

[4] L. Zhang, D. Wang & A. Vincent, “Reliability Measurement of Disparity Estimates for Intermediate view Reconstruction”, ICIP, Vol.3, pp. 837-840, Rochester NY, USA, September 2002.

[5] L. B. Stelmach & W.J. Tam, “Stereoscopic Image Coding: Effect of Disparate Image-Quality in Left- and Right-Eye Views”, Signal Processing: Image Communication, Special Issue on 3D Video Technology, Vol.14, pp.111-117, 1999.

[6] W. J. Tam, L. B. Stelmach, D. Meegan, and A.Vincent, “Bandwidth Reduction for Stereoscopic Video Signals,” Proceedings of the SPIE: Stereoscopic Displays and Virtual Reality Systems VIII, Vol. 3957, pp. 33-40, San Jose, CA, USA, 22-28 January 2000.

[7] W. J. Tam, L. B. Stelmach, S. Subramaniam, “Stereoscopic Video: Asymmetrical Coding with Temporal Interleaving”, Proceedings of the SPIE: Stereoscopic Displays and Virtual Reality Systems VIII, Vol. 4297, pp. 299-306, San Jose, CA, USA, 20-26 January 2001.

[8] W. J. Tam, L. B. Stelmach, F. Speranza, & R. Renaud, “Cross-switching in Asymmetrical Coding for Stereoscopic Video,” Proceedings of the SPIE: Stereoscopic Displays and Virtual Reality Systems IX, Vol. 4660, pp. 95-104, San Jose, CA, USA, 19-25 January 2002.

[9] ITU-R Recommendation BT.500-7 (1974-1997). Methodology for the subjective assessment of the quality of television pictures.