Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2014) 22: 663 – 678
© TÜBİTAK
doi:10.3906/elk-1209-53

Perceptual quality evaluation of asymmetric stereo video coding for efficient 3D rate scaling

Nükhet ÖZBEK[1,*], Gizem ERTAN[2], Oktay KARAKUŞ[3]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Ege University, İzmir, Turkey
[2]International Computer Institute, Ege University, İzmir, Turkey
[3]Department of Electrical and Electronics Engineering, Faculty of Engineering, İzmir Institute of Technology, İzmir, Turkey

Received: 11.09.2012 • **Accepted:** 22.11.2012 • **Published Online:** 21.03.2014 • **Printed:** 18.04.2014

Abstract: In 3D perception, the binocular suppression of the human vision, perceiving high quality 3D video in the case of a view in higher quality, can be exploited in asymmetric stereo video coding for efficient 3D rate scaling. Hence, the best stereo rate-visual distortion performance may be gained by asymmetric coding, which is the reduction of the spatial and/or quantization resolution of the low-quality view, while keeping the high-quality view in full resolution. However, how to determine the level of asymmetry and what type of scaling should be chosen are still in question. In this work, we try to assess the overall performance of the scalability options with several test contents. The test videos are encoded at critical bitrates with symmetric options and spatial or signal-to-noise ratio (SNR) asymmetric coding, and are subjectively evaluated in a stereo-polarized projection 3D display system. Two different types of evaluation methodologies are used: the Double-Stimulus Continuous-Quality Scale (DSCQS) and Subjective Evaluation of Stereo Video Quality (SESVIQ). Dense visual tests show that the spatial scaling is generally inferior when compared to SNR scaling, except that high motion scenes and symmetric SNR are more preferable for a higher bitrate. The characteristics of the video content should be taken into consideration for efficient stereo rate scaling.

Key words: Asymmetric stereo video coding, 3D video quality assessment, DSCQS, SESVIQ, blur, blockiness

1. Introduction

Recently, 3D video has made a significant jump in terms of its establishment in large-scale screening technologies. Stereoscopic 3D became mature and is widely accepted from capture to display; thus, the developments are believed to be sustainable. Such systems mostly rely on approaches like stereo or multiview video for representation, simulcast, or multiview video coding [1] for coding and transmission. However, it is a big challenge to deliver a highly satisfying viewing experience under certain technical requirements.

The binocular suppression theory of the human visual system (HVS) is widely accepted, which assumes that the perceived 3D quality is constructed by the domination of the higher quality view [2–6]. The theory can be exploited in stereo video encoding to reach the best stereo rate-visual distortion (RD) performance. This is performed by asymmetric coding, which is the reduction of the spatial and/or quantization resolution of the auxiliary view, while the reference view is kept in full or higher resolution [7–14]. The objective of asymmetric coding is to maximize the perceived 3D video quality subject to an overall bitrate constraint by efficient stereo

*Correspondence: ozbek.nukhet@gmail.com

rate scaling. Nevertheless, how to determine the level of asymmetry and what type of scaling should be chosen are still in question.

The most appealing part of this problem is to imagine an adequate quality measure for the level of asymmetry. Conventional metrics, like peak signal-to-noise ratio (PSNR), are not adequate for video qualities at different spatial and/or SNR resolutions and in 3D [15,16]. Modern image quality assessment techniques have already been revealed for monoscopic video as visual quality metrics, i.e. blockiness, blur, structure similarity metric, and video quality metric. Some of them are standardized based on a benchmark by the Video Quality Experts Group. To the best of our knowledge, there are a few papers on stereoscopic image [17] and video [18]. Hence, subjective video quality measurement is still a valid way to evaluate perceptual 3D video quality. The Double Stimulus Continuous Quality Scale (DSCQS), Single Stimulus Impairment Scale (SSIS), and Single Stimulus Continuous Quality Evaluation (SSCQE) are recommendations of the International Telecommunication Union Radiocommunication Sector (ITU-R), which are described in BT.500-11 [19]. Some others are recommended by the European Broadcasting Union, such as the Absolute Category Rating (ACR) [20] and Subjective Assessment of Video Quality (SAMVIQ) [21,22].

In psychovisual studies of stereoscopic vision, it has been demonstrated that the blur in a distorted image presented to one view is masked by a sharp image presented to the other view and does not affect the perceived depth [2]. Stelmach et al. [3] first started to study the ways of asymmetric coding and the quality assessment thereof. In [4], they compared low-pass filtering with discrete cosine transform (DCT)-based quantization and reported that in a low-pass filtering experiment, the mean quality scores were close to the individual scores of the unfiltered reference view. On the other hand, in a quantization experiment, the image quality scores turned out to be an approximate average of the individual scores of the reference and auxiliary views. The results for the third test, which combined low-pass filtering and quantization, revealed that the perceived image quality was dominated by quantization, and thus low-pass filtering had very little effect compared to quantization. Note that one of the views was at original quality for the low-pass filtering and the combined tests; however, the 2 views were both degraded for the quantization test.

The results in [6] showed that blockiness distortion is much more disturbing compared to blur distortion. The image quality of stereo images with different amounts of blur distortions in the reference and auxiliary view images is considered as dominated by the high quality view. However, if both views have blockiness distortion, the perceived 3D image quality is considered as an average of the image quality of the 2 views.

In [12], asymmetric rate scaling using scalable video coding was evaluated on 2 different types of 3D display technologies: polarized projection and parallax barrier. The experimental results showed that users with full spatial resolution 3D displays, like polarized projection, should prefer SNR scalability, while users with half-resolution 3D displays, such as parallax barrier, should favor the spatially scaled stream. In [13] and [14], the authors made the following conclusions: 1) at high bitrates, asymmetric encoding by scaling SNR resolution performs the best in perceived quality; 2) at low bitrates, asymmetric encoding by scaling spatial resolution performs the best in perceived quality; and 3) between these, symmetric encoding is preferred over asymmetric encoding.

In this paper, we present the experimental results of extensive subjective tests for several videos that are asymmetrically coded and displayed on a polarized stereo projection system. The objective of this study is to search appropriate scalability options for asymmetric coding towards the best 3D rate scaling. As an alternative to DSCQS, we develop novel software for an interactive multistimuli methodology, in which the tested condition may be compared against the other tested conditions. We use 8 stereo sequences with different

content characteristics to understand how the effect of the scaling option relates to the content. The test points are critically selected after subjective evaluation with another program, which is developed and dedicated to expert use. We conduct dense experiments in order to cross-check using the same stimuli with the 2 quality evaluation methods. We also conduct a monocular quality evaluation and present the relation between the monocular and binocular subjective test results.

