SIViP (2015) 9:331–345
DOI 10.1007/s11760-013-0439-0

ORIGINAL PAPER

Subjective quality assessment of asymmetric stereoscopic 3D video

Payman Aflaki · Miska M. Hannuksela ·
Moncef Gabbouj

Received: 15 May 2012 / Revised: 12 December 2012 / Accepted: 22 February 2013 / Published online: 16 March 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract In asymmetric stereoscopic video compression, the views are coded with different qualities. According to the binocular suppression theory, the perceived quality is closer to that of the higher-fidelity view. Hence, a higher compression ratio is potentially achieved through asymmetric coding. Furthermore, when mixed-resolution coding is applied, the complexity of the coding and decoding is reduced. In this paper, we study whether asymmetric stereoscopic video coding achieves the mentioned claimed benefits. Two sets of systematic subjective quality evaluation experiments are presented in the paper. In the first set of the experiments, we analyze the extent of downsampling for the lower-resolution view in mixed-resolution stereoscopic videos. We show that the lower-resolution view becomes dominant in the subjective quality rating at a certain downsampling ratio, and this is dependent on the sequence, the angular resolution, and the angular width. In the second set of the experiments, we compare symmetric stereoscopic video coding, quality-asymmetric stereoscopic video coding, and mixed-resolution coding subjectively. We show that in many cases, mixed-resolution coding achieves a similar subjective quality to that of symmetric stereoscopic video coding, while the computational complexity is significantly reduced.

Keywords Mixed resolution · Asymmetric stereoscopic · Stereoscopic 3D video · Subjective quality

1 Introduction

Stereoscopic video compression has gained importance during the recent years thanks to the recent advances in display technology. In many stereoscopic 3D video services and applications, the challenge is that the available bitrate or storage space is similar to that for monoscopic video, while the perceived temporal and spatial quality should also be similar to those for monoscopic video. Recent advances in video compression have alleviated the mentioned challenge to some extent. For example, the inter-view prediction enabled by the Multiview Video Coding (MVC) [1] annex of the widely used Advanced Video Coding (H.264/AVC) standard [2] has been shown to improve compression efficiency significantly compared to independent coding of the views. As an example, Merkle et al. [3] reported gains up to 3.2 dB and an average gain of 1.5 dB in terms of average luma peak signal-to-noise ratio (PSNR). However, further compression without compromising the visual quality is desirable in order to meet the bitrate and quality expectations of many applications. There are several other examples for video coding methods that aim to provide higher performance encoding to video content, for example, High Efficiency Video Coding (HEVC) [4] and a depth enhanced extension for MVC, abbreviated MVC+D, specifying encapsulation of MVC-coded texture and depth views into a single bitstream [5,6].

This work was supported by the Academy of Finland, (application number 129657, Finnish Programme for Centres of Excellence in Research 2006–2011).

P. Aflaki (✉) · M. Gabbouj
Department of Signal Processing,
Tampere University of Technology,
Tampere, Finland
e-mail: payman.aflaki@tut.fi

M. Gabbouj
e-mail: moncef.gabbouj@tut.fi

M. M. Hannuksela
Nokia Research Center, Tampere, Finland
e-mail: miska.hannuksela@nokia.com

Video compression is commonly achieved by removing spatial, frequency, and temporal redundancies. Different types of prediction and quantization of transform-domain prediction residuals are jointly used in many video coding standards to exploit both spatial and temporal redundancies. In addition, as coding schemes have a practical limit in the redundancy that can be removed, spatial and temporal sampling frequency as well as the bit depth of samples can be selected in such a manner that the subjective quality is degraded as little as possible.

One branch of research for obtaining compression improvement in stereoscopic video is known as asymmetric stereoscopic video coding, in which there is a quality difference between the two coded views. This is attributed to the binocular suppression theory [7]. It is assumed according to the binocular suppression theory that the HVS fuses the two images with different levels of sharpness such that the perceived quality is close to that of the sharper view [8]. This is because, in normal vision, there is some additional fusion to impulses from corresponding points of the two retinas. The correspondence of the retinal elements is completely rigid and un-changing; however, one of a pair of corresponding points tends to suppress the other and create the binocular suppression. In the next sections, we will cover several studies which have been exploiting binocular suppression in asymmetric stereoscopic video coding.

Asymmetry in quality between the two coded views can be achieved by one or more of the following methods:

(a) Mixed-resolution (MR) stereoscopic video coding, first introduced in [9], also referred to as resolution- asymmetric stereoscopic video coding. One of the views is low-pass filtered and hence has a smaller amount of spatial details or a lower spatial resolution. Furthermore, the low-pass filtered view is usually sampled with a coarser sampling grid, that is, represented by fewer pixels.

(b) Mixed-resolution chroma sampling [10]. The chroma pictures of one view are represented by fewer samples than the respective chroma pictures of the other view.

(c) Asymmetric sample-domain quantization [11]. The sample values of the two views are quantized with a different step size. For example, the luma samples of one view may be represented with the range of 0–255 (i.e., 8 bits per sample), while the range may be scaled to the range of 0–159 for the second view. Thanks to fewer quantization steps, the second view can be compressed with a higher ratio compared to the first view. Different quantization step sizes may be used for luma and chroma samples. As a special case of asymmetric sample-domain quantization, one can refer to bit-depth-asymmetric stereoscopic video when the number of quantization steps in each view matches a power of two.

(d) Asymmetric transform-domain quantization. The transform coefficients of the two views are quantized with a different step size. As a result, one of the views has a lower fidelity and may be subject to a greater amount of visible coding artifacts, such as blocking and ringing.

(e) A combination of different encoding techniques above.

The aforementioned types of asymmetric stereoscopic video coding are illustrated in Fig. 1. The first row presents the higher quality view which is only transform-coded. The remaining rows present several encoding combinations which have been investigated to create the lower quality view using different steps, namely, downsampling, sample-domain quantization, and transform-based coding. It can be observed from the figure that downsampling or sample-domain quantization can be applied or skipped regardless of how other steps in the processing chain are applied. Likewise, the quantization step in the transform-domain coding step can be selected independently of the other steps. Thus, practical realizations of asymmetric stereoscopic video coding may use appropriate techniques for achieving asymmetry in a combined manner as illustrated in Fig. 1e. Moreover, in [12], the subjective quality of mixed temporal resolution was assessed and compared to mixed spatial resolution on two test sequences having a resolution of 720 × 480. The paper concluded that at 1/2 temporal resolution, mixed temporal resolution performed worse than mixed spatial resolution with different downsampling ratios. Due to its inferior performance, mixed temporal resolution is not considered in the subsequent parts of this paper.

This paper attempts to provide answers to two research questions: Firstly, to what extent downsampling can be applied for mixed resolution stereoscopic video? Secondly, what are the constraints which limit the preference of utilizing asymmetric coding achieved with different coding schemes compared to symmetric coding? These research questions were studied using systematic subjective testing, because no commonly acceptable objective metrics are available for approximating the perceived quality of asymmetric stereoscopic video.

The rest of this paper is organized as follows: A brief overview of the relevant literature is presented in Sect. 2. Section 3 presents a study of downsampling constraints for MR stereoscopic video. Asymmetric stereoscopic video achieved by mixed-resolution coding or asymmetric transform-domain quantization is subjectively assessed and compared to symmetric stereoscopic video coding in Sect. 4. The primary target in the study presented in Sect. 4 is to reveal whether asymmetric stereoscopic video coding outperforms symmetric stereoscopic video coding in terms of subjective quality when the same bitrate is used for both. Furthermore, the study compares the subjective quality achieved by the

