



Author(s) Aflaki, Payman; Hannuksela, Miska M.; Homayouni, Maryam; Gabbouj, Moncef

Title Cross-asymmetric mixed-resolution 3D video compression

Citation Aflaki, Payman; Hannuksela, Miska M.; Homayouni, Maryam; Gabbouj, Moncef 2012. Cross-asymmetric mixed-resolution 3D video compression. 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), Zurich, 15-17 October 2012. 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video Piscataway, NJ, 1-4.

Year 2012

DOI <http://dx.doi.org/10.1109/3DTV.2012.6365439>

Version Post-print

URN <http://URN.fi/URN:NBN:fi:ty-201409231438>

Copyright © 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

CROSS-ASYMMETRIC MIXED-RESOLUTION 3D VIDEO COMPRESSION

Payman Aflaki^a, Miska M. Hannuksela^b, Maryam Homayouni^a, Moncef Gabbouj^a

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

^bNokia Research Center, Tampere, Finland;

ABSTRACT

Conventional mixed-resolution (MR) stereoscopic video where one view has full resolution (FR) and the other view has a lower resolution has shown to provide similar subjective quality compared to symmetric FR stereoscopic video while decreasing the encoding complexity considerably. In this paper, we propose a new cross-asymmetric mixed-resolution scheme where both views have a lower resolution compared to FR but downsampling is applied in horizontal direction for one view while the other view is vertically downsampled. Subjective results comparing the proposed scheme with the conventional MR and the symmetric FR schemes show that the perceived quality of the proposed scheme is higher than that of the two other schemes. Moreover, the computational complexity and memory requirements also are reduced thanks to the decreased number of pixels involved in the encoding and decoding processes.

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

1. INTRODUCTION

Two approaches for compressing stereoscopic video are common nowadays: frame-compatible stereoscopic video and Multiview Video Coding (MVC) [1]. The latter was standardized as an annex to Advanced Video Coding (H.264/AVC) standard [2]. In frame-compatible stereoscopic video, a spatial packing of a stereo pair into a single frame is performed at the encoder side as a pre-processing step for encoding and then the frame-packed frames are encoded with a conventional 2D video coding scheme. The encoder side may indicate the used frame packing format for example by including one or more frame packing arrangement supplemental enhancement information (SEI) messages as specified in the H.264/AVC standard into the bitstream. The decoder unpacks the two constituent frames from the output frames of the decoder and upsamples them to revert the encoder side's downsampling process and render the constituent frames on a 3D display. In contrast to frame packing, MVC enables any spatial resolution to be used in encoding and facilitates plain H.264/AVC decoders to produce single-view output without additional processing. Moreover, inter-view prediction presented in MVC provides a considerable compression improvement compared to frame packing and stereoscopic video representation using H.264/AVC simulcast. However, due to increased amount of data compared to conventional 2D video, further compression without perceivable subjective quality degradation is required in many applications.

One potential approach to achieve a better compression is to provide left and right views with different qualities referred to as asymmetric quality video where 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 [3] that the Human Visual System (HVS) fuses the two images such that the perceived quality is close to that of the higher quality view. Quality difference can be achieved by utilizing coarser quantization steps for one view and/or presenting stereoscopic video with MR where one view is downsampled prior to encoding. Considering that a smaller number of samples are involved in the coding/decoding process of MR stereoscopic video, it is expected to have lower processing complexity compared to the FR scheme.

Asymmetric stereoscopic video coding has been studied extensively over the years. For example, in [4] a set of subjective tests on encoded FR and MR stereoscopic videos were performed under the same bitrate constraint. The results show that the MR stereoscopic video with downsampling ratio 1/2, applied both vertically and horizontally, performed similarly to the FR in most cases. In [5], the MR approach was compared with a quality-asymmetric approach, in which the bigger steps were utilized for transform coefficients while coding one of the views. Results confirmed that perceived quality of the MR videos were close to that of the FR view. The impact of quantization was verified in [6], 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.

To approximate the perceived quality of the stereoscopic video, objective quality metrics often perform well. However, in the case of asymmetric stereoscopic video, there are two views with different qualities, and it has been found that objective quality assessment metrics face some ambiguity on how to approximate the perceived quality of asymmetric stereoscopic video [7].

This paper first describes a new cross-asymmetric MR compression technique and then evaluates its performance with a set of subjective tests. The proposed method is compared to compressed FR and conventional MR videos with different downsampling ratios applied to one view and along both coordinate axes. JM 17.2 reference software [8] of H.264/AVC is utilized as the encoder and the comparison is performed under the same bitrate constraints for two different bitrates.

