Subjective Study on Asymmetric Stereoscopic Video with Low-pass Filtered Slices

Sima Valizadeh, *Member, IEEE*, Maryam Azimi, *Member, IEEE*, Xiaokang Li, Lino E. Coria, *Senior Member, IEEE,* and Panos Nasiopoulos, *Member, IEEE*

***Abstract*— This paper proposes a new method for asymmetric filtering of stereoscopic video. In traditional asymmetric stereoscopic video coding, the quality of one of the views is reduced while the other view is of original quality. However, this approach is not fair for people with one dominant eye. We propose to address this problem by reducing the quality of horizontal slices in both views. In our approach, we applied low-pass filtering to slices of both views while the corresponding slice in the other view is untouched. Subjective tests show that the quality, sharpness and depth of the low-passed filtered stereo video are close to the original one. We tested a different number of horizontal slices and various levels of filtering for two representative video sequences. Results show that low-pass filtering of the horizontal slices of both views with smoothing on the slice edges is an effective technique for asymmetric stereoscopic videos.**

***Index Terms*— stereoscopic video, asymmetric coding, binocular suppression.**

I. INTRODUCTION

3D television has been largely commercialized and is becoming very popular. 3D videos can be either coded as multiview or video plus depth in which a stereo pair can be rendered at the decoder. A special case of the former, known as stereoscopic video (where number of views =2) [1], shows two slightly different views of the same scene: one for the left eye and one for the right eye. Thus, the amount of data required to represent stereoscopic video is twice as much as the information needed to represent 2D (monoscopic) video. Therefore, effective compression techniques for stereoscopic video are of great importance, especially when the available bandwidth or memory is limited.

The goal of 2D video compression schemes is to remove spatial and temporal redundancies within a sequence of frames. In the case of stereoscopic video, the similarities between the two views are also taken into account to improve the compression efficiency [2], [3]. An auspicious technique for stereoscopic video compression known as asymmetric coding is based on mixed-resolution which involves reducing the quality of one of the views while keeping the other view at the original quality [4], [5]. This is based on the suppression theory of the binocular vision [6], which states that the human visual system (HVS) perceives high-detailed 3D content even when just one of the views is of high quality. Sharp edges in the high quality image mask the blur in the low quality view and the overall depth impression is close to the sharper view. Asymmetric coding for stereoscopic video is based on this characteristic of the human visual system.

Perkins [4] introduced the concept of mixed-resolution coding in 1992. In this approach, a low-resolution picture is presented to one eye and a high-resolution picture is presented to the other eye. The low-resolution picture is obtained by downsampling the rows and columns of the original picture by a factor of 4. Thus, the bitrate required for this mixed-resolution coding approach is only 6% more than the bitrate needed for a single high resolution sequence. The proposed approach uses bilinear interpolation for reconstructing the low-resolution image, since it is easy to implement. The brain fuses the mixed-resolution stereo pairs. According to subjective tests the final quality and sharpness perception is similar to that of the high-resolution picture.

In [5], the quality of mixed-resolution stereo video is examined to determine how close it is to that of an equal-resolution pair. It was shown that coding efficiency was improved by this approach. Mixed-resolution sequences using quarter and half resolution for one of the views were compared to full resolution stereo videos. Temporal averaging and drop-and-repeat frame modes were also tested for stereo and non-stereo videos. The performed subjective tests showed that, for the spatial filtered sequences, quality and sharpness were rated higher for stereoscopic videos than for monoscopic videos with the same amount of filtering. In addition, the quality and sharpness ratings for mixed-resolution stereoscopic video were close to the ratings for full resolution stereoscopic video. For the case of temporal filtering, subjective tests showed that there is a noticeable quality drop. The authors conclude that spatial filtering is more promising than temporal filtering.

Another asymmetric approach for 3D video was proposed in [7]. In this case, in addition to degrading the luminance quality of one of the views, the chroma information from this view is not transmitted. Instead, it is reconstructed at the decoder end using data from the chroma and luma channels of the other view. Results showed significant bitrate reduction while users found the quality of the asymmetric 3D video sequences comparable to those of the original stereoscopic videos.