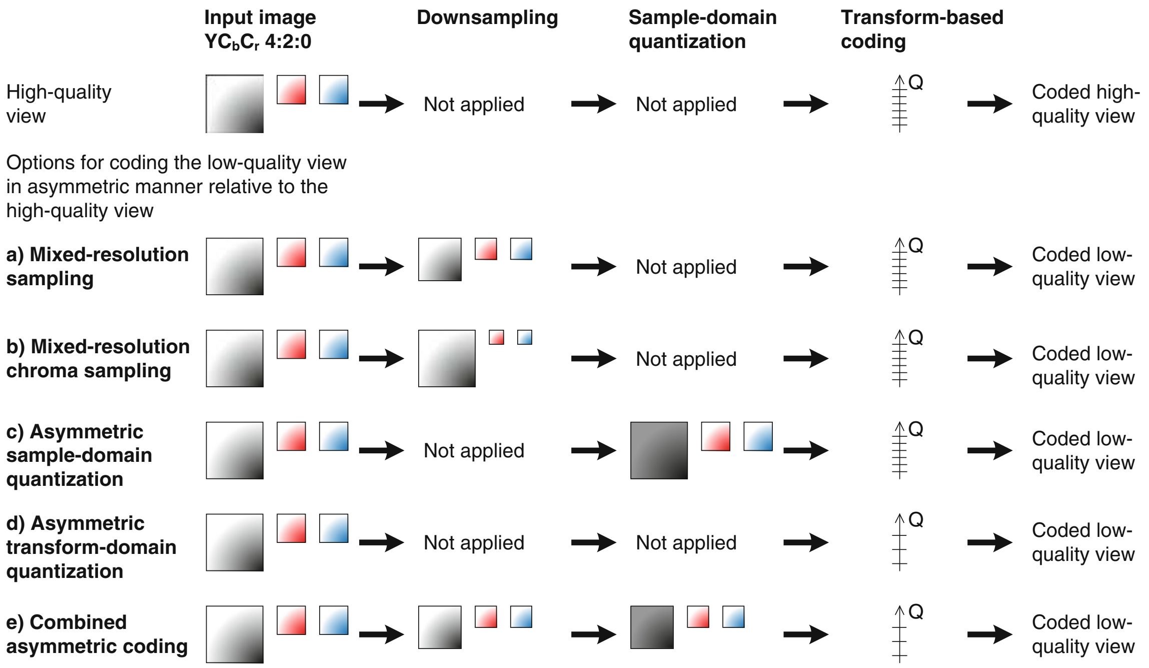

Input image
YC_bC_r 4:2:0
Downsampling
Sample-domain quantization
Transform-based coding
High-quality view
Not applied
Not applied
Q
Coded high-quality view
Options for coding the low-quality view in asymmetric manner relative to the high-quality view
a) Mixed-resolution sampling
Not applied
Coded low-quality view
b) Mixed-resolution chroma sampling
Not applied
Coded low-quality view
c) Asymmetric sample-domain quantization
Not applied
Coded low-quality view
d) Asymmetric transform-domain quantization
Not applied
Not applied
Coded low-quality view
e) Combined asymmetric coding
Coded low-quality view

Fig. 1 Illustrative examples of different types of asymmetric stereoscopic video coding

mentioned two asymmetric stereoscopic video coding methods. Finally, conclusions are provided in Sect. 5.

2 Literature review

2.1 Uncompressed mixed-resolution stereoscopic video

The subjective impact of uncompressed MR sequences at downsampling ratios of 1/2 and 1/4 applied both horizontally and vertically was studied in [12]. A combination of a data projector and shutter glasses were used as the viewing equipment with a viewing distance equal to 4H, where H was 91.5 cm. It was found that the perceived sharpness and the subjective image quality of the MR image sequences were nearly transparent at the downsampling ratio of 1/2 along both coordinate axes but dropped slightly at the ratio of 1/4.

The study presented in [13] included a subjective evaluation for full- and mixed-resolution stereo video on a 32-inch polarization stereo display and on a 3.5-inch mobile display. One of the views in the MR sequences was downsampled to half the resolution both horizontally and vertically. The results revealed that uncompressed full-resolution (FR) sequences were preferred in 94 and 63 % of the test cases for the 32- and 3.5-inch displays, respectively. Moreover, different resolutions for the symmetric stereo video and the higher-resolution view of the MR videos were tried out, while the downsampling ratio in the MR videos was always 1/2 both horizontally and vertically. It was found that the higher the resolution, the smaller the subjective difference is between FR and MR stereoscopic video. An equivalent result was also discovered as a function of the viewing distance by changing the distance from 1 to 3 m—the greater the viewing distance, the smaller the subjective difference becomes between FR and MR.

An obvious question related to MR stereoscopic video is whether people having a different ocular dominance perceive the quality of the same MR stereoscopic image sequence differently. However, it has been discovered in several studies, such as [14] and [15], that subjective ratings of MR image sequences are not statistically impacted by eye dominance.

In this paper, along with providing results completing those included in [12] and [13] under our test setup, we also determine the extent of the downsampling ratio that can be applied to one view before the low-resolution view starts to dominate in the perceived quality.

2.2 Compressed asymmetric stereoscopic video

The quantization of transform coefficients may result into perceivable coding artifacts and also often suppresses high-frequency transform coefficients and hence essentially reduces spatial resolution. Consequently, there is a tradeoff between spatial resolution of images used as input for the encoding and the quantization step size. The tradeoff between the selections of spatial resolution and the quantization step size in JPEG coding of monoscopic images was studied in [16].

Saygili et al. [17] addressed the questions what should be the level of asymmetry and whether asymmetry should be achieved by spatial resolution reduction or SNR reduction by presenting subjective assessment results. They used two test setups. The first setup included polarized glasses and a pair of projectors each having resolution of 1,024 × 768. The viewing distance was set to approximately 3 m from the screen. In the second setup, a parallax barrier autostereoscopic display was used. The authors concluded that when the reference view is encoded at a sufficiently high quality, the auxiliary view can be encoded above a low-quality threshold without a noticeable degradation on the perceived quality. This low-quality threshold was 31 and 33 dB in terms of average luma PSNR for the parallax barrier and the polarized projection displays, respectively. Moreover, their results showed that, at high bitrates, asymmetric coding with SNR scaling achieved the best perceived quality, while at low bitrates, asymmetric coding with spatial scaling achieved the best perceived quality. In between these two thresholds, symmetric coding was preferred over asymmetric coding.

Tam [18] compared the MR approach 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 was 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 the quantization of transform coefficients was verified in [15], where it was 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 comparison among different compression methods was presented in [19] among which MR and symmetric stereoscopic video coded with H.264/AVC were compared. Forty-seven subjects assessed 6 sequences at two bitrates typically suitable for mobile devices. The downsampling ratio of 1/2 was used for the MR bitstreams. The viewing was performed on a mobile autostereoscopic display. At the higher bitrate, symmetric stereoscopic video outperformed MR in terms of subjective acceptance and satisfaction, while the methods performed similarly at the lower bitrate.

In Sect. 4 of this paper, a systematic subjective quality evaluation test comparing different methods of asymmetric stereoscopic video coding and symmetric stereoscopic video coding are presented. The results provide some indications under which bitrates and other conditions asymmetric stereoscopic video coding is beneficial and which parameter values, such as which downsampling ratios for MR stereoscopic video, should be used. This paper therefore supplements the earlier findings reviewed above.

3 Extent of downsampling for mixed-resolution stereoscopic video

3.1 Introduction

It is evident that there are limits on the amount of asymmetry that binocular fusion can successfully mask so that the perceived quality is closer to the quality of the higher-fidelity view. It is presumably easier to discover such limits in subjective tests when only one type of asymmetry is applied. Hence, studying uncompressed MR stereoscopic video in subjective tests makes it possible to assess such limits in resolution asymmetry between views and avoids the difficulty of analyzing the results of subjective experiences when views undergo multiple types of asymmetry. In this section, we seek to clarify as follows: "*under which viewing conditions uncompressed mixed-resolution stereoscopic video is similar to full-resolution symmetric stereoscopic video in terms of subjective quality*." The research question was tackled by performing a subjective quality evaluation study and analyzing the results. This section extends the discussion of the subjective experiment as reported in [20] by providing more technical detail, for example, angular width, visual horizontal angle, subjective scores, and PSNR of test materials. Section 3.2 introduces the used test material, while Sect. 3.3 presents the test setup. The results are presented and analyzed in Sect. 3.4.

3.2 Test material

A subjective test was performed to evaluate the subjective quality of MR stereoscopic video. The test was carried out using five sequences: undo dancer, dog, pantomime, champagne tower, and newspaper. All these sequences, presented in Fig. 2, 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. The duration of all sequences in all experiments was limited to 10 s. The user perception of video quality may vary between different content types; for example, viewers may perceive action sequences differently from slow moving sequences. In order to characterize the content of the sequences, spatial and temporal perceptual information were determined using spatial information (SI) and temporal information (TI) metrics [21], although they may not always correlate well with individual's perception experience. Considering these values, one can have a general approximation on the amount of details available in the video and how much temporal movement is expected during the content playback. The obtained SI and TI results are reported in Table 1.

