

SUBJECTIVE STUDY ON COMPRESSED ASYMMETRIC STEREOSCOPIC VIDEO

Payman Aflaki^a, Miska M. Hannuksela^b, Jukka Häkkinen^{b,c}, Paul Lindroos^d, Moncef Gabbouj^a

^aDepartment of Signal Processing, Tampere University of Technology, Tampere, Finland;

^bNokia Research Center, Tampere, Finland;

^cDepartment of Media Technology, Aalto University School of Science and Technology, Espoo, Finland

^dDepartment of Psychology, University of Helsinki, Helsinki, Finland

ABSTRACT

Asymmetric stereoscopic video coding takes advantage of the binocular suppression of the human vision by representing one of the views with a lower quality. This paper describes a subjective quality test with asymmetric stereoscopic video. Different options for achieving compressed mixed-quality and mixed-resolution asymmetric stereo video were studied and compared to symmetric stereo video. The bitstreams for different coding arrangements were simulcast-coded according to the Advanced Video Coding (H.264/AVC) standard. The results showed that in most cases, resolution-asymmetric stereo video with the downsampling ratio of 1/2 along both coordinate axes provided similar quality as symmetric and quality-asymmetric full-resolution stereo video. These results were achieved under same bitrate constrain while the processing complexity decreased considerably. Moreover, in all test cases, the symmetric and mixed-quality full-resolution stereoscopic video bitstreams resulted in a similar quality at the same bitrates.

Index Terms— Asymmetric stereoscopic video, mixed resolution, subjective evaluation

1. INTRODUCTION

Multiview Video Coding (MVC) [1], recently standardized as an annex to the widely used Advanced Video Coding (H.264/AVC) standard [2], has potential to become a popular compression format for stereoscopic video. While MVC provides a considerable compression improvement compared to H.264/AVC simulcast mainly thanks to inter-view prediction, further compression without compromising the visual quality is desirable for many applications. One branch of research for further compression improvement in stereoscopic video is known as asymmetric stereoscopic video coding, in which one of the two views is coded with a lower quality compared to the other one. This is attributed to the widely believed assumption of the binocular suppression

theory that the Human Visual System (HVS) fuses the two images such that the perceived quality is close to that of the higher quality view. In mixed-resolution (MR) stereoscopic video, one of the views is downsampled before its compression.

In [3], the mixed-resolution approach was compared with a quality-asymmetric approach, in which the transform coefficients of one of the coded views were quantized coarsely. It was found that the perceived quality of the mixed-resolution videos were close to that of the higher-resolution view, while the perceived quality of the quality-asymmetric video was approximately equal to the average of the perceived qualities of the two views. The impact of quantization was verified in [4], which concluded that the perceived quality of coded equal-resolution stereo image pairs was approximately the average of the perceived qualities of the high-quality image and the low-quality image of the stereo pairs. A subjective evaluation for full- and mixed-resolution stereo video coding on a 32-inch polarization stereo display and on a 3.5-inch mobile display was presented in [5]. The results revealed that full-resolution (FR) was rated better for uncoded sequences in 79% of the test cases, while MR where one of the views was downsampled to a quarter resolution was rated better in 57% of the test cases for H.264/AVC-coded sequences. Moreover, it was concluded that 30% to 35% of the overall bitrate should be devoted to the lower resolution view.

This paper presents a set of subjective tests of quality- and resolution-asymmetric stereoscopic video coded with H.264/AVC. The tests attempt to reveal whether asymmetric stereoscopic video brings a compression improvement compared to symmetric stereoscopic video and whether mixed-quality or mixed-resolution is subjectively preferred. The comparison was performed using multiple bitrates to reveal performance dependencies on bitrate. Furthermore, limits of acceptable downsampling ratios in mixed-resolution stereoscopic video were tested.

This paper is organized as follows. In Section II, the test setup is explained. Section III presents and discusses the results. Finally, Section IV provides the conclusions.