While the proposed asymmetric coding approaches have shown to reduce the amount of bandwidth or memory required for the transmission or storage of stereoscopic video, they fail to consider the issue of eye dominance [8]. If the high quality view is seen through the subject's weak eye, the overall impression of the 3D video will not resemble that of the high quality sequence. In addition, sustained imbalance in the two views may cause visual fatigue. One way of reducing this imbalance is to interleave low and high quality views in time [9], i.e., to reduce the quality of one view for a certain number of frames and then reduce the quality of the other view for the following time interval. However, results in [9] show that the cross-switch is noticeable and annoying unless it occurs at scene cuts.

In this paper, we propose a novel method for asymmetrically reducing the quality of stereoscopic videos. We divide the frames of each view into slices and apply low-pass filtering to the odd slices of one of the views and the even slices of the other view. We ran subjective tests to quantify the perceived sharpness, depth and quality of the resulting stereoscopic videos and compared these results to those of the original videos. Performance evaluations showed that despite the quality degradation (up to a threshold) in both views, the perceived sharpness, quality and depth of the stereo pair was close to those of the original-quality stereo pair.

The rest of the paper is organized as follows: Section II describes our method. Section III explains our test setup. Then, in section IV, we present and discuss our results. Finally, conclusions are made in section V.

II. OUR PROPOSED METHOD

In our method, we examined the perceived sharpness, depth and quality of stereoscopic videos after low-pass filtering alternate horizontal slices in the right and left views. A large variety of filter levels and sizes of horizontal slices were considered. Fig 1 shows one example where the odd slices of the left view and the even slices of the right view are filtered. We considered only horizontal slices since the horizontal disparity between the two views has the potential of causing the same object to be filtered in both views in the case of other directions (e.g., vertical).

To reduce the discontinuity at slice edges, we smoothed the edges by reducing the strength of filtering along the edges. In order to do that, we made the filtering strength follow a bell-shaped pattern so that the center of the slice is strongly filtered compared to the edges.

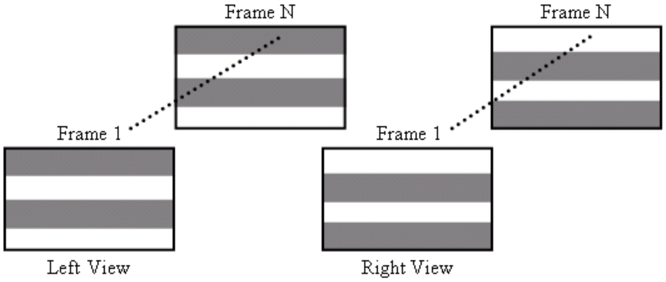


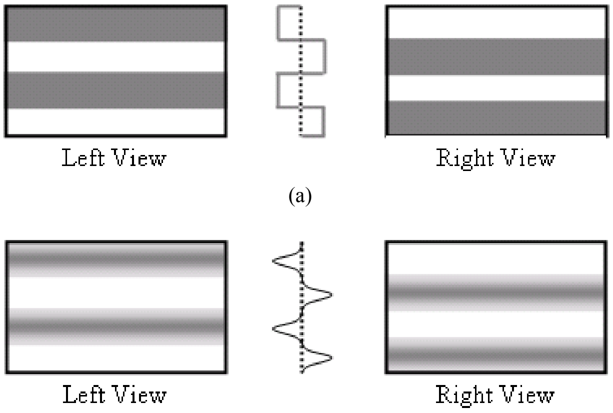
Fig. 1. Illustration of an example of our filtering pattern. Grey areas represent filtered slices in each view of the stereoscopic video.

In our study, we considered several numbers of horizontal slices in each frame: 2, 4, 10, 40 and 72. Performance evaluations have shown that 10 slices per frame provide the best visual quality for all different levels of low-pass filtering, with a fair distribution of reduced quality in both views. An excessive number of slices – such as 40 and 72 - seemed to annoy the viewers.