For each sequence, we had the possibility to choose between several camera separations or view selections. This was studied first in a pilot test of 9 subjects. The test pro-

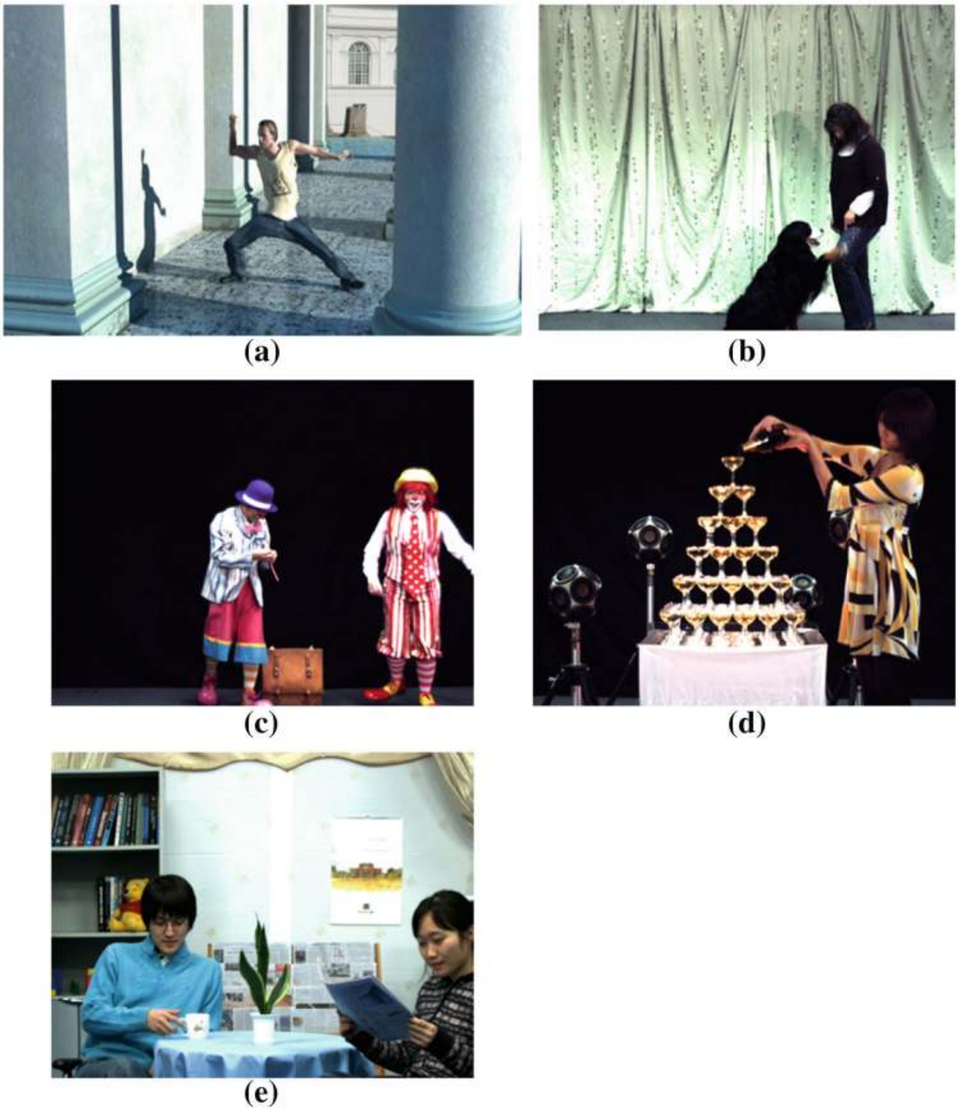


Fig. 2 **a** Undo dancer, **b** dog, **c** pantomime, **d** champagne tower, **e** newspaper

Table 1 Spatial and temporal complexity of sequences calculated using SI and TI metrics

Sequence	SI	TI
Undo dancer	98.6	23.0
Dog	90.7	23.6
Pantomime	108.3	47.0
Champagne tower	107.0	24.8
Newspaper	77.6	15.4

cedure of the pilot test was similar to that of the actual test presented in Sect. 3.3. The best average subjective viewing experience rating for undo dancer was obtained with the camera separation of 4 cm, while in the other tests, separations of 2, 6, 8, 14, and 26 cm dropped the average subjective viewing experience rating by less than 1 point on a 7-point scale. For other sequences, camera separations of 5, 10, 15, and 20 cm were tested and 5 cm separation provided the highest subjective ratings for all sequences.

Test clips were prepared as follows. Both the left and the right view image sequences were first downsampled from their original resolution to the "full" resolution presented in Table 2. The "full" resolution was selected to occupy the largest possible area on the used monitor (see Sect. 3.3) with a downsampling ratio of 1/2, 5/8, or 3/4. Moreover, the same downsampling ratio was along both directions to keep the pixel aspect ratio unchanged. To achieve the full-resolution (FR) sequences, downsampling ratio 1/2 and 3/4, were applied in both directions for undo dancer and newspaper, respectively, and 5/8 for the rest of the sequences. No cropping was applied in the conversion from the original resolution to the "full" resolution.

Two sets of test sequences were then generated, differing in whether the left view or the right view was downsampled

Table 2 Spatial resolutions and angular widths of sequences

	Original	Full	1/2	3/8	1/4	Angular width
Undo dancer	1,920 × 1,080	960 × 540	480 × 270	360 × 202	240 × 135	40.4°
Dog	1,280 × 960	800 × 600	400 × 300	300 × 225	200 × 150	34.1°
Pantomime	1,280 × 960	800 × 600	400 × 300	300 × 225	200 × 150	34.1°
Champagne	1,280 × 960	800 × 600	400 × 300	300 × 225	200 × 150	34.1°
Newspaper	1,024 × 768	768 × 576	384 × 288	288 × 216	192 × 144	32.8°

and subsequently upsampled. In other words, in the first set of sequences, the left view was downsampled to 1/2, 3/8, or 1/4 resolution and subsequently upsampled for rendering on the display, while the right view was kept at "full" resolution. In the second set, the right view was downsampled and subsequently upsampled, while the left view was kept at "full" resolution. This arrangement of preparing two sets of sequences was done so that we could study the effect of eye dominance on the subjective quality of asymmetric stereoscopic sequences. The tested downsampling factors were 1/2, 3/8, and 1/4 symmetrically along both coordinate axes. The resolutions of the test sequences are provided in Table 2. The filters included in the JSVM reference software of the scalable video coding standard were used in the downsampling and upsampling operations [22]. The default method 0 for down and upsampling was enabled for the process. For downsampling, a sine-windowed sinc-function designed to support an extended range of spatial scaling ratios, as required by Extended Spatial Scalability (ESS), was applied. For upsampling the Scalable Video Coding (SVC), normative upsampling method designed to support ESS was applied. This filter includes a 4 tap filter with coefficients [-3, 19, 19, -3] which is originally derived from the Lanczos-3 filter. This interpolation supports any inter-layer scaling ratios, which can also be different in horizontal and vertical.

3.3 Test setup

The sequences were displayed un-scaled with a black background on a Hyundai P240W with a 24" polarizing stereoscopic screen having a total resolution of 1,920 × 1,200 pixels and a resolution of 1,920 × 600 per view when used in stereoscopic mode. The viewing distance was set to 70 cm because in a trial test, it yielded slightly better subjective ratings with smaller quality variation compared to those of the viewing distance of 110 cm. Since the image height was slightly different and the images were displayed un-scaled, the viewing distance of 70 cm corresponded to the range of 2.1–2.4 H for different sequences, where H is the image height. Table 3 reports the visual angle in pixels per degree (PPD) for the test setup. Moreover, Table 2 reports the angular widths in degrees.