This paper is organized as follows. In Section 2, the proposed MR scheme is presented while the test material and procedure are explained in section 3. Section 4 presents and discusses the results, and section 5 concludes the paper.

2. PROPOSED CROSS-ASYMMETRIC MIXED-RESOLUTION SCHEME

2.1 Overview

The traditional MR scheme performs downsampling on one view while the other view remains untouched having FR. In order to apply inter-view prediction between views of different resolution, a resampling process is required in the coding and decoding loop. As no such resampling is available in H.264/MVC, we present the proposed scheme in the context of H.264/AVC simulcast. Consequently, we also avoid any influence of non-standardized resampling and inter-view prediction algorithms on the results. Hence, we can be sure that the results are trustable and different performances are only due to utilization of different MR schemes, rather than different performance of utilized MR adaptive H.264/MVC codec. While the proposed method is presented for H.264/AVC simulcast environment, a frame packing scheme can also be designed, or a multiview codec with in-loop resampling and inter-view prediction can be applied.

2.2 Proposed MR scheme

In the proposed cross-asymmetric MR scheme, different resolutions for left and right views are utilized. Unlike the conventional MR scheme, we intend not to utilize any of two views in FR but downsample both views asymmetrically in such a way that horizontal and vertical downsampling ratios differ for the same view and the choice of the horizontal and vertical downsampling ratios is reversed for the other view. In other words, one view is downsampled more in the vertical direction while keeping more horizontal spatial information in that view. The other view is downsampled more in the horizontal direction. On the basis of the binocular suppression theory we expect the human visual system to perceive the picture in such a way that the higher quality information in each direction from the view where less downsampling along that direction was applied prevails. Figure 2 presents the general block diagram of the proposed MR scheme. W and H represent the width and height of the FR views, respectively, while a_1 , a_2 , b_1 , and b_2 are downsampling coefficients. In the proposed scheme, it is required that $a_1 > a_2$ and $b_1 < b_2$. Since the encoding is applied on downsampled views, after decoding and prior to the final presentation of stereoscopic video, the views will be upsampled to the FR.

Considering that eye dominance was shown not to have an impact on the perceived quality of MR stereoscopic videos [9], it is proposed that the decision on which the view should be more downsampled in horizontal/vertical direction is made based on spatial information (SI) [10] along each direction of each view. To calculate SI, a 3×3 Sobel filter is utilized (1) to emphasize horizontal edges using the smoothing effect by approximating a vertical gradient. To emphasize vertical edges, the transpose of the filter (h') will be applied.

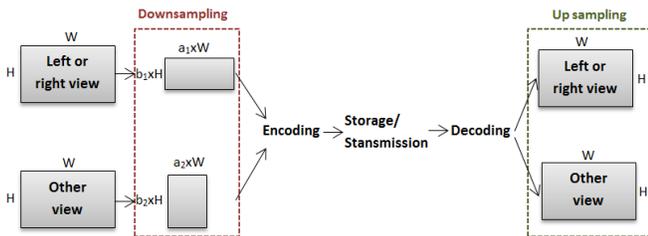


Figure 2. Proposed MR encoding/decoding scheme

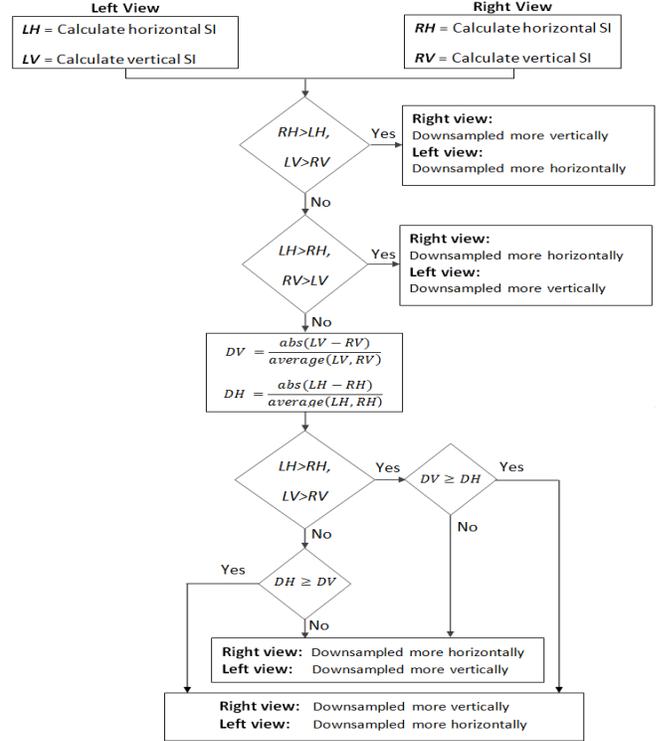


Figure 3. Flowchart of selecting the downsampling direction for left and right view.