Our choice of a low-pass filter was a 15x15 Gaussian filter which has the following form:

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} exp^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

We used Gaussian filter since it is a typical and well-known low-pass filter. We picked a window size of 15x15 in order to generate filters for the different slice sizes with less and more strength by just changing the sigma in (1). In the first set of our tests, the strength of our filter follows a pulse-train pattern, which allows us to apply it to every other slice while keeping the in-between slices at their original quality. In the second set of our tests, we ensure a smooth transition between filtered and unfiltered slices by applying weaker filtering at the edges compared to the center of the slice. We control the strength of our Gaussian filter generating different sigma values from a Bell function. For the pixels near the slice edges, we applied weak filters with sigma close to zero. As the pixels get further from the edges, the sigma of the filter increases such that the strongest filtering is applied to the pixels in the center of every slice. Fig 2(a) and Fig 2(b) show the filtering pattern of the right and left views in the first and the second set of tests, respectively.



(a)

(b)

Fig. 2. Smoothing edges. (a) shows left and right frames with unsmoothed edges. (b) shows left and right frames with smoothed edges.

III. TEST SETUP

A. Video Sequences

We considered two representative stereo video sequences for our tests. The first video, "Mother and Kid", includes a lady standing with her kid. This video was taken outdoors and it contains different levels of details such as human faces and textures such as bushes and grass. The second video, "Two Dolls", includes two dolls being moved in front of the camera.

This video was shot indoors and it has less detail than the "Mother and Kid". Both sequences are silent, Each sequence is 10 seconds long with 30 fps. The resolution of each view is 1920 × 1080 pixels. Our test videos were captured with two identical HD camcorders (1080i, 60Hz, NTSC) set up in parallel. Figs. 3(a) and 3(b) show the first frame of the right view of Mother and Kid and Two Dolls, respectively.

(a)

(b)

Fig. 3. The first frame of our two video sequences. (a)First video sequence: Mother and Kid (b) Second video sequence: Two Dolls. These frames are the right view of our stereoscopic videos

B. *Test Design*

We applied four different levels of Gaussian filter which were produced by four different sigmas of 1, 3, 10 and 30, to each of our test sequences. By adding up the original video to these four low-pass filtered videos, we came up with five different versions for each sequence, resulting in ten unique stereo sequences.

In order to see how people perceive the quality degradation obtained by the same filters in 2D and 3D videos, we used the left view of each sequence as their 2D version and applied the above-mentioned filters to them. This resulted in 5 unique non-stereo test videos for each sequence.

As a next step, in order to have smoothed slice edges, we considered three different sigma values (3, 10 and 30) for the maximum filter strength. That strength is applied at the center row of each slice and gradually decreases to the minimum sigma value of 1 for pixel rows closer to the edges. This implementation follows the pattern mentioned in Section II. This step brought about three unique test videos for each sequence. In summary, all the above test videos resulted in eight stereoscopic test sequences as well as five monoscopic ones for each sequence. We asked the viewers to rate the overall quality, depth and sharpness of the test sequences after viewing the original one. All test sequences were shown in random order and the subjects were not aware of the test objectives. Table 1 summarizes the parameters used in our experiments.

Table 1. Parameters employed for our experiments

Filter type	Gaussian	
Number of horizontal slices	10	
Display method	Stereo and non-stereo	
Video sequences	*Mother and Kid* (outdoor) *Two Dolls* (indoor)	
Filter strength	Pulse Pattern Sigma of 1, 3, 10 and 30	Bell Shape Max sigma of 3, 10 and 30

C. *Viewers*

We showed our tests to 14 viewers. These viewers were between 23 and 38 years old with mean age of 28. Gender distribution was not controlled. All viewers were screened for visual acuity, color vision and contrast sensitivity. Only viewers who passed the screening participated in the experiment.

D. *Display*

We used a 65" 3D HD TV with 16:9 aspect ratio to show the videos to the viewers. We inserted a 10-second grey field between test sequences to allow the viewers' eyes to rest and also to give them enough time to rate perceived sharpness, depth and quality of the videos. The viewers' distance from the display was 4 times the height of the display. The room, in which we conducted the tests, was consistent with the ITU-R recommendation 500. The duration of the test was approximately 12 minutes for each participant.