Table 3 Visual angle (in pixels per degree)

Downsampling ratio	Visual horizontal angle
1	22.8
1/2	11.4
3/8	7.6
1/4	5.7

Ten subjects with an average age of 21 years and without substantial prior experiences on stereoscopic video participated in the test. As we intended to confirm the previously achieved results regarding the eye dominance effect on the perceived visual quality of asymmetric stereoscopic video, half of the viewers were right-eye-dominant, while the other half were left-eye-dominant. Prior to the experiment, the viewers were subject to a thorough vision screening. The participants were screened for far and near visual acuity of each eye with a rejection criterion of 20/40 tested with Lea Numbers [23], stereoacuity criterion was 60 arcsec tested with the TNO stereo test. Criteria for near horizontal phoria, tested with the Maddox Wing test [24], were 13D for exophoria and 7D for esophoria, and 1D for vertical phoria. All participants had a stereoscopic acuity of 60 arc sec or better. The following visual tests were conducted for all participants: far and near visual acuity, stereoscopic acuity (Randot test), contrast sensitivity (Functional Acuity Contrast Test), near point of accommodation and convergence RAF gauge test [25], and the interpupillary distance. Viewers who were found not to have normal visual acuity and stereopsis were rejected. The duration of subjective test was limited to 45 min to prevent eye strain and fatigue in subjects. D50 white point, ambient illuminance level of ∼200 lux, and 20 % image surround reflectance were fixed as the viewing conditions of all experiments. Moreover, the background noise level was kept equal or less than 30 dBA. The subjective test started with a combination of anchoring and training. The participants were shown both extremes of the quality range of a stimulus to familiarize the participants with the test task, the contents, and the variation in quality to be expected in the actual test that followed. The test sequences were presented one at a time in a random order and appeared twice in the test session. Each sequence was rated independently after its presentation utilizing an on-screen scoring

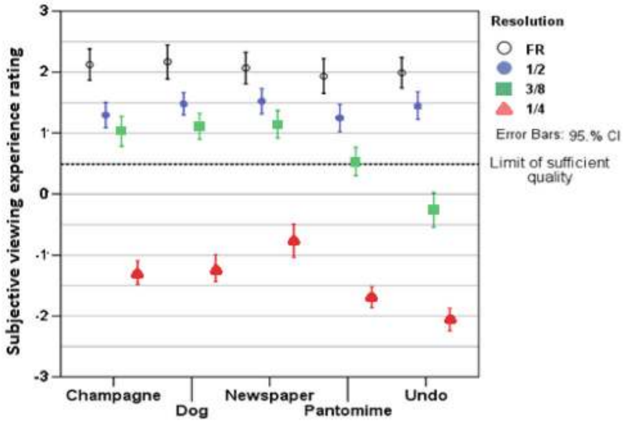


Fig. 3 Average subjective viewing experience ratings and the 95 % CI

scroll bar. After each rating, the next sequence started, and hence, the time used for rating was not limited in any of the experiments.

In this experiment, 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 and explained orally by the test coordinator 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.” Moreover, the viewers were asked to estimate the limit of sufficient quality [26] with a line on the general image quality scale after viewing each test sequence. This value estimated the minimum subjective rating over which the quality was acceptable for the viewers. Observers were allowed to keep the limit of the sufficient quality at the same point for the whole experiment.

3.4 Results

3.4.1 Limit of downsampling ratio

Figure 3 presents the average values and the 95 % confidence interval (CI) of the subjective viewing experience ratings. Furthermore, it displays the average limit of sufficient quality, which did not vary very much between sequences. It can be seen that the FR stereoscopic video sequences outperformed the MR sequences in all test cases. The quality of all MR stereoscopic image sequences downsampled by 1/2 both horizontally and vertically was clearly above the limit of sufficient quality. For three of the sequences, the downsampling ratio of 3/8 provided a quality higher than the limit of sufficient quality, while the quality of the MR sequences with the downsampling ratio of 1/4 was clearly unacceptable in terms of subjective image quality. Moreover, we observed that 70 % of the total rating interval was covered by the average subjective viewing experience ratings.

When compared to earlier studies [12,13], the performance of the MR sequences relative to the respective FR sequences was worse. This might be explained by the chosen viewing distance in relation to the physical size of a pixel. It has also been established that when the angular resolution (e.g. in pixels per degree) stays unchanged, the greater the angular size of the display, the more contrast sensitivity the HVS has [27]. Thus, the threshold angular resolution for mixed-resolution stereoscopic video may also depend on the angular size of the display. In the viewing conditions used in this test, downsampling ratios 1/2, 3/8, and 1/4 corresponded to 11.4, 7.6, and 5.7 PPD (of viewing angle), respectively, in the lower-resolution view. As a comparison, the downsampling ratios of 1/2 and 1/4 in [12] corresponded to more than 15 and close to 10 PPD, respectively, as far as we could conclude from the information provided in the paper. The exact values for pixels per viewing angle could not be concluded from the information given in [13], but the authors discovered equivalently to our results that the subjective difference between FR and MR was a descending function of the resolution in terms of the number of pixels.

Moreover, we analyzed whether the subjective image quality ratings had any correlation to the average luma PSNR of the lower-resolution view. The downsampled views were first upsampled to the FR, and the PSNR values were derived against the FR sequences. Then, a least square estimate was derived for the relation of the subjective image quality ratings and the obtained average luma PSNR values. Finally, a Pearson’s correlation coefficient was derived between the least square estimate and the PSNR values. A large Pearson’s correlation value can be assumed to indicate that the lower-resolution view contributed more heavily to the image quality rating. Table 4 provides the PSNR of the left view and the corresponding subjective score.

A comparison between the PSNR values and the subjective viewing experience ratings of the views downsampled by ratio 1/2 resulted in Pearson’s correlation coefficient equal to 0.10, indicating that there was practically no correlation between the subjective image quality rating and the average luma PSNR of the downsampled view. The data points and the resulting least square fit for downsampling ratios 3/8 and 1/4 are presented in Fig. 4. Interestingly, the slope of the linear estimations for downsampling ratios 3/8 and 1/4 was similar and equal to 0.30 and 0.28, respectively. Along with obvious similarity of the subjective scores and the linear estimations, we further confirmed the correlation by deriving the root mean square error values, 0.25 and 0.11, and the Pearson’s correlation coefficients, 0.88 and 0.97, for downsampling ratios 3/8 and 1/4, respectively. This analysis indicates that the PSNR of the lower-resolution view correlated with subjective perception at downsampling ratios of 3/8 and 1/4. As full-reference objective quality metrics, such as PSNR, were not applicable for the full-resolution view, no analysis

Table 4 The average luma PSNR of the left view and the average subjective viewing experience rating for different downsampling ratios

Downsampling ratio	1/2	3/8	1/4
	PSNR in dB–SSIM (average subjective rating)		
Dog	37.60–0.985 (1.47)	32.79–0.970 (1.11)	29.80–0.948 (−1.21)
Pantomime	35.62–0.990 (1.24)	33.42–0.979 (0.53)	28.74–0.965 (−1.68)
Champagne	36.32–0.993 (1.29)	33.96–0.988 (1.02)	29.04–0.983 (−1.28)
Newspaper	36.93–0.972 (1.52)	34.54–0.943 (1.14)	31.06–0.912 (−0.76)
Undo dancer	32.82–0.887 (1.45)	30.01–0.825 (−0.26)	26.44–0.778 (−2.05)

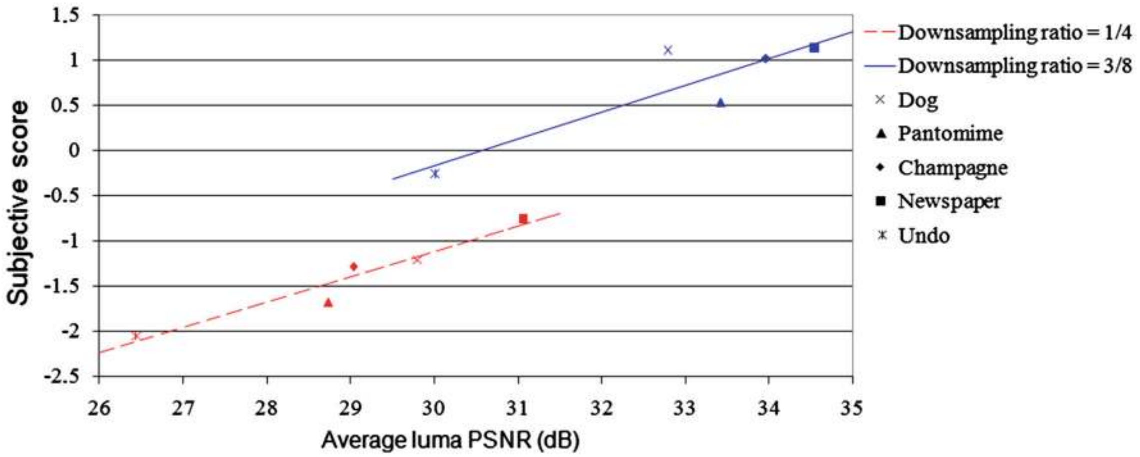


Fig. 4 Correlation of the average luma PSNR of the lower-resolution view and the subjective viewing experience ratings, *blue* = downsampling ratio 3/8, *red* = downsampling ratio 1/4 (color figure online)

on the subjective impact of the full-resolution view was feasible with a similar method. It would therefore require further studies to verify whether the full-resolution view was dominant in the subjective quality ratings for downsampling ratio 1/2 and similarly whether the lower-resolution view was dominant at downsampling ratios 3/8 and 1/4 for the viewing conditions and the sequences used in this experiment.