2. TEST SETUP

2.1. Preparation of the Test Stimuli

The tests were carried out using four sequences: Undo dancer, Dog, Pantomime, and Newspaper. Undo dancer, exemplified in Fig. 1, is a synthetically created photorealistic multiview sequence including a dancing person, reproduced from a motion capture. The other three sequences are common test sequences in the 3D Video (3DV) ad-hoc group of the Moving Picture Expert Group (MPEG). No audio track was available for any of the test sequences. Views corresponding to camera separations of 4 cm and 5 cm were selected for Undo dancer and the rest of the sequences, respectively, based on a trial test.

Three types of sequences were tested:

1. Full-resolution with symmetric quality in both views
2. Full-resolution with asymmetric quality between the views caused by different quantization steps of transform coefficients
3. Mixed-resolution with asymmetric quality

The uncompressed full-resolution sequences were generated by downsampling both the left and right view image sequences from their original resolution to the “Full” resolution mentioned in Table I. The “full” resolution was selected to occupy as large area as possible on the used monitor with a reasonable downsampling ratio from the original resolution.

The mixed-resolution uncompressed sequences were generated from the full resolution ones by further downsampling the left view. Downsampling ratios 1/2 and 3/8 were symmetrically applied horizontally and vertically. As eye dominance was not shown to have an impact which view is provided with a better quality [4], only one set of MR sequences was prepared. The filters included in the JSVM reference software of the Scalable Video Coding standard were used in the downsampling and upsampling operations [6].

Views were independently coded using H.264/AVC in order to treat the FR and MR cases as equally as possible. Moreover, no inter-view prediction has been standardized

Table I. Resolutions of the sequences

	<i>Full</i>	$\frac{1}{2}$	$\frac{3}{8}$
Undo Dancer	960x576	480x288	360x216
Dog	768x576	384x288	288x216
Pantomime	768x576	384x288	288x216
Newspaper	768x576	384x288	288x216

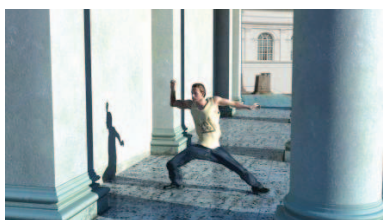


Fig. 1 a frame of one of the views of Undo dancer sequence

Table II. QP selection for asymmetric stereo bitstreams. a-b represents QP=a for left view and QP=b for right view

a) FR asymmetric bitstreams

<i>QP</i>	<i>39-39</i>	<i>35-35</i>	<i>30-30</i>	<i>25-25</i>
Undo Dancer	42-36	38-32	32-28	27-23
Dog	41-37	27-33	32-28	27-23
Pantomime	42-36	37-33	33-27	28-22
Newspaper	42-36	37-33	32-28	27-23

b) MR bitstreams with downsampling ratio of 1/2

<i>QP</i>	<i>39-39</i>	<i>35-35</i>	<i>30-30</i>	<i>25-25</i>
Undo Dancer	33-36	30-32	25-28	20-23
Dog	33-37	30-33	24-28	19-23
Pantomime	34-36	31-32	24-28	20-22
Newspaper	33-36	30-32	24-28	20-23

for a MR coding scheme, and we wanted to avoid the use of non-standardized codecs to provide as generally applicable results as possible.