E. *Test Assessment*

We asked the viewers to rate the sharpness, overall quality and depth perception for the stereoscopic video sequences, whereas for non-stereo videos, they were asked to rate only the quality and sharpness on a vertical rating scale.

The actual scale used in the tests had five equal-length labels: Excellent, Good, Fair, Poor and Bad. We used linear transformation from the scale to numbers between 0 and 100. These numbers were used to calculate the average ratings over the viewers for each video sequence. Ratings were made using the double-stimulus continuous-quality method described in ITU-R recommendation 500. The original video was shown to viewers prior to each modified video.

IV. RESULTS AND DISCUSSION

Fig. 4 shows the result of our subjective test for the "Mother and Kid" video sequence (video 1). The vertical axis shows the averaged ratings for both the stereo and non-stereo video sequences. The horizontal axis shows the sigma value of the Gaussian filter applied to the sequences. Sharpness and quality are shown in the top and bottom plots, respectively, while depth perception for the stereo video sequence is shown in the middle figures. Viewers' evaluations of the non-stereoscopic (2D) videos provide an indication of how strong the filtering is and how it affects the quality and sharpness of the non-stereo tested content.

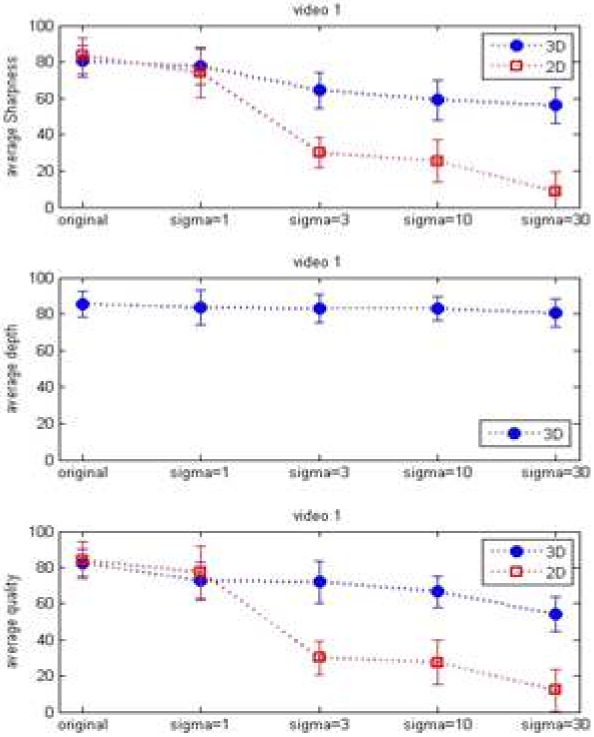


Fig. 4. Sharpness, depth and quality of video 1 averaged over the viewers.

An overall observation for video 1, "Mother and Kid", is that the quality and sharpness of the low-pass filtered stereo are much better than those of the low-pass filtered non-stereo videos. This is because the high quality slices in one view mask the blur in the low-pass filtered slices in other view in the case of stereoscopic videos. This does not apply to monoscopic videos since there is only one view.

Fig. 4 also implies that the quality and sharpness of the low-pass filtered stereo video are rated close to those of the original video up to a threshold in filtering strength. These results indicate that we can low-pass filter alternate slices of both views without significantly reducing the overall perceived quality of stereo pair. In this case, we may conclude that a Gaussian filter of 15x15 size, and sigma = 3 is a safe bound where most people cannot perceive the quality degradation. Beyond this point, we observe degradation in the perceived quality and sharpness.

Fig. 5 shows the results for the second video sequence, "Two Dolls". Similar to the case of the "Mother and Kid", the sharpness and quality of the low-pass filtered stereo videos are rated higher than those of the low-pass filtered non-stereo counterparts. Additionally, sharpness and quality of filtered stereo pairs are rated close to original stereo for up to filter-strength sigma = 3. Here the sharpness quality seems to be just a bit higher than the one for the "Mother and Kid" This may be due to the less relevant details (faces and textures) that this video has compared to "Mother and Kid".