3.4.2 Eye dominance

As explained above, there were both left- and right-eye-dominant participants in the test which included two sets of test sequences, differing in whether the left view or the right view was downsampled and subsequently upsampled. Both left and right-eye dominant subjects scored the two sets of test sequences. Figure 5 presents the average ratings given by the left- and right-eye-dominant viewers, separately. The labels of the horizontal axis identify which view was downsampled and the downsampling factor. It can be observed that there is always an overlap of the 95 % confidence interval for all the respective scores, hence indicating that the eye dominance of the viewers had no significant impact on the perceived quality of the MR sequences used in the test. However, at the downsampling ratio of 1/4 along both coordinate axes, the average rating of the MR sequences where the full-resolution view was the same as the dominant eye of the viewer was slightly higher than the average rating of the other sequences of the same downsampling ratio.

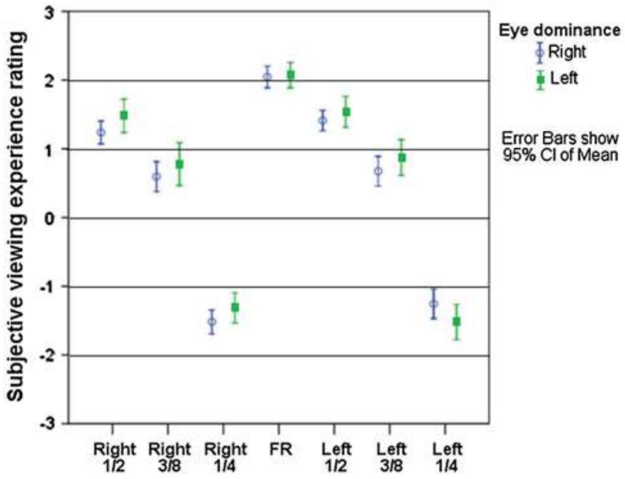


Fig. 5 Impact of eye dominance versus downsampled view

We also performed statistical significance comparison achieved by the Wilcoxon signed-rank test on the results. The scores from the left- and right-eye-dominant observers were tested against each other in order to find out whether their evaluations of the sequences differ in any case. All test cases achieved a P value equal to 1 except champagne and dog sequences at downsampling ratios of 1/2 and 3/8, respectively, for which the P values were 0.86 and 0.885, respectively. In other words, there were no significant differences of ratings between the left- and right-eye-dominant viewers based on these results. Our results therefore confirmed the earlier findings in [14] and [15] that eye dominance has no statistically significant impact on how MR sequences are rated subjectively.

4 Subjective quality assessment of asymmetric stereoscopic video coding

4.1 Introduction

Asymmetric stereoscopic video is perceived by the HVS in such a way that the lower quality of one view, due to compression artifacts, might be masked by the higher quality view. Therefore, we seek to assess the subjective quality of asymmetric stereoscopic videos with different quality combinations. For single-view video, there are a number of objective quality measures which can be used [28]. However, when it comes to stereoscopic video, objective quality assessment metrics may face some ambiguity as how to perform the joint assessment fairly, since there are two views involved with different qualities. In this section, we seek an answer to the following question: "*Does asymmetric stereoscopic video coding make sense from a subjective quality point of view?*" The approach to reach a conclusion is based on subjective quality assessment of symmetric and asymmetric stereoscopic videos having the same bitrate. Furthermore, the impact of downsampling ratio in mixed-resolution stereoscopic video coding is analyzed in terms of encoding computational complexity. This section further extends our preliminary results in [29].

4.2 Test material

The tests were carried out using four sequences: undo dancer, dog, pantomime, and newspaper. Three types of sequences were tested as follows:

1. Full-resolution with symmetric quality in both views
2. Full-resolution with asymmetric quality between the views caused by different quantization step 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 5. The mixed-resolution uncompressed sequences were generated from the FR ones by downsampling the left view further. Downsampling ratios 1/2 and 3/8 were symmetrically applied horizontally and vertically. As in Sect. 3.4.2, we confirmed that eye dominance was not shown to have an impact which view is provided with a better quality, only one set of MR sequences was prepared. Views were independently coded using H.264/AVC in order to treat the FR and MR cases as equally as possible and prevent affecting the results by different performance of inter-view prediction depending on the downsampling ratios. Moreover, since no inter-view prediction has been standardized for a MR coding scheme, we specifically avoided the use of non-standardized codecs to provide as generally applicable results as possible. Examples of coding arrangements enabling mixed-resolution stereoscopic video with inter-view prediction have been proposed, for example, in [30] and [31].

Table 5 Spatial resolutions of different sequences

	Full	1/2	3/8
Undo dancer	960 × 576	480 × 288	360 × 216
Others	768 × 576	384 × 288	288 × 216

The duration of a viewing session was limited to less than 1 h to avoid viewers becoming exhausted. Hence, the experiment was split into two sessions, where 9 and 7 naïve 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 quality and bitrate of H.264/AVC bitstreams are 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. The horizontal axis of Fig. 6 displays 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 6a 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 6b, c summarize the QP selections for the downsampling ratio of 1/2