$$h = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (1)$$

Based on the direction of the applied Sobel filter to the luma values of each image, SI will be calculated for the vertical or horizontal direction averaging over the magnitudes of the filtered image. Considering LV , LH , RV , and RH presenting SI of Left view in Vertical direction, Left view in Horizontal direction, Right view in Vertical direction, and Right view in Horizontal direction, respectively. The flowchart presented in Figure 3 shows how the decision of the downsampling direction of the left and right views is made. If LV is greater than RV and LH is smaller than RH , then the right view will be downsampled more in the vertical direction and the left view will be downsampled more in the horizontal direction. On the other hand, if LV is smaller than RV and LH is greater than RH , then the right view will be downsampled more in the horizontal direction while the left view is downsampled more in the vertical direction. If none of the above mentioned cases is valid, i.e. the left view has a higher SI in both directions ($LV > RV$ and $LH > RH$) or the right view has a higher SI in both directions ($LV < RV$ and $LH < RH$), the decision is made based on the normalized absolute difference levels, defined next. Considering the case where the left view has a higher SI in both directions, let us define the normalized absolute difference values as:

$$DV = \frac{abs(LV - RV)}{average(LV, RV)}$$

$$DH = \frac{abs(LH - RH)}{average(LH, RH)} \quad (2)$$

where $average(LX, RX) = \frac{LX+RX}{2}$ and DV and DH present the normalized absolute difference between the left and right views in vertical and horizontal direction, respectively. If $DV \geq DH$, then

the left view will be downsampled more in the horizontal direction and the right view will be downsampled more in the vertical direction. In the case where the right view has a higher SI in both directions, if $DV \geq DH$ then the right view will be downsampled more in the horizontal direction and the left view will be downsampled more in the vertical direction. The main idea behind the use of SI for downsampling the left and right views is that the highest combined amount of information in the downsampled MR stereoscopic video is preserved.

3. TEST SETUP

3.1 Test material

The tests were carried out using four sequences: Ballet, Breakdancer [11], Alt Moabit, and Book Arrival [12] with resolution 1024×768.

Five types of encoding schemes based on the resolution of left and right views were selected for the subjective test:

1. Anchor scheme: Full-resolution in both views (AS)
2. Conventional MR Scheme with downsampling ratio = 1/2, i.e. half resolution for one view in both directions and FR in the other view (CS1/2)
3. Proposed MR Scheme with downsampling ratio = 1/2, i.e. one view is downsampled only in vertical direction while the other view is only downsampled in the horizontal direction, the downsampling ratio is set to 1/2 for both cases (PS1/2)
4. Conventional MR Scheme with downsampling ratio = 1/4, (CS1/4)
5. Proposed MR Scheme with downsampling ratio = 1/4, (PS1/4)

The filters included in the JSVM reference software of the Scalable Video Coding standard were utilized in the downsampling and upsampling operations [13]. Moreover, views were independently coded using the reference JM 17.2 software in order to treat the FR and MR cases as equally as possible, as described in sub-section 2.1.

The quality and bitrate of H.264/AVC bitstreams is controlled by the quantization parameter (QP). In order to get results from a larger range of qualities and compressed bitrates, two constant QP values, 34 and 38, were selected for encoding in AS. Other schemes were encoded having a bitrate within 4% of the bitrate of the corresponding AS. The QP for left and right view was selected in such a way that bitrate ratio between the left and right view was close to one. This was due to the fact that we did not

want to affect the experiment by the selection of different QPs but limit the study to evaluate the performance of different applied downsampling schemes. The uncompressed FR sequences were included in the viewed sequences to obtain a reference point for the highest perceived quality of each particular sequence.

3.2 Test procedure

Test clips were displayed on a 46" polarizing stereoscopic screen having a total resolution of 1920×1200 pixels and a resolution of 1920×600 per view when used in the stereoscopic mode. Sequences were presented un-scaled with black background on the display fixing the viewing distance to 1.63 meter that is 4 times the height of the videos.

The duration of a viewing session was limited to ~35 minutes to avoid viewers becoming exhausted. In total 20 subjects (17 male and 3 female) attended the test. The average age of subjects was 26.5 years. All the participants were naïve users who had no previous experience on 3D video processing.