This paper is organized as follows: Section 2 explains our research method and test setup in detail. Experimental results are given and discussed in Section 3. Finally, the paper ends with the conclusions in Section 4.

2. Research method and test setup

The experiments were conducted under controlled laboratory conditions, which were set according to ITU-R BT.500-11 recommendations [19]. This section presents an explanation of our research methodology, stereo projection display system, test stimuli, and participants.

2.1. Testing environment

Our stereo projection display system, of which a screenshot is given in Figure 1, consists of a pair of Sharp PG-D4010X (4200 American National Standards Institute lumens) digital light processing projector devices, a dielectric screen (silver), a pair of polarized filter glasses, and a PC to drive the projectors. One polarization filter applies a linear clockwise direction and the other a counter-clockwise direction for the right and left eye views, respectively. A virtual desktop of 2048 $\times$ 768-pixel resolution on the PC is set up so that each projector displays one half of the virtual desktop on the dielectric screen on top of each other. To be able to ensure that the appropriate viewing distance is satisfied according to the screen size, the subjects sit and watch from about 3 m away.

Figure 1. Stereo projection display laboratory: a special-purpose dark room consistent with the specifications given in ITU-R Rec. 500-11 [19].

2.2. Content selection and stimuli preparation

The tests are carried out using 8 sequences: Flower2, Flowerpot, Horse, and Car are included in the first group (Group 1). Flower3, Soccer2, Pantomime, and Dog are used for the second group (Group 2) [23]. The characteristics vary in terms of their spatial details, object/camera motion, amount of depth, and natural/studio light. Resolutions are 640 $\times$ 352 for Flower2, Horse, Car, and Flower3 and 720 $\times$ 480 for Flowerpot, Soccer2, Pantomime, and Dog. Thumbnail images of the representative frames of the test sequences are given in Figure 2.

Figure 2. Thumbnail images of the representative frames of the test sequences.

The reference (V0) and the auxiliary view (V1) are encoded independently as a single/base layer (H.264/AVC) using the Joint Scalable Video Model (JSVM) reference software of the Scalable Video Coding standard [24]. For the spatial asymmetric option, V1 is encoded after down-sampling by a factor of 2, both horizontally and vertically. The JSVM is also used for down-sampling and up-sampling filters. We construct RD curves with the quantization parameter (QP) values changing gradually between 20 and 50 for the full resolution of V0 and full and quarter resolution of V1, which are depicted in Figures 3 and 4. The QP value is fixed throughout the sequence to avoid quality variance in terms of the quantization SNR.

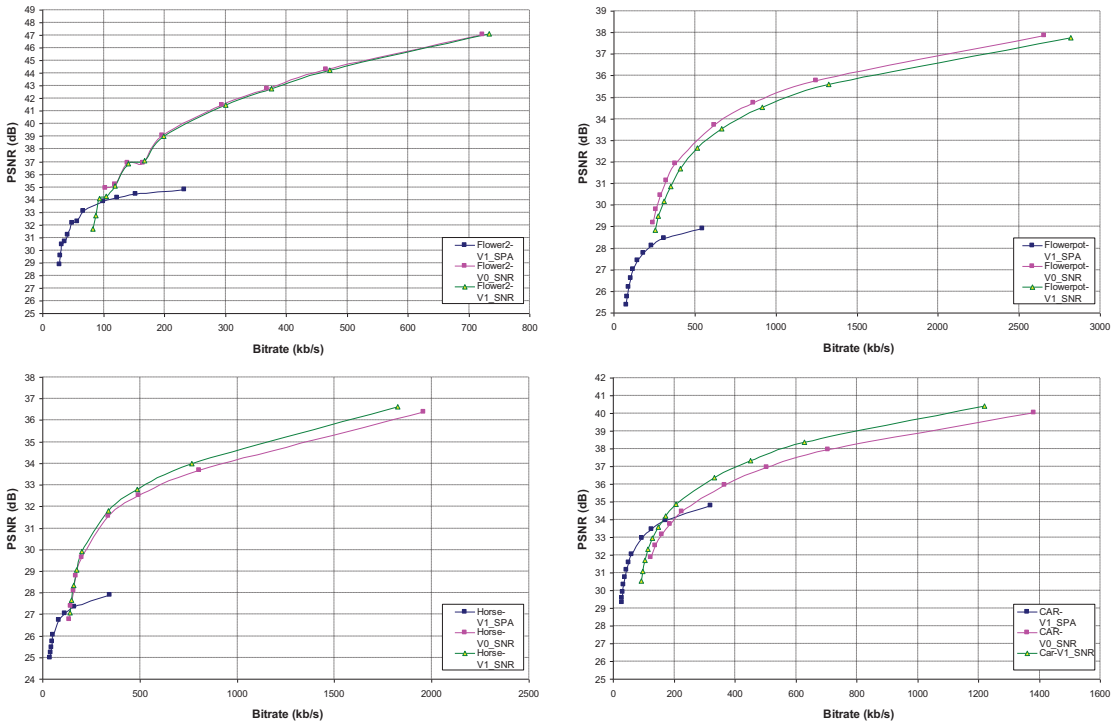

Figure 3. RD curves of the test videos for spatial and SNR scaling for Group 1.