The quality and bitrate of H.264/AVC bitstreams is controlled by the quantization parameter (QP). In order to get results from a large range of qualities and compressed bitrates, four constant quantization parameter (QP) values, 25, 30, 35, and 39, were selected for symmetrically compressed FR sequences. Fig. 2 includes the bitrates for different test sequences resulting from this QP value selection. A number of candidate asymmetric FR and MR bitstreams were generated, each having a bitrate within 5% of the bitrate of the corresponding symmetric full-resolution bitstream. The QP of a view was kept unchanged throughout the sequence in order to avoid any consequences of time-varying quality on the results. FR sequences with asymmetric quality were created by decreasing the QP for one view and increasing it for the other one. Table II (a) presents these selected QP values. Consequently, a large variety of compressed MR combinations were considered and the best combinations were selected in expert viewing for the actual subjective viewing test by naive viewers. Table II (b) summarizes the QP selections for the downsampling ratio of 1/2. These selections of QP values caused the bitrates of the lower resolution view to vary from 33% to 39% relative to the bitrate of both views together. In addition, the uncompressed FR and MR sequences were included in the viewed sequences to obtain a reference point for the highest perceived quality of a particular sequence.

2.2. Test procedure

Test clips were displayed on a 24" polarizing stereoscopic screen having a total resolution of 1920x1200 pixels and a resolution of 1920x600 per view when used in the stereoscopic mode. The on-screen lighting was less than 20

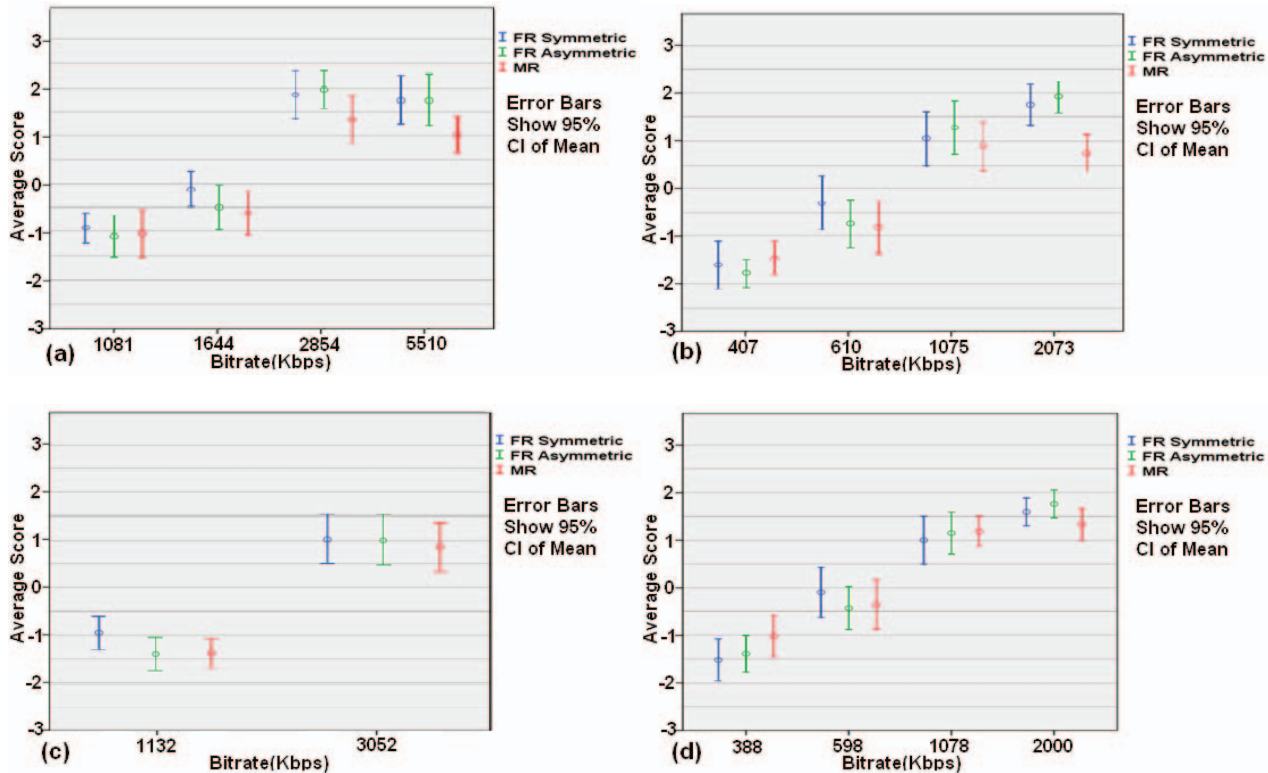


Fig. 2 Results of compressed MR subjective tests for sequences: a) Undo dancer, b) Newspaper, c) Pantomime, d) Dog

lux in order to prevent reflections on the screen. The viewing distance was fixed to 70 cm.