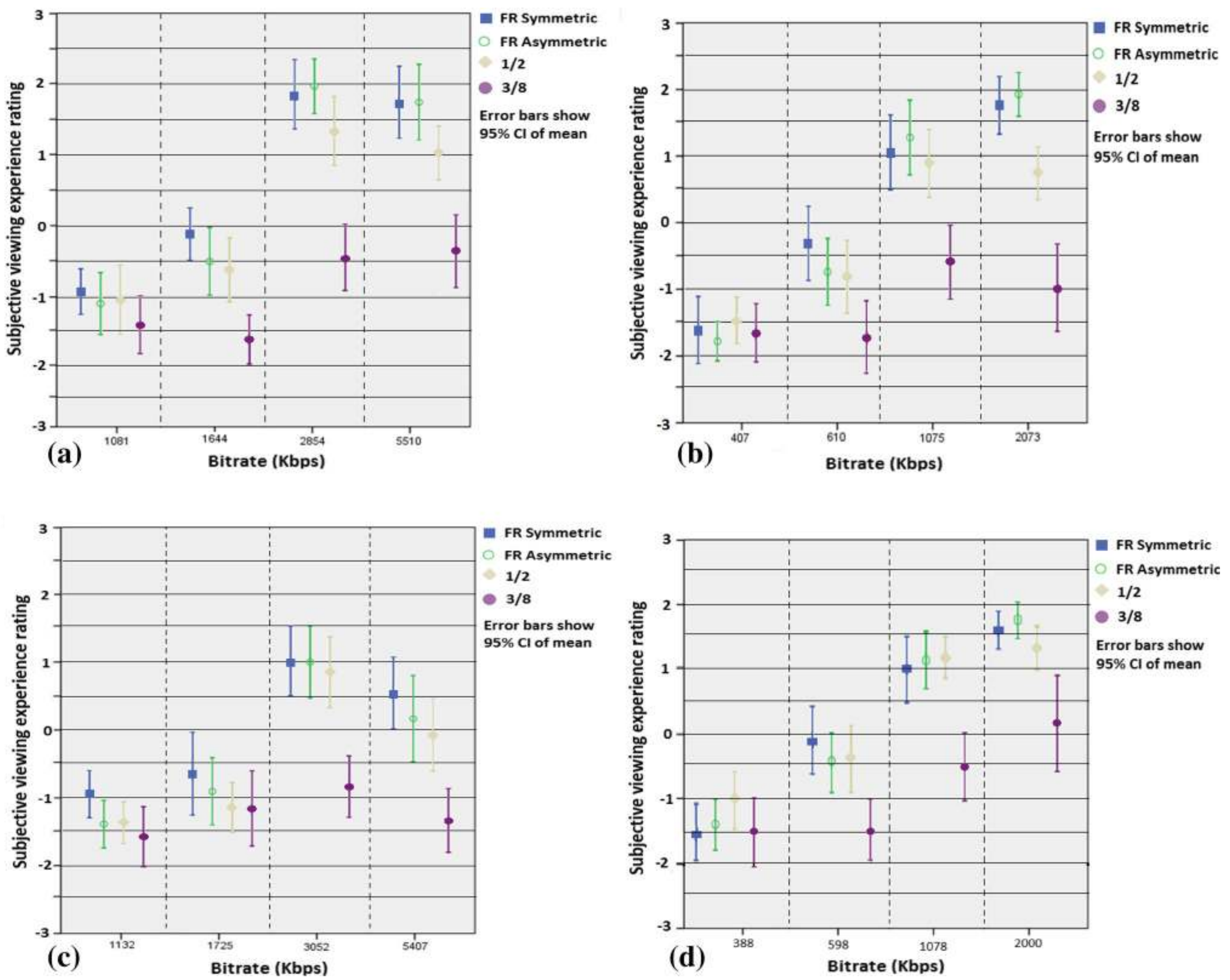


Fig. 6 Results of compressed MR subjective tests for sequences: **a** undo dancer, **b** newspaper, **c** pantomime, **d** dog

and 3/8, respectively. 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 the 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.

4.3 Results and discussion

The average subjective viewing experience ratings are presented in Fig. 6. The 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 subjective quality of MR clips with downsampling ratio 3/8 along both axes is clearly inferior to the subjective quality of all other corresponding test cases. Thus, the results of downsampling ratio 3/8 are not discussed further. Moreover, although the confidence intervals overlap for the two highest bitrates in Fig. 6c, the average subjective ratings of the highest bitrate are slightly lower than the second highest bitrate. This is due to the fact that the experiment was divided to two sessions, and as a result, all four bitrates are not comparable. The highest bitrate and second lowest bitrate were included in the same session while the two other bitrates in another session.

Figure 6 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, there is an overlap of the 95 % confidence intervals of the subjective ratings of FR symmetric, FR asymmetric, and MR with downsampling ratio 1/2 for each test sequence. However, the use of mixed-resolution coding can be justified in many applications by its lower computational complexity. Furthermore, it can be observed from Fig. 6 that the performance of mixed-resolution coding of downsampling ratio 1/2 depends on the input sequence to some extent.

Table 6 QP selection (left-right) for asymmetric stereo bitstreams. a represents QP for FR asymmetric quality, while b and c represent QP selection where the left view is downsampled with ratio of 1/2 and 3/8, respectively

QP	39-39	35-35	30-30	25-25
(a) FR asymmetric bitstreams				
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				
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
(c) MR bitstreams with downsampling ratio of 3/8				
Undo dancer	32-36	29-32	24-28	19-23
Dog	32-36	29-32	24-27	19-22
Pantomime	32-36	29-32	24-27	19-21
Newspaper	31-36	28-32	24-27	20-22

Objective quality metrics were applied to the sequences to analyze the subjective viewing results as follows. Since to our knowledge, no widely adopted objective metrics for stereoscopic video are available, we verified the results with two common metrics: PSNR and structured similarity (SSIM) [32,33]. The average luma PSNR was derived for each view of each bitstream. For mixed-resolution bitstreams, a decoded view of a lower-resolution was upsampled before the PSNR calculation to have comparable results with full-resolution bitstreams. In the following, the PSNR of the left (L) and right (R) views of the full-resolution symmetric, full-resolution quality-asymmetric, and mixed-resolution bitstreams are marked with $P_{\mathrm{SFRL}}, P_{\mathrm{SFRR}}, P_{\mathrm{AFRL}}, P_{\mathrm{AFRR}}, P_{\mathrm{MRL}}$, and P_{MRR}, respectively. SSIM values were also derived for each view of each bitstream similarly to PSNR. In the following, the SSIM values are marked in a similar fashion as, that is, $S_{\mathrm{SFRL}}, S_{\mathrm{SFRR}}, S_{\mathrm{AFRL}}, S_{\mathrm{AFRR}}, S_{\mathrm{MRL}}$, and S_{MRR}.

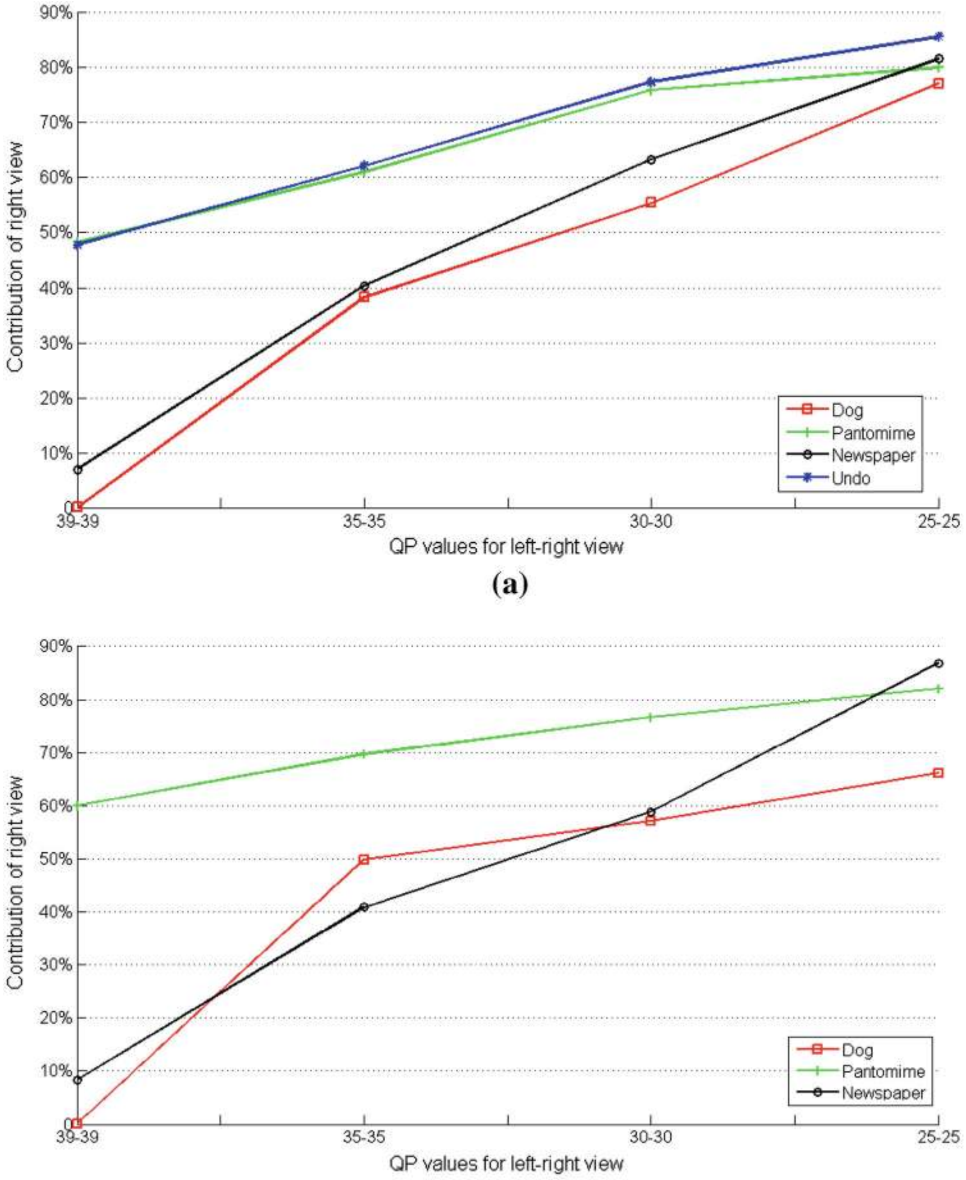
In the case of MR stereoscopic video, both blurring and blocking are involved. We analyzed the relative contribution of the views of MR bitstreams to the overall subjective quality with both PSNR and SSIM as follows. It was assumed that the average objective quality (PSNR or SSIM) of the symmetric FR bitstreams reflects the overall subjective quality. Furthermore, we assumed that when a weighted average of the objective quality values between the left and right view of an MR bitstream matches the average objective quality of the respective symmetric FR bitstream having the same subjective quality rating, the weights for the weighted averaging reveal the relative contribution of left and right views to the subjective quality. In other words, for those MR bitstreams that had approximately equal subjective quality as the respective FR bitstreams, we derived weights W that minimized the mean square error of the difference between the weighted average of the objective quality of the left and right views and that of the FR:

$$\mathrm{mse} = (W \times P_{\mathrm{MRR}} + (1 - W) \times P_{\mathrm{MRL}} - P_{\mathrm{SFR}})^2 \qquad (1)$$