Similar to the study in [14], we develop a comparator test program and investigate the perceptually noticeable level of asymmetry to pick up a critical threshold for the test points. In order to find a critical level

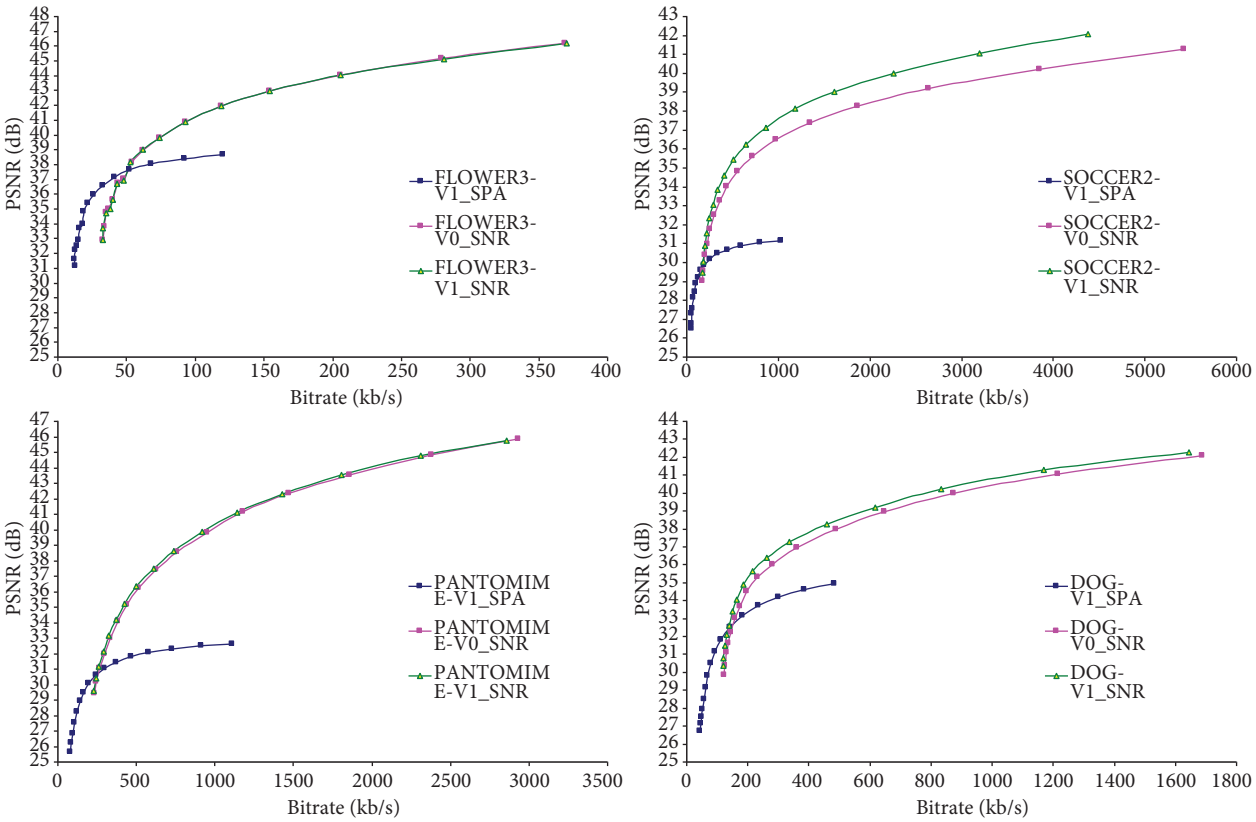

Figure 4. RD curves of the test videos for spatial and SNR scaling for Group 2.

of asymmetry, the comparator test program is utilized to visually test all of the QP options that construct the RD curves at full resolution. The test starts with both views at highest quality and then the subjects decrease the quality of the auxiliary view down to a threshold at which the perceptual quality maps are slightly annoying or annoying. The threshold PSNR is about 31 dB for Flowerpot, whereas it is 35 dB for Flower2 according to expert assessors. Since we try to assess the overall performance of the scalability options with several test contents, half of the test points should be definitely critical and the other half should be moderately critical, according to [19]. One rate point is chosen below the threshold and another test point is picked up above it. Test points under the threshold, called low bitrates, correspond to definitely critical parts of the test material, whereas the others are of moderately critical parts and are called high bitrates. Note that high bitrate points are restricted to the spatial scaling curves of the auxiliary view (see Figures 3 and 4). The bitrate in Kbps, PSNR in dB, and QP values are given in Tables 1 and 2 for the test points of Groups 1 and 2, respectively, where Asym = asymmetric, Sym = symmetric, Spa = spatial, Lo = low, and Hi = high. In Table 1, since the left and right QP values selected for the asymmetric SNR option turned out to be very close to each other, it was called Asym/Sym SNR-Hi.

2.3. Visual tests: testing procedure

Prior to the subjective tests, the following tests are conducted for all of the assessors: far visual acuity, Randot test, and contrast sensitivity. An anchoring and training session is held in order to familiarize the participants with the contents used, extremes of quality range, and evaluation process. The training session is another

program such that for each video test sequence, the subjects are shown the lowest and the original quality and are explicitly requested to compare them using the DSCQS method.

Table 1. Group 1 test points.

		Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym/Sym SNR-Hi
Flower2	QP-V1	30	42	40	24	36
	Rate-V1	123	119	141	233	199
	PSNR-V1	34.15	35.10	36.87	34.81	39.04
	QP-V0	36	36	38	32	32
	Rate-V0	196	196	165	295	295
	PSNR-V0	39.07	39.07	36.91	41.46	41.46
Flowerpot	QP-V1	30	44	38	24	34
	Rate-V1	233	256	349	547	516
	PSNR-V1	28.13	28.86	30.88	28.92	32.64
	QP-V0	34	34	36	32	32
	Rate-V0	478	478	378	619	619
	PSNR-V0	32.83	32.83	31.93	33.72	33.72
Horse	QP-V1	30	44	40	24	32
	Rate-V1	114	140	158	344	336
	PSNR-V1	27.05	27.08	28.36	27.89	31.81
	QP-V0	36	36	38	30	30
	Rate-V0	200	200	172	492	492
	PSNR-V0	29.66	29.66	28.80	32.53	32.53
Car	QP-V1	32	44	38	24	32
	Rate-V1	95	114	172	319	332
	PSNR-V1	32.94	32.33	34.20	34.80	36.38
	QP-V0	36	36	40	28	28
	Rate-V0	226	226	160	706	706
	PSNR-V0	34.46	34.46	33.15	37.97	37.97

2.3.1. DSCQS

All of the test pairs, including the originals, show up randomly in the test session. Each participant completes 24 experimental pairs (4 sequences and 6 conditions, including the original) for the Group 1 test while 21 pairs (3 sequences, excluding Dog, and 7 conditions, including the original) are used for the Group 2 test. Since the number of algorithms is incremented by 1, only 3 videos of Group 2 are included. Before voting, each test video and its original version are displayed 2 times. Since their order is also random, the subject cannot have knowledge of which one is the original. During the voting period, grading is done on a continuous scale for both stimuli according to the DSCQS. The assessors are asked to rate the overall quality of the test stimulus considering the perceived depth, sharpness, and naturalness. We also asked the assessors about impairment of visual comfort, but no assessors complained about visual discomfort.