Subjective quality assessment was done according to Double Stimulus Impairment Scale (DSIS) method [14] and discrete unlabeled quality scale from 0 to 10 was used for quality assessment. Prior to the actual test, subjects were familiarized with test task, test sequences and with the variation in quality they could expect in the actual tests. The viewers were instructed that 0 stands for the lowest quality and 10 for the highest. The test clips were presented in a random order each clip was rated independently after its presentation. Prior to the participation in subjective viewing experiment, candidates were subject to a thorough vision screening. All participants had a stereoscopic acuity of at least 60 arc sec.

4. RESULTS AND DISCUSSION

The average and 95% confidence interval (CI) of subjective scores are presented in Figure 4. The naming of the encoding schemes is according to sub-section 3.1 and O represents the original FR uncompressed stereoscopic video.

It can be judged from the mean scores and confidence intervals presented in Figure 4 that the subjective quality of the higher bitrate was rated better in general compared to the lower bitrate. Moreover, the original uncompressed video had superior quality compared to other schemes. The observation on significant differences between the encoding schemes was further analyzed using statistical analysis as presented in the paragraphs below.

Non-parametric statistical analysis methods, Friedman's and

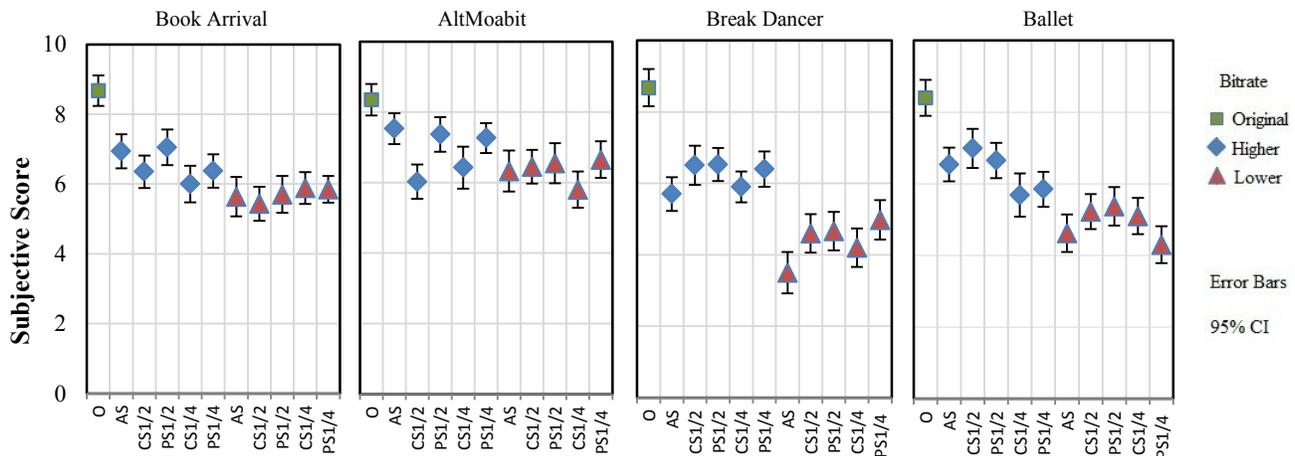


Figure 4. Viewing experience ratings

Table 1. Pairwise performance comparison of different coding schemes over all content

Coding scheme	Higher bitrate			Lower bitrate		
	Better	Similar	Worse	Better	Similar	Worse
AS	5	8	3	0	12	4
CS1/2	3	9	4	2	14	0
PS1/2	9	7	0	4	12	0
CS1/4	0	7	9	1	12	3
PS1/4	3	9	4	3	10	3

Table 2. Per view and total number of pixels involved in the coding/decoding process for different coding schemes

Coding Scheme	Number of Pixels		
	One view	Other view	Total
AS	$W * H$	$W * H$	$2 * W * H$
CS1/2	$\frac{W}{2} * \frac{H}{2}$	$W * H$	$\frac{5}{4} * W * H$
PS1/2	$\frac{W}{2} * H$	$W * \frac{H}{2}$	$W * H$
CS1/4	$\frac{W}{4} * \frac{H}{4}$	$W * H$	$\frac{17}{16} * W * H$
PS1/4	$\frac{W}{4} * H$	$W * \frac{H}{4}$	$\frac{W * H}{2}$

Wilcoxon's tests, were used as the data did not reach normal distribution (Kolmogorov-Smirnov: $p < .05$). Wilcoxon's test is applicable to measure differences between two related and ordinal data sets [15]. A significance difference level of $p < 0.05$ was used in our analysis.