We can also notice from the depth plots of both Fig. 4 and Fig. 5 that low-pass filtering will not affect the perceived depth in stereo pairs and it has remained unchanged for all levels of filtering that we applied.

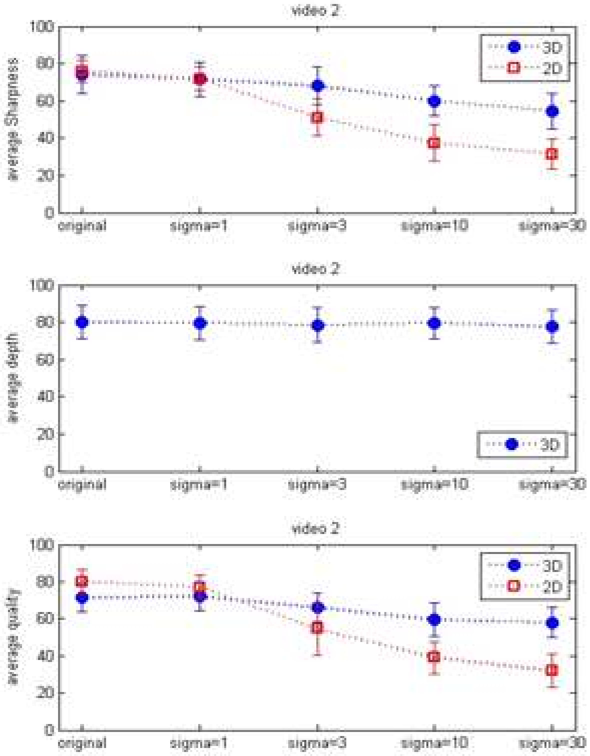


Fig. 5. Sharpness, depth and quality of video 2 averaged over the viewers.

Our next test is designed to determine if better ratings can be achieved for quality and sharpness of the videos by gradually smoothing the slice edges. To achieve this, we applied strong filtering in the middle of each slice and reduced the amount of filtering as we got closer to the edges. Fig. 6 shows the results of our subjective tests on sharpness, depth and video quality for the first video sequence (Mother and Kid) with smooth edges and compares them to those with the original filter (unsmoothed edges). We observe that smoothing the slice edges results in slightly better stereo video quality, sharpness.

Fig.7 shows the results for the second video sequence (Two Dolls). For this sequence as well, our subjects rated the sharpness and quality of videos with smoothed edges slightly better than those without smoothing.

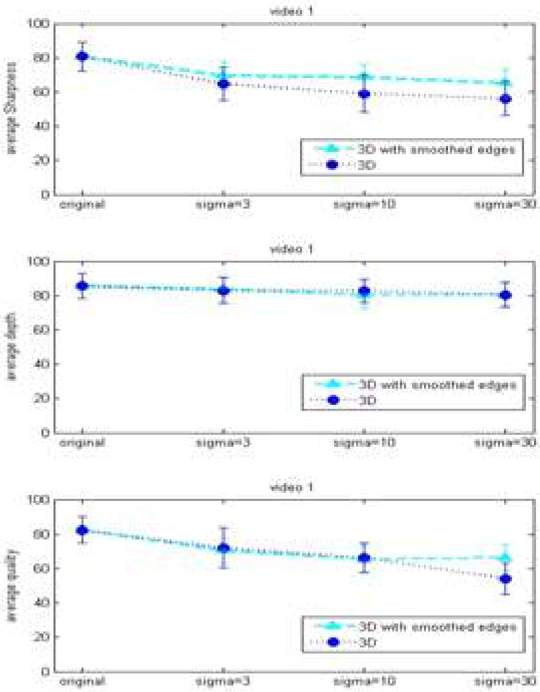


Fig. 6. Sharpness, depth and quality of video 1 averaged over the viewers. Smoothing the edges of the slices provide better quality and sharpness.