For the analysis of the results, an individual score is calculated as the difference between the scores of the source video and the test conditions. By calculating all of the scores, they are normalized to values between 0 and 100. The mean opinion score (MOS) refers to the average of the normalized scores of the assessors who have not been rejected due to scoring outside of safety margins [19]. Finally, to calculate the score of each test condition, the averages of the MOS for that test condition over all of the videos are taken.

Table 2. Group 2 test points.

		Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym SNR-Hi	Sym SNR-Hi
Flower3	QP-V1	30	48	46	28	42	38
	Rate-V1	66	68	70	82	80	96
	PSNR-V1	36.6	33.8	34.8	37.18	35.64	37.06
	QP-V0	40	40	44	34	34	36
	Rate-V0	86	86	76	124	124	106
	PSNR-V0	36.71	36.71	35	38.99	38.99	38.16
Soccer2	QP-V1	32	46	38	30	42	36
	Rate-V1	195	196	338	254	244	406
	PSNR-V1	29.86	30.9	33.84	30.19	32.36	34.6
	QP-V0	36	36	40	32	32	34
	Rate-V0	441	441	304	718	718	559
	PSNR-V0	34.03	34.03	32.5	35.6	35.6	34.81
Pantomime	QP-V1	34	48	40	34	48	36
	Rate-V1	242	242	373	242	242	501
	PSNR-V1	30.62	30.38	34.21	30.62	30.38	36.37
	QP-V0	36	36	40	32	32	36
	Rate-V0	516	516	383	761	761	516
	PSNR-V0	36.26	36.26	34.1	38.57	38.57	36.26
Dog	QP-V1	34	42	36	34	42	36
	Rate-V1	94	138	186	94	138	186
	PSNR-V1	31.17	32.62	34.88	31.17	32.62	34.88
	QP-V0	36	36	38	32	32	34
	Rate-V0	197	197	174	280	280	232
	PSNR-V0	34.52	34.52	33.7	36.04	36.04	35.3

In the DSCQS method, the tested condition is always compared to the source. Thus, the assessor has to remember and then judge his/her previous tested conditions and the related scorings. Inevitably, this causes the misjudgment of the assessor in the preceding pairs of the test, as well as among the reference and test pairs. Moreover, with a contextual effect, such as if a low-quality test is followed by a high-quality test, the score may be lower than that where a low-quality test is followed by a low- or medium-quality test [25]. The DSCQS is effective while comparing stimuli that have small differences and it provides a global quality score for a short display duration.

On the other hand, the ACR and SSCQE are suitable for stimuli with detectable differences within a wide quality range. The 5/9/11-point scale is used for the ACR, whereas a 0–100 continuous scale is used for the SSCQE. The SAMVIQ is inspired by the standardized DSCQS method and uses a nonsequential voting process that avoids misjudgments due to the ephemeral nature of the DSCQS method. In [26], it was shown that SAMVIQ scores have greater accuracy and better differentiate stimuli than ACR scores for the same number of assessors. Since we have heterogeneous stimuli and/or small and large differences, it is appropriate and safe to use the SAMVIQ according to [21].

2.3.2. SESVIQ

Figure 5 gives a screenshot of the SESVIQ test method. The SESVIQ was developed as an extension of the SAMVIQ for stereoscopic videos and presented in [27] by comparing it to the DSCQS. The interactive evaluation

is carried out scene after scene. We use 4 different stereo scenes in each group of tests and 5 and 6 algorithms at the test setup for Groups 1 and 2, respectively. In order to get an anonymous display, we use buttons labeled with one letter. The number of letter-labeled buttons is equal to the number of algorithms to test plus one, for the original. Additionally, the button labeled 'Ref' is included as well.

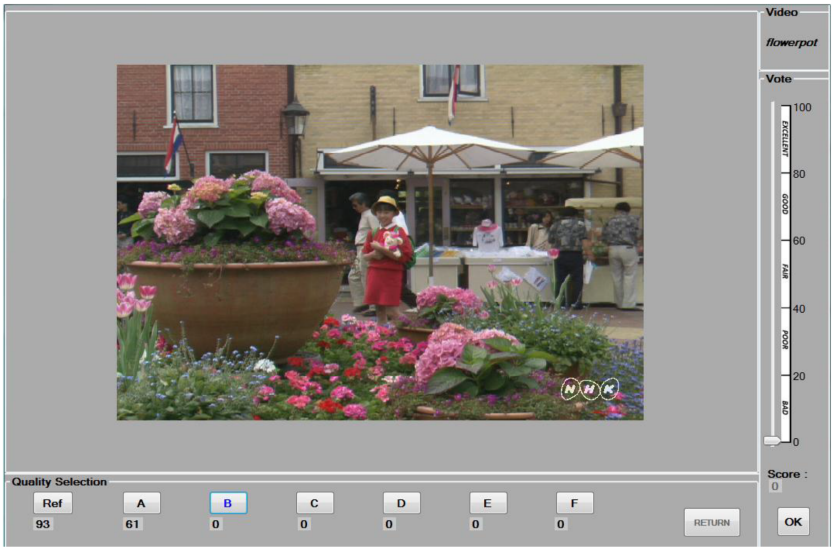

Figure 5. Flowerpot scene of Group 1 under assessment with the SESVIQ program.

One anchor (the original of the scene) is accessed under the Ref button serving as an explicit reference and the other is accessed randomly under any of the algorithm buttons ('A'–'F' in Figure 5) as a hidden reference. These 2 anchors help to stabilize the results and gain more reliability. The experimental results show that including an explicit reference decreases the standard deviation when compared to a hidden or no reference [22]. The hidden reference is also added to evaluate the intrinsic quality of the reference video. For each scene, the SESVIQ test program includes 7 buttons to be scored for the Group 1 test and 8 buttons for the Group 2 test.

While evaluating the current scene, the assessor can play and score any algorithm in any order, and it can then replay and rescore. To test the next scene, all of the algorithms of the current scene must be scored. From one scene to another, the algorithm's access is randomized so that the assessors can be avoided to attempt scoring similarly according to the button labels. To score, the assessor moves a slider on the 0–100 impairment scale corresponding to 5 quality items divided linearly, such as excellent, good, fair, poor, and bad. For the analysis of the results, the difference scores of the hidden reference and test conditions are taken into account.