Table 1 reports the performance analysis results of each coding scheme, as achieved by Wilcoxon's test, in a pairwise comparison to other schemes. For each coding scheme three values are reported per bitrate. First, the value in column *Better* provides the total number of cases in which the associated scheme was ranked significantly better than the other schemes. The second number reports the total number of cases where similar subjective quality to the other schemes was reported (*Similar*). Finally, the third value in column *Worse*, reports the number of cases in which the referred coding scheme provided a significantly worse rating compared to the other schemes. The next paragraphs discuss the performance of different coding schemes based on the statistics reported in Table 1.

In the higher bitrate the performance of PS1/2 was clearly superior, since in no comparison it was ranked worse than other schemes and in the majority of cases it was ranked better compared to other schemes. Moreover, CS1/4 performed worse since in the majority of comparisons it was ranked worse than other schemes while in none of the comparisons it was ranked better.

In the lower bitrate, the coding schemes performed closer to each other while the majority of comparisons resulted in a similar subjective quality. AS performed slightly inferior to others since it was never ranked better. Moreover, PS1/2 and CS1/2 were never ranked worse compared to other schemes; nevertheless, PS1/2 had slightly better performance compared to CS1/2 since it was ranked better in four cases compared to two cases for CS1/2.

In general, the results show that utilization of FR videos (as in AS) in that lower bitrate was not subjectively preferred and applying downsampling through different schemes provided a higher perceived quality. This is in agreement with the conclusion achieved in [4, 16]. Moreover, the subjective results confirm that PS1/2 performed the best in the lower and higher bitrates.

Next we compare the complexity of the coding/decoding schemes based on the number of pixels involved in the

coding/decoding process. If the width and the height of the FR views is represented with W and H , respectively, the total number of pixels for both views can be calculated as shown in Table 2. Based on the results presented in Table 2 the proposed methods (PS1/2 and PS1/4) introduce the least number of pixels for the coding/decoding process. Hence, along with superior subjective quality, lower complexity is another important advantage which justifies the utilization of the proposed coding scheme.

5. CONCLUSIONS

The paper proposes a mixed-resolution (MR) stereoscopic video coding scheme, where one view is horizontally downsampled while the other view is vertically downsampled at different rates. The proposed scheme was compared with symmetric full-resolution (FR) stereoscopic video as well as the conventional MR coding, where one view is downsampled along both coordinate axes while the other view is maintained at its original resolution. A series of subjective tests was conducted comparing the proposed scheme with conventional MR and symmetric FR schemes. The results show that proposed method outperforms the other methods while decreasing the computational complexity and memory requirements of the codec.

6. 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
- [2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," Mar. 2009.
- [3] R. Blake, "Threshold conditions for binocular rivalry," Journal of Experimental Psychology: Human Perception and Performance, vol. 3(2), pp. 251-257, 2001.
- [4] P. Aflaki, et al, "Subjective study on compressed asymmetric stereoscopic video," Proc. of Int. Conf. on Image Proc., Sep. 2010.
- [5] W. J. Tam, "Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV," Joint Video Team document JVT-W094, Apr. 2007.
- [6] P. Seuntjens, L. Meesters, and W. IJsselstein, "Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation," ACM Trans. on Applied Perception, vol. 3, no. 2, pp. 95-109, Apr. 2006.
- [7] P. W. Gorley, N.S. Holliman, "Stereoscopic image quality metrics and compression", Stereoscopic Displays and Virtual Reality Systems XIX, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 6803, January 2008
- [8] JM reference software: <http://iphome.hhi.de/suehring/tml/download>
- [9] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, "Impact of downsampling ratio in mixed-resolution stereoscopic video," Proc. of 3DTV-Conference, Jun. 2010.
- [10] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," 1999.
- [11] <http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload>
- [12] <ftp://ftp.hhi.de/HHIMPEG3DV/sequences/>
- [13] JSVM Software http://ip.hhi.de/imagecom_G1/save/downloads/SVC-Reference-Software.htm
- [14] ITU-R Rec. BT.500-11, Methodology for the subjective assessment of the quality of television pictures, 2002
- [15] H. Cooligan "Research methods and statistics in psychology," (4th ed.). London: Arrowsmith., 2004.
- [16] H. Brust, A. Smolic, K. Müller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices" 3DTV Conference, May 2009.