The duration of a viewing session had to be limited to one hour to avoid viewers becoming exhausted. Hence, the experiment was split into two sessions, where 9 and 7 subjects attended the assessment tests respectively. None of the viewers attended both sessions. Test clips having the bitrate corresponding to QP values 30 and 39 were tested in one session, whereas the remaining test clips were tested in the other test session.

The gender distribution was not controlled, as in our experience; there are no statistical gender differences in image and video quality estimation. A majority of the test subjects in all tests were female with an average age of 21 years. All the participants were untrained, having no significant prior experience on stereoscopic video.

The following visual tests were conducted for all participants: far and near visual acuity, stereoscopic acuity (Randot test), contrast sensitivity (Functional Acuity Contrast Test), horizontal and vertical phoria (Maddox wing test [7]), near point of accommodation and convergence RAF gauge test [8], and the interpupillary distance. Viewers who were not found to have normal visual acuity and stereopsis were rejected.

The subjective test started with a combination of anchoring and training. Subjects were shown the extremes of the quality range of the stimuli to familiarize the participants with the test task, the contents, and the variation in quality they could expect in the actual tests that followed.

The test clips were presented one at a time in a random order and appeared twice in the test session. Each clip was rated independently after its presentation, and the time used for rating was not limited. The test subjects were asked to rate the test stimulus according to three criteria separately: the general image quality, naturalness and perceived depth [9]. An integer scale in the range of -3 to 3 was used for the rating. At the beginning of the test, the scales were presented orally by the test leader to the participants until they understood everything thoroughly. The viewers were instructed that -3 means “very bad” or “not natural”, 0 is “mediocre”, and 3 stands for “very good” or “very natural”.

3. RESULTS AND DISCUSSION

The average general image quality ratings are presented in Fig. 2. The other rating scales had a significant correlation to the general image quality ratings. Results of both testing sessions are merged into the same figure, even though they are not fully comparable due to different test stimuli and participants. The MR test clips of downsampling ratio 3/8 along both axes were clearly inferior to all other respective test cases and did not reach acceptable quality on any test sequence. Thus, the results of downsampling ratio 3/8 are not discussed further and are not included in Fig. 2 in order to keep the figure more readable. It turned out that the subjective ratings for Pantomime were not consistent across the test sessions. For clarity, the Pantomime results of only the first session are presented in Fig. 2.

Fig. 2 indicates that mixed-resolution stereoscopic video of downsampling ratio 1/2 along both coordinate axes performed close to full-resolution symmetric stereoscopic video. Moreover, it confirms that except for the highest bitrate of Newspaper, the rest of the test sequences achieved an overlap on their 95% confidence interval. However, the use of mixed-resolution coding can be justified in many applications by its lower computational complexity. It can be observed from Fig. 2 that the performance of mixed-resolution coding of downsampling ratio 1/2 depends on the input sequence to some extent. Further studies are needed to conclude the objective characteristics of the test sequences under which mixed-resolution stereoscopic video of downsampling ratio 1/2 can be used without any remarkable impact on the subjective quality. Moreover, we plan to study the impact of the compression gain provided by inter-view prediction, firstly, on subjective ratings relative to the ratings for the H.264/AVC simulcast and, secondly, on the relative performance between symmetric FR, quality-asymmetric FR, and MR stereoscopic video.