2.4. Participants

Two groups of assessors participate in the experiments. One group rates 4 test sequences, i.e. Flower2, Flowerpot, Horse, and Car, with both the DSCQS and SESVIQ methods. This group consists of 15 viewers, including 12 males and 3 females with mean ages of 26.3 and 27.7, respectively. The other group rates the other 4 test sequences, i.e. Flower3, Soccer2, Pantomime, and Dog with SESVIQ, but Dog is excluded in the DSCQS test because of a time constraint. The latter group consists of 15 viewers, including 10 males and 5 females with mean ages of 24.2 and 30.0, respectively. The aim of the 2 groups of assessors is to get more statistical data and more reliable subjective evaluation. However, a few of the volunteers participate in both groups of tests. All of the assessors are recruited from Yaşar University, mostly graduate and undergraduate senior students, and a

number of academicians. They are checked to confirm that they have normal stereo depth perception, which is the Randot test, and normal far and near visual acuity, color vision, and contrast sensitivity. Only the assessors who pass the screening tests participate in the experiments; thus, the one person who failed the Randot test is not enrolled. Assessors are not aware of the purpose of the experiment and none of them are experts in video quality assessment.

3. Experimental results and discussion

Aside from the testing procedure, ITU-R BT.500-11 also includes recommendations for the statistical analysis of the collected data [19]. A common approach in single and double stimulus test methods is to examine the distribution of the integer values changing between 0 and 100 and calculate a MOS value. In a paired comparison, the difference is taken between the scores of the test and original algorithms, and then the differences in the values are normalized to 0–100. Last, 95% confidence interval values are calculated. The MOS value should be presented with the associated confidence interval in order to describe the accuracy of the subjective assessment tests. A larger MOS indicates a poorer level of perceived video quality. Likewise, a larger confidence interval causes an inference of a poorer level of reliability.

3.1. Results of the subjective tests with the DSCQS and SESVIQ methods

In Tables 3 and 4, the MOS values with 95% confidence intervals for the DSCQS tests of Groups 1 and 2 are presented, while Figures 6 and 7 show the SESVIQ test results for Groups 1 and 2, respectively. First, 3 options have almost the same bitrate, which is called the low bitrate scenario and the other options are for the high bitrate scenario. Remember that the selected bitrate and QP values are given in Tables 1 and 2 for Groups 1 and 2 in Section 2.2., respectively. From the results of Group 1, it is clear that Asym/Sym SNR is better than Asym Spa for all of the videos at a high bitrate. In the low bitrate scenario for Flower2, the Flowerpot and Horse sequences Asym SNR is better than Sym SNR and Sym SNR is better than Asym Spa in the DSCQS, whereas Sym SNR is better than Asym SNR in the SESVIQ. However, the Car sequence Asym Spa turns out to be very close to Sym SNR, but better than Asym SNR. We conclude from the results that the video content is a very important factor, where, for example, spatial scaling is preferable for the Car sequence since high motion content is more vulnerable to blocking artifacts.

Table 3. Group 1 DSCQS MOS values with 95% confidence interval (CI) of the mean.

	Flower2	Flowerpot	Horse	Car
Original	11.2 ± 0.7	9.8 ± 0.5	9.3 ± 0.5	10.8 ± 0.8
Asym Spa-Lo	50.3 ± 3.8	58.8 ± 2.4	58.6 ± 3.9	56.4 ± 3.5
Asym SNR-Lo	39.9 ± 3.0	53.2 ± 3.5	45.0 ± 3.0	64.6 ± 3.9
Sym SNR-Lo	49.9 ± 2.7	56.1 ± 2.6	50.6 ± 3.1	56.6 ± 2.9
Asym Spa-Hi	47.7 ± 3.3	60.8 ± 4.0	51.4 ± 2.6	45.3 ± 3.3
Asym/Sym SNR-Hi	23.6 ± 1.7	30.4 ± 2.7	30.3 ± 3.5	28.0 ± 3.7

Among the Group 1 sequences, only Car has complex camera and object motion; thus, highly noticeable blockiness is the main artifact leading the scores to turn out with the Spa Asym being the best for the low bitrate. Flowerpot and Horse have relatively high luminance, contrast, and spatial detail. Blur, the loss of spatial detail, is the reason why Spa Asym turned out the worst for the low bitrate.

Table 4. Group 2 DSCQS MOS values with 95% CI of the mean.

	Flower3	Soccer2	PANTOMIME
Original	25.3 ± 2.8	29.2 ± 3.2	24.6 ± 2.6
Asym Spa-Lo	55.9 ± 3.2	52.7 ± 5.9	49.8 ± 3.4
Asym SNR-Lo	60.9 ± 3.8	55.8 ± 3.3	54.8 ± 4.5
Sym SNR-Lo	74.9 ± 5.7	53.2 ± 4.8	40.8 ± 3.4
Asym Spa-Hi	43.9 ± 2.9	44.8 ± 2.3	43.9 ± 3.7
Asym SNR-Hi	46.7 ± 4.1	39.8 ± 3.2	51.8 ± 4.5
Sym SNR-Hi	55.2 ± 4.4	49.9 ± 4.4	34.6 ± 3.2