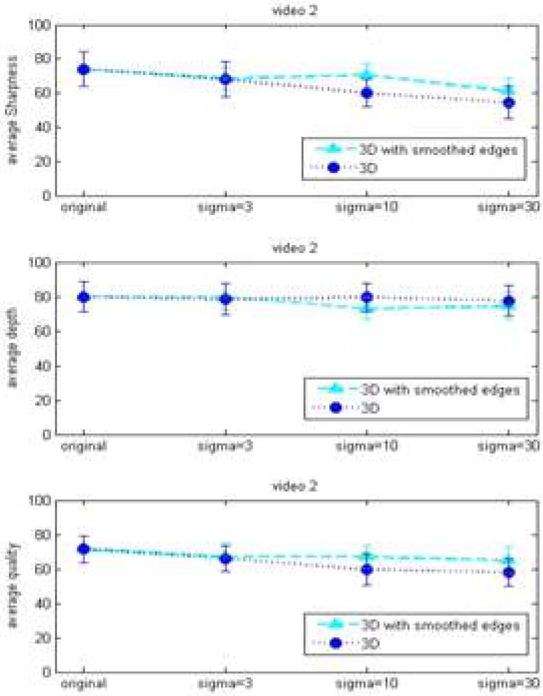


Fig. 7. Sharpness, depth and quality of video 2 averaged over the viewers. Smoothing the edges of the slices provide slightly better quality and sharpness.

V. CONCLUSION

We proposed a modified scheme for asymmetric coding of the stereoscopic video in which the videos are divided into horizontal slices in both the left and right views. Half of these slices are low-pass filtered while the corresponding slices in the other view are of original quality. We tested the perceived sharpness, quality and depth of the video sequences subjectively. Viewers rated sharpness and quality of our modified asymmetric videos close to those of the original stereoscopic videos up to a filtering strength threshold (Gaussian 15x15 with sigma = 3) while the same amount of filtering was quite apparent in the monoscopic videos. Our implementation of asymmetric video coding has the advantage of being fair for viewers with one dominant eye over the conventional asymmetric methods, because we divided the filtered slices between both views. Based on the fact that low-pass filtering is an effective technique to reduce transmission bandwidth and memory required for transmission and storage of the videos, our next step is to combine the proposed method with H.264 and MVC compression to determine the actual resulting bitrate performance.

VI. REFERENCES

[1] A. Smolic, K. Mueller, P. Merkle, P. Kauff, T. Wiegand, "An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution," Picture Coding Symposium, 2009. PCS 2009 , vol., no., pp.1-4, 6-8 May 2009.

[2] ISO/IEC 13818-2 AMD 3, "Mpeg-2 Multiview Profile, ISO/IEC JTC1/SC29/WG11 document No. N1366, September 1996.

[3] B. Bai, P. Boulanger, J. Harms, "An Efficient Multiview Video Compression Scheme," Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, vol., no., pp.836-839, 6-6 July 2005.

[4] M. G. Perkins, "Data compression of stereopairs," IEEE Trans. Commun., vol. 40, pp. 684–696, Apr. 1992.

[5] L. Stelmach, W. J. Tam, D. Meegan, A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution, "Circuits and Systems for Video Technology, IEEE Transactions on, vol.10, no.2, pp.188-193, Mar 2000.

[6] B. Julesz, "Foundations of Cyclopean Perception," Chicago, IL: Univ. Chicago Press, 1971.

[7] Shao Feng, Jiang Gangyi, Wang Xu, Yu Mei, Chen Ken, "Stereoscopic video coding with asymmetric luminance and chrominance qualities," Consumer Electronics, IEEE Transactions on, vol.56, no.4, pp.2460-2468, November 2010.

[8] Michael Reiss, and Gilfe Reiss, "Ocular Dominance: Some Family Data," Laterality Journal, Psychology Press, volume 2, issue 1, pages 7-16, 1997.

[9] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: Asymmetrical coding with temporal interleaving," Stereoscopic Displays and Virtual Reality Systems VIII, Vol. 4297, pp. 299-306, 2001.