We derived the average luma peak signal-to-noise ratio (PSNR) for each bitstream. A decoded view of a lower resolution was upsampled before PSNR calculation. The average luma PSNR over both views of the quality-asymmetric full-resolution bitstreams was found to be very close to that of the symmetric full-resolution bitstreams, the absolute difference being only 0.1 dB on average. This finding is aligned with the earlier conclusions in [3] and [4] that the perceived quality of the quality-asymmetric video was approximately the mean of the perceived qualities of the two views. In the case of MR stereoscopic video, both blurring and blocking are involved. We analyzed the contribution of the average luma PSNR of the higher-resolution view of a MR bitstream relative to the average luma PSNR of both views of the respective symmetric FR bitstream, when the subjective qualities of the bitstreams were approximately equal. It can be seen from Fig. 3 that in low QP values, the right view contributed more to the quality measured by PSNR, which supports the conclusions in [3] that the perceived quality of the mixed-resolution videos were close to that of the higher-resolution view. The higher the QP value became, the more contribution the left

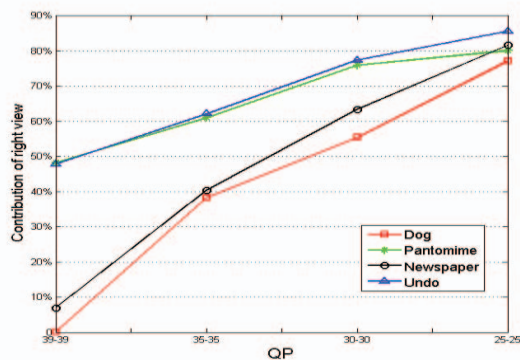


Fig. 3 Right view contribution to quality measured by PSNR

view had on the overall quality. Moreover, as can be seen in Fig. 3, the relative contribution of the right view depends on the sequence. We acknowledge that PSNR is not a proper measure for analyzing blocking and blurriness jointly and plan to perform further studies with other objective quality measures. However, the provided analysis of PSNR values was found to support the findings of earlier studies.

4. CONCLUSIONS

A set of subjective tests comparing symmetric full-resolution, quality-asymmetric full-resolution, and mixed-resolution stereoscopic video coding were presented. All bitstreams were simulcast-coded with H.264/AVC. The performance of symmetric and quality-asymmetric full-resolution bitstreams was approximately equal. Mixed-resolution stereoscopic video with downsampling ratio 1/2 along both coordinate axes performed similarly to the full-resolution bitstreams in most of the cases. Due to the lower required processing complexity, the use of mixed-resolution stereoscopic video can be considered in many applications. Mixed-resolution stereoscopic video with downsampling ratio 3/8 along both coordinate axes was found to be clearly inferior to all other tested coding arrangements and did not yield acceptable quality at any bitrate. Further work is planned to analyze the test results with selected objective metrics.

5. REFERENCES

- [1] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 786015, 13 pages, 2009. doi:10.1155/2009/786015.
- [2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," Mar. 2009.
- [3] W. J. Tam, "Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV," Joint Video Team document JVT-W094, Apr. 2007.
- [4] P. Seuntjens, L. Meesters, and W. IJsselstein, "Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Transactions on Applied Perception*, vol. 3, no. 2, pp. 95–109, Apr. 2006.
- [5] H. Brust, A. Smolic, K. Müller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices" 3DTV Conference, May 2009.
- [6] JSVM Software http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm
- [7] M. Rosenfield and N. Logan (Eds.), "Optometry: science, techniques and clinical management," Elsevier, 2009.
- [8] J. C. Neely, "The R.A.F. near-point rule," *British Journal of Ophthalmology*, vol. 40(10), pp. 636-637, Oct. 1956.
- [9] W. IJsselstein, H. de Ridder, and R. Hamberg, "Perceptual factors in stereoscopic displays: the effect of stereoscopic filming parameters on perceived quality and reported eyestrain," *Proc. SPIE*, vol. 3299, pp. 282-291, Jan. 1998.