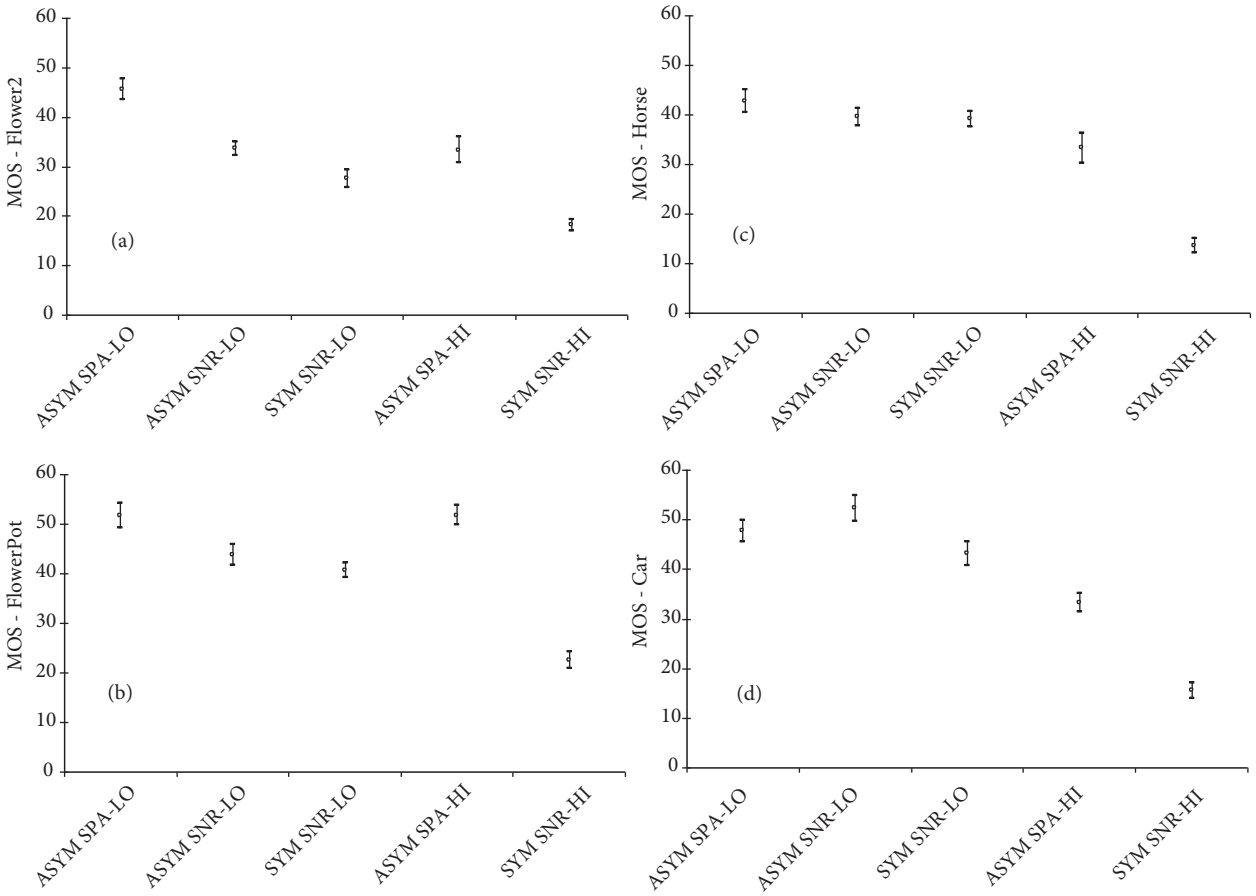

Figure 6. SESVIQ MOS values of the a) Flower2, b) Flowerpot, c) Horse, and d) Car sequences.

The results of Group 2 show that for the Flower3 sequence, at low and high bitrates, Asym Spa is better than both Asym SNR and Sym SNR. Similar to the Car sequence, for Flower3, the blocking artifact is very disturbing due to high motion content. For the Soccer2 sequence, at a low bitrate, Asym Spa is better than Asym/Sym SNR, but at a high bitrate, it is not so clear. In the case of the Pantomime sequence, for low and high bitrates, Sym SNR is the winner for all of the cases. Asym Spa is also not bad at high and low bitrates. Symmetry is important and the matching of blocking artifacts is necessary since it is a sequence that has special color and structures on the black background. The Dog sequence having less color and light may be the reason why the high bitrate results are not so discriminative, while Sym SNR is the winner at a low bitrate.

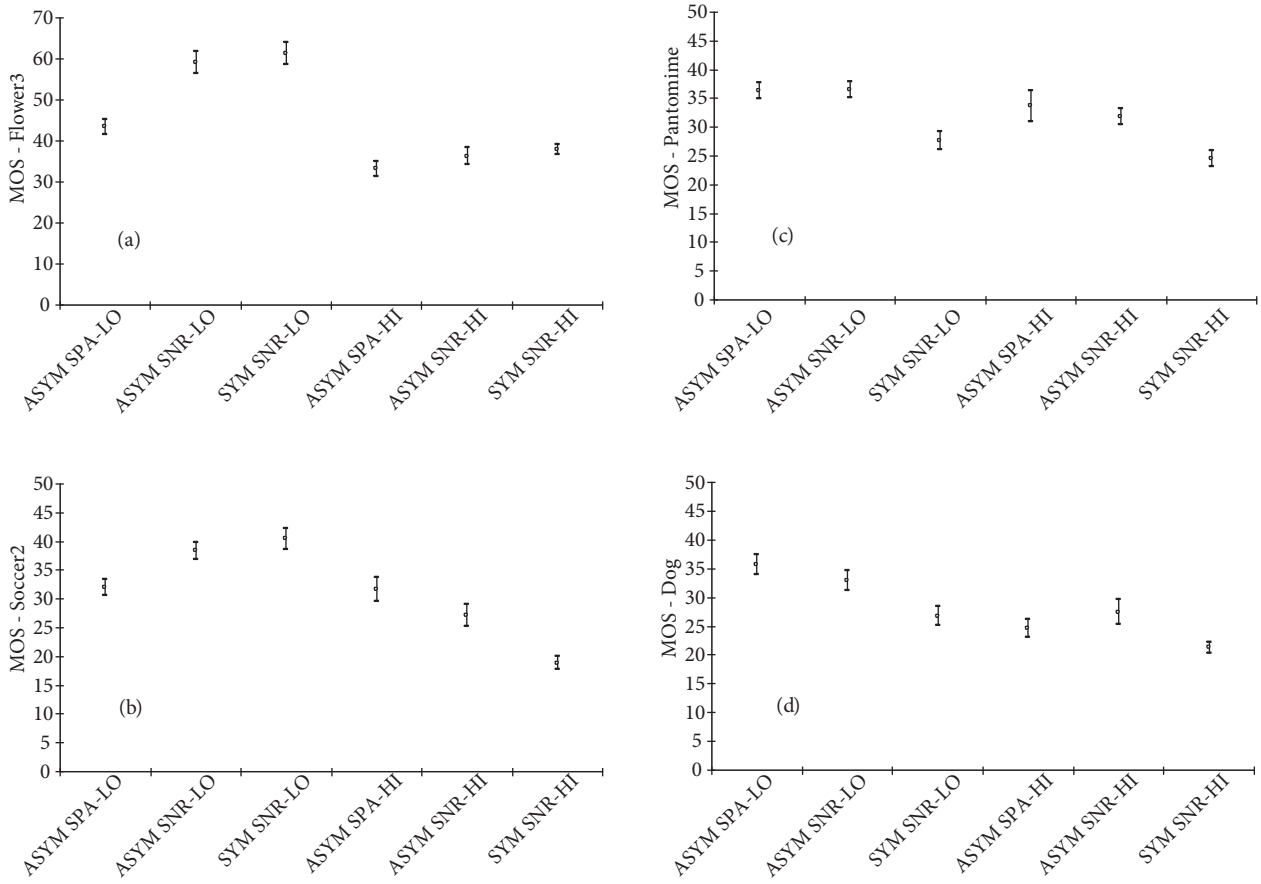

Figure 7. SESVIQ MOS values of the a) Flower3, b) Soccer2, c) Pantomime, and d) Dog sequences.