In Eq. (1), $W \times P_{\mathrm{MRR}} + (1 - W) \times P_{\mathrm{MRL}}$ reflects the weighted average of MR bitstreams and mse is minimized by changing the weight (W) over the quality of left and right views. Assuming that $P_{\mathrm{MRL}} < P_{\mathrm{SFR}} < P_{\mathrm{MRR}}$, which is typically true because only the left view is downsampled and due to the downsampling, the right view gets a lower QP value compared to the right view of symmetric FR, the above expression reaches its minimum when

$$W = (P_{\mathrm{SFR}} - P_{\mathrm{MRL}}) / (P_{\mathrm{MRR}} - P_{\mathrm{MRL}}) \qquad (2)$$

The same reasoning can be applied for SSIM. Figure 7a, b indicate the contribution of the right view to the overall quality, that is, W, for different QP values and sequences, derived from PSNR and SSIM, respectively. The results of undo dancer were not included in Fig. 7b because the MATLAB implementation of the SSIM index, utilizing the suggested empirical formula [33], seemed to fail in estimating its subjective quality. SSIM provided very close values for the left and right views for undo dancer as derived from Eq. (2). A full 100 % contribution was assigned to the right view for the three highest QP values. This was not the case for the



(a)

(b)

Fig. 7 Contribution of the FR view (*right*) to the overall quality of mixed-resolution stereoscopic video measured by **a** PSNR **b** SSIM, that is, the value of W as derived with Eq. (2)

other sequences, perhaps due to the synthetic nature of the undo dancer sequence. It can be seen from Fig. 7a, b that the contribution of the right view increased when blocking decreased and that the higher the QP value became, the more contribution the left view had on the overall quality. Moreover, Fig. 7 appears to be in agreement with the conclusions in [7] that the perceived quality of the mixed-resolution videos was close to that of the higher-resolution view. This behavior was not biased by QP selection for the left and the right view for different bitrates since as reported in Table 6b, the QP difference between the left and the right view for all MR videos was kept equal or close to three. It can also be seen in Fig. 7 that the relative contribution of the right view was dependent on the sequence.

The average luma PSNR over both views of the quality-asymmetric full-resolution bitstreams, that is, $(P_{\mathrm{AFRL}} + P_{\mathrm{AFRR}})/2$, was found to be very close to that of the symmetric full-resolution bitstreams, that is, $P_{\mathrm{SFR}} = (P_{\mathrm{SFRL}} + P_{\mathrm{SFRR}})/2$, the absolute difference being only 0.1 dB on average. The same analysis for SSIM metric resulted in an absolute difference of 0.005 on average. This finding is aligned with the earlier conclusions in [7] and [15] that the perceived quality of the quality-asymmetric video was approximately the mean of the perceived qualities of the two views. The same analysis, as reported for MR stereoscopic video in Fig. 7, was performed for quality-asymmetric full-resolution sequences. The results are provided in Fig. 8 for both PSNR and SSIM objective metrics showing that both

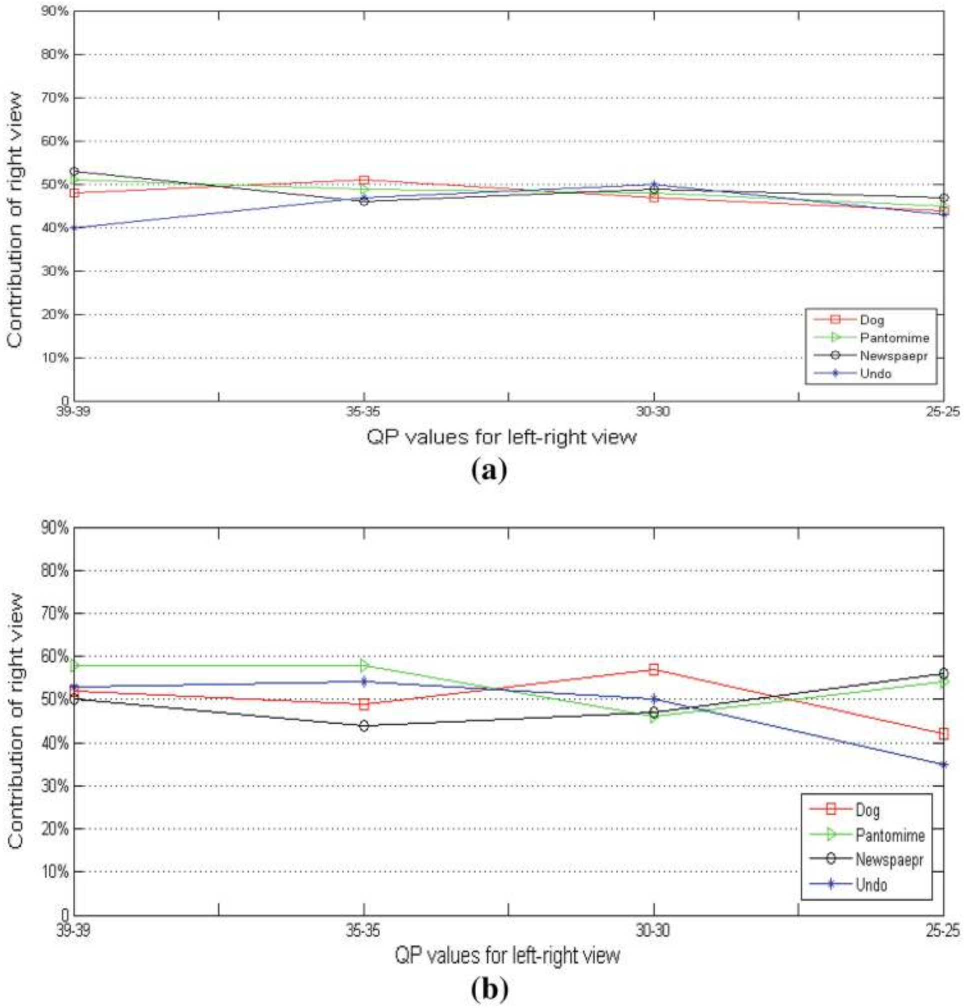


Fig. 8 Contribution of the right view to the overall quality of quality-asymmetric full-resolution stereoscopic video measured by **a** PSNR **b** SSIM

views contributed almost equally to the final quality of the stereoscopic video.

As discussed above, MR coding did not provide a better subjective quality compared to FR coding. However, due to the smaller spatial resolution, the use of MR coding may be justified. A complexity comparison for encoding the full and lower-resolution views in our experiments is presented in Fig. 9. The experiments were performed on Windows OS with a dual-core CPU having a clock rate of 3.16 GHz. The execution time for the FR view consisted of the encoding time, and for the lower-resolution view, it included both the encoding and the downsampling times. Since the encoding time varied depending on the ongoing processes of the PC, an average value of seconds per frame over five different QP values for full-length videos was calculated. As illustrated in Fig. 9 by decreasing the spatial resolution by ratio 1/2 and 3/8 both vertically and horizontally, the encoding time decreased on average to 36 and 21 % of the encoding time for the FR sequences, respectively.