Of the 2 monocular inputs presented, which have different qualities, the HVS will give more importance to the input that contains more information, because the edges are rich information sources and thus are blocking artifacts. One reason for the reduction in the high-frequency components of the spatial information is low-pass filtering, and the other is DCT-like compression algorithms. At higher values of the QP for the transform coefficients, blocking artifacts may occur, whereas at lower values of the QP, a reduction of the high-frequency components of the spatial information is the most significant artifact of the DCT-like compression. As an example, for some test sequences, the Asym Spa option shows the blocking artifact at a high bitrate, whereas it shows the blurring artifact at a low bitrate. We propose to choose QP values for the auxiliary view where the DCT causes the blurring artifact rather than the blocking artifact, which is in fact content-dependent. When both views are compressed at bitrates in which blocking artifacts start to occur, we propose to match the blocking artifacts in each view, which is in parallel with what we observed in Sym SNR as the winner in general for the moderately critical bitrate.

3.2. 2D subjective test results

To further examine the effect of monocular distortion on binocular perception, 2D tests are conducted using the SESVIQ method in such a way that the left and right view videos are the same. The participants of the 2D visual tests comprise 11 male and 4 female assessors. In Tables 5 and 6, the crosses show the selected sequences for the 2D stimuli from the set of 3D test stimuli.

Table 5. QP values of the 2D test stimuli selected from the stimuli of the Group 1 3D test.

Flower2	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym/Sym SNR-Hi
QP-V1	30	42	40	24	36
QP-V0	36	36	38	32	32
Flowerpot	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym/Sym SNR-Hi
QP-V1	30	44	38	24	34
QP-V0	34	34	36	32	32
Horse	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym/Sym SNR-Hi
QP-V1	30	44	40	24	32
QP-V0	36	36	38	30	30
Car	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym/Sym SNR-Hi
QP-V1	32	44	38	24	32
QP-V0	36	36	40	28	28

Table 6. QP values of the 2D test stimuli selected from the stimuli of the Group 2 3D test.

Flower3	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym SNR-Hi	Sym SNR-Hi
QP-V1	30	48	46	28	42	38
QP-V0	40	40	44	34	34	36
Soccer2	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym SNR-Hi	Sym SNR-Hi
QP-V1	32	46	38	30	42	36
QP-V0	36	36	40	32	32	34
Pantomime	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym SNR-Hi	Sym SNR-Hi
QP-V1	34	48	40	34	48	36
QP-V0	36	36	40	32	32	36
Dog	Asym Spa-Lo	Asym SNR-Lo	Sym SNR-Lo	Asym Spa-Hi	Asym SNR-Hi	Sym SNR-Hi
QP-V1	34	42	36	34	42	36
QP-V0	36	36	38	32	32	34

Tables 7–10 show the V0, V1, and 3D MOS values for all of the tested options and the best correlation values found. A search is made in Tables 7–10 for the determination of α, which holds the best match between the weighted 3D MOS [Eq. (1)] and the 3D SESVIQ scores. Correlation values are calculated by the Pearson

linear correlation. The α values found are given in Tables 7–10.

$$MOS_{3D} = \alpha \cdot MOS_{V0} + (1 - \alpha) \cdot MOS_{V1} \quad (1)$$

From the experiments, for the Car and Flower3 sequences, for all of the scaling options and bitrates, equal weighting of the monocular inputs is observed. These are the sequences that are the most sensitive to blockiness distortion due to high motion content. The results of the Flower2 and Soccer sequences are similar to those of Car and Flower3 for the Sym and Asym SNR options, but for the Asym Spa-Hi option, α is 0.7, which means that a higher quality view tolerates the blur in the lower quality view. For Soccer2 Asym Spa-Lo, the toleration is a little less ($\alpha = 0.6$). For the Pantomime sequence, the case is different than with Flower2 and Soccer2; since $\alpha = 0.4$ for the Asym Spa low and high bitrates, a higher quality view cannot tolerate blur in the lower quality view. Note that Pantomime has special content. For the Flowerpot and Horse sequences, the results show underweighting of the higher quality input (α between 0.2 and 0.4). They have higher spatial detail and more light. More specifically, the Flowerpot sequence has more spatial detail compared to the Horse sequence, so it cannot tolerate spatial scaling artifacts at low or high bitrates.

Table 7. 2D SESVIQ and the weighted 3D MOS values for Flower2 and Flowerpot.

	Flower2				Flowerpot			
	MOS_{V1}	MOS_{V0}	MOS_{3D}	α	MOS_{V1}	MOS_{V0}	MOS_{3D}	α
Asym Spa-Lo	64.3	19.2	41.8	0.5	61.8	21.3	53.7	0.2
Asym SNR-Lo	42.9	19.2	31.1	0.5	60.7	21.3	41.0	0.5
Sym SNR-Lo	34.0	25.0	29.5	0.5	39.8	27.8	33.8	0.5
Asym Spa-Hi	62.8	19.2	32.3	0.7	66.9	18.7	57.3	0.2
Asym SNR-Hi	19.2	19.2	19.2	0.5	21.3	18.7	20.0	0.5
Correlation			0.99				0.97	

Table 8. 2D SESVIQ and the weighted 3D MOS values for Horse and Car.

	Horse				Car			
	MOS_{V1}	MOS_{V0}	MOS_{3D}	α	MOS_{V1}	MOS_{V0}	MOS_{3D}	α
Asym Spa-Lo	57.3	14.8	44.6	0.3	58.3	22.8	40.6	0.5
Asym SNR-Lo	52.5	14.8	37.4	0.4	60.9	22.8	41.9	0.5
Sym SNR-Lo	34.1	19.4	29.7	0.3	38.2	37.3	37.8	0.5
Asym Spa-Hi	63.6	11.6	32.4	0.6	54.7	12.8	33.8	0.5
Asym SNR-Hi	11.6	11.6	11.6	0.5	18.1	12.8	15.5	0.5
Correlation			0.94				0.98	

Table 9. 2D SESVIQ and the weighted 3D MOS values for Flower3 and Soccer2.