To reduce the amount of time-taking subjective experiments, it is preferred to estimate the subjective quality of asymmetric stereoscopic video by a reliable model depending on available information, for example, the characteristics of the viewing conditions, the used asymmetric coding scheme, and the viewed video content. In [34], we tried to estimate the subjective quality of asymmetric stereoscopic video taking into account the number of pixels per degree of viewing angle. The results showed high correlation between subjective ratings and pixels per degree values but were obtained with a relatively small amount of subjective test data. In order to verify the results of [34] and to develop the model further, we plan to conduct extensive subjective tests under multiple test setup conditions, different asymmetric coding schemes, and various video clips.

5 Conclusions

In this paper, we attempted to discover suitable methods and configurations for asymmetric stereoscopic video coding through two sets of systematic subjective quality evaluation experiments. We studied the subjective impact of downsam-

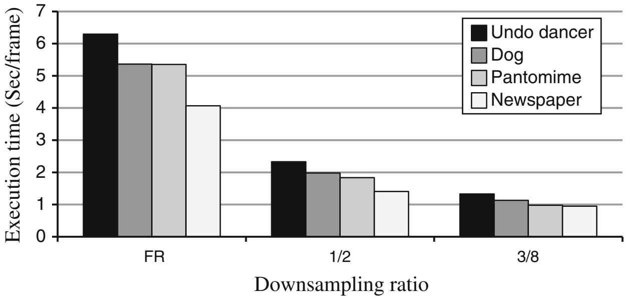


Fig. 9 Encoding time comparison for FR view and downsampled views

pling applied for one of the views in an uncompressed mixed-resolution (MR) stereoscopic video. In our experiment, FR sequences always outperformed MR sequences. However, the quality of the MR sequences where one view was downsampled by a factor of 1/2 horizontally and vertically was clearly acceptable. We found that the lower-resolution view appeared to become dominant in the subjective quality rating at a certain downsampling ratio, which seemed to depend on the sequence, the angular resolution, and the angular width.

A subjective test comparing symmetric full-resolution, quality-asymmetric full-resolution, and mixed-resolution stereoscopic video coding was also presented. The performance of symmetric and quality-asymmetric full-resolution bitstreams was found to be 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 test 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.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Chen, Y., Wang, Y.-K., Ugur, K., Hannuksela, M.M., Lainema, J., Gabbouj, M.: The emerging MVC standard for 3D video services. EURASIP J. Adv. Signal Process. **2009**, 13 (2009); Article ID 786015. doi: 10.1155/2009/786015
2. ITU-T Recommendation H.264.: Advanced Video Coding for Generic Audiovisual Services (March 2009)
3. Merkle, P., Smolic, A., Muller, K., Wiegand, T.: Efficient prediction structure for multiview video coding. IEEE Trans. Circuits Syst. Video Technol. **17**(11), 1461–1473 (2007)
4. Sullivan, G.J., Ohm, J.-R., Han, W.-J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol. **22**, 1649–1668 (2012)
5. Suzuki, T., Hannuksela, M.M., Chen, Y., Hattori, S., Sullivan, G.J. (eds.): MVC extension for inclusion of depth maps draft text 4. Joint Collaborative Team on 3D Video Coding Extension Development, document JCT3V-A1001 (July 2012)
6. Suzuki, T., Hannuksela, M.M., Chen, Y., Hattori, S., Sullivan, G.J. (eds.): MVC extension for inclusion of depth maps draft text 6. Joint Collaborative Team on 3D Video Coding Extension Development, document JCT3V-C1001 (Mar. 2013)
7. Blake, R.: Threshold conditions for binocular rivalry. J. Exp. Psychol. Human Percept. Perform. **3**(2), 251–257 (2001)
8. Julesz, B.: Foundations of Cyclopean Perception. University of Chicago Press, Chicago (1971)
9. Perkins, M.G.: Data compression of stereopairs. IEEE Trans. Commun. **40**(4), 684–696 (1992)
10. Aksay, A., Bilen, C., Bozdagi Akar, G.: Subjective evaluation of effects of spectral and spatial redundancy reduction on stereo images. In: 13th European Signal Processing Conference, EUSIPCO-2005, Turkey (Sep. 2005)
11. Aflaki, P., Hannuksela, M.M., Hakala, J., Häkkinen, J., Gabbouj, M.: Joint adaptation of spatial resolution and sample value quantization for asymmetric stereoscopic video compression: a subjective study. In: Procedings of the International Symposium on Image and Signal Processing and Analysis (Sep. 2011)
12. Stelmach, L., Tam, W.J., Meegan, D., Vincent, A.: Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Trans. Circuits Syst. Video Technol. **10**(2), 188–193 (2000)
13. Brust, H., Smolic, A., Mueller, K., Tech, G., Wiegand, T.: Mixed resolution coding of stereoscopic video for mobile devices. In: Proceedings of the of 3DTV Conference (May 2009)
14. Meegan, D.V., Stelmach, L.B., Tam, W.J.: Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery. J. Exp. Psychol. Appl. **7**(2), 143–153 (2001)
15. Seuntiens, P., Meesters, L., IJsselsteijn, A.: Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation. ACM Trans. Appl. Percept. **3**(2), 96–109 (2006)
16. Bruckstein, A.M., Elad, M., Kimmel, R.: Down-scaling for better transform compression. IEEE Trans. Image Process. **12**(9), 1132–1144 (Sep. 2003)
17. Saygili, G., Gürler, C.G., Tekalp, A.M.: Quality assessment of asymmetric stereo video coding. In: Proceedings of the of IEEE Internationl Conference on Image Processing (Sep. 2010)
18. Tam, W.J.: Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV. In: Joint Video Team document JVT-W094 (Apr. 2007)
19. Strohmeier, D., Tech, G.: Sharp, bright, three-dimensional: open profiling of quality for mobile 3DTV coding methods. In: Proceedings of the SPIE International Society for Optical Engineering, vol. 7542 (Jan. 2010)
20. Aflaki, P., Hannuksela, M. M., Häkkinen, J., Lindroos, P., Gabbouj, M.: Impact of downsampling ratio in mixed-resolution stereoscopic video. In: Proceedings of the of 3DTV Conference (June 2010)
21. ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications (1999)
22. JSVM Software: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm
23. Hyvärinen, L.: Lea Numbers 15-Line distance chart instructions. Web page: leatest.net/en/vistests/instruct/2711/index.html. Accessed 01 Jan 2011, Created 2009
24. Rosenfield, M., Logan, N. (eds.): Optometry: Science, Techniques and Clinical Management. Elsevier, Amsterdam (2009)
25. Neely, J.C.: The R.A.F. near-point rule. British J. Ophthalmol. **40**(10), 636–637 (Oct. 1956)

26. Nyman, G., Häkkinen, J., Koivisto, E.-M., Leisti, T., Lindroos, P., Orenius, O., Virtanen, T., Vuori, T.: Evaluation of the visual performance of image processing pipes: information value of subjective image attribute. In: Proceedings of the SPIE, vol. 7529 (Jan. 2010)
27. Barten, P.G.J.: The effects of picture size and definition on perceived image quality. IEEE Trans. Electron. Devices **36**(9), 1865–1869 (Sep. 1989)
28. You, J., Reiter, U., Hannuksela, M.M., Gabbouj, M., Perkis, A.: Perceptual-based quality assessment for audio-visual services: a survey. Signal Process. Image Commun. **25**(7), 482–501 (2010)
29. Aflaki, P., Hannuksela, M.M., Häkkinen J., Lindroos P., Gabbouj M.: Subjective study on compressed asymmetric stereoscopic video. In: Proceedings of IEEE International Conference on Image Processing (ICIP) (Sep. 2010)
30. Chen, Y., Liu, S., Wang, Y.-K., Hannuksela, M.M., Li, H., Gabbouj, M.: Low-complexity asymmetric multiview video coding. In: Proceedings of the IEEE International Conference on Multimedia & Expo (ICME) (June 2008)
31. Brust, H., Tech, G., Mueller, K., Wiegand, T.: Mixed resolution coding with interview prediction for mobile 3DTV. In: Proceedings of 3DTV Conference (June 2010)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
33. https://ece.uwaterloo.ca/~z70wang/research/ssim/
34. Aflaki, P., Hannuksela, M.M., Hakala, J., Häkkinen, J., Gabbouj, M.: Estimation of subjective quality for mixed-resolution stereoscopic video. In: Proceedings of the of 3DTV-Conference (May 2011)