	Flower3				Soccer2			
	MOS_{V1}	MOS_{V0}	MOS_{3D}	α	MOS_{V1}	MOS_{V0}	MOS_{3D}	α
Asym Spa-Lo	45.4	33.8	39.6	0.5	56.3	21.0	35.1	0.6
Asym SNR-Lo	74.0	33.8	53.9	0.5	65.3	21.0	34.3	0.5
Sym SNR-Lo	69.6	52.0	60.8	0.5	35.5	35.5	35.5	0.5
Asym Spa-Hi	46.5	21.0	33.8	0.5	56.3	16.7	32.5	0.7
Asym SNR-Hi	52.0	21.0	36.5	0.5	57.7	16.7	29.0	0.5
Sym SNR-Hi	33.8	21.0	27.4	0.5	34.3	16.7	22.0	0.5
Correlation			0.94				0.93	

Table 10. 2D SESVIQ and the weighted 3D MOS values for Pantomime and Dog.

	Pantomime				Dog			
	MOS_{V1}	MOS_{V0}	MOS_{3D}	α	MOS_{V1}	MOS_{V0}	MOS_{3D}	α
Asym Spa-Lo	48.7	15.1	35.3	0.4	52.6	17.8	35.2	0.5
Asym SNR-Lo	51.9	15.1	33.5	0.5	53.8	17.8	32.2	0.6
Sym SNR-Lo	24.2	24.2	24.2	0.5	17.8	17.8	17.8	0.5
Asym Spa-Hi	48.7	15.6	35.5	0.4	52.6	16.7	23.9	0.8
Asym SNR-Hi	51.9	15.6	33.8	0.5	53.8	16.7	24.1	0.8
Sym SNR-Hi	15.1	15.1	15.1	0.5	17.8	16.7	17.3	0.5
Correlation			0.92				0.92	

4. Conclusions

Subjective quality evaluation of asymmetrically coded 3D videos is presented for carefully selected critical bitrates. Dense visual tests show that spatial scaling is generally inferior when compared to SNR scaling, except high motion scenes. For higher bitrates, Sym SNR is usually preferable. In other words, below the critical threshold, Sym SNR is not always the winner; spatial scaling has very big potential for high motion videos. Furthermore, above the critical threshold, Asym SNR wins in general, but Sym SNR is preferable for the types of videos in which the matching of blocking artifacts has great importance. From the experiments, equal weighting of monocular inputs for high motion content and underweighting of higher quality for high spatial detail content is observed.

The results indicate that the scalable coding of each view is necessary for efficient rate scaling. Simulcast coding, where each view is separately encoded as scalable, is not efficient. Thus, scalable multiview video coding is promising for dynamic rate adaptation [28]. For the best perceived 3D video quality, content-adaptive rate scaling in asymmetric coding is inevitable. In future works, we plan to extend the evaluation by including some objective quality metrics, and the monocular tests will also be conducted using the DSCQS method.

Acknowledgments

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under project number 109E145. We thank our reviewers for their helpful comments.

References

[1] K. Müller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, T. Wiegand, "Multi-view video coding based on H.264/MPEG4-AVC using hierarchical B pictures", Proceedings of the International Picture Coding Symposium, 2006.

[2] B. Julesz, Foundations of Cyclopean Perception, Chicago, IL, USA, University of Chicago Press, 1971.

[3] L.B. Stelmach, W.J. Tam, "Stereoscopic image coding: effect of disparate image-quality in left-and right-eye views", Signal Processing: Image Communication, Vol. 14, pp. 111–117, 1998.

[4] L.B. Stelmach, W.J. Tam, D.V. Meegan, A. Vincent, P. Corriveau, "Human perception of mismatched stereoscopic 3D inputs", Proceedings of the International Conference on Image Processing, Vol. 1, pp. 5–8, 2000.

[5] L. Stelmach, W.J. Tam, D. Meegan, A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, pp. 188–193, 2000.

[6] D. Meegan, L. Stelmach, W.J. Tam, "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery", Journal of Experimental Psychology: Applied, Vol. 7, pp. 143–153, 2001.

[7] N. Ozbek, M. Tekalp, "Unequal inter-view rate allocation using scalable stereo video and an objective stereo video quality measure", IEEE International Conference on Multimedia and Expo, pp. 1113–1116, 2008.

[8] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G.B. Akar, M.R. Civanlar, A.M. Tekalp, "Temporal and spatial scaling for stereoscopic video compression", Proceedings of the 14th European Signal Processing Conference, pp. 1–5, 2006.

[9] A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G. Bozdagi Akar, M.R. Civanlar, A.M. Tekalp, "End-to-end stereoscopic video streaming with content-adaptive rate and format control", Signal Processing: Image Communication, Vol. 22, pp. 157–168, 2007.

[10] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, J. Kim, "Asymmetric coding of stereoscopic video for transmission over T-DMB", Proceedings of the 3DTV Conference, pp. 1–4, 2007.

[11] P. Aflaki, M.M. Hannuksela, J. Hakkinen, P. Lindroos, M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video", Proceedings of the 17th IEEE International Conference on Image Processing, pp. 4021–4024, 2010.

[12] G. Saygılı, C.G. Gürler, A. M. Tekalp, "3D display dependent quality evaluation and rate allocation using scalable video coding", Proceedings of the 16th IEEE International Conference on Image Processing, pp. 717–720, 2009.

[13] G. Saygılı, C.G. Gürler, A.M. Tekalp, "Quality assessment of asymmetric stereo video coding", Proceedings of the 17th IEEE International Conference on Image Processing, pp. 4009–4012, 2010.

[14] G. Saygılı, C.G. Gürler, A.M. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming", IEEE Transactions on Broadcasting, Vol. 57, pp. 593–601, 2011.

[15] S.F. Chang, A. Vetro, "Video adaptation: concepts, technologies, and open issues", Proceedings of the IEEE, Vol. 93, pp. 148–158, 2005.

[16] L.M.J. Meesters, W.A. IJsselsteijn, P.J.H. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp. 381–391, 2004.

[17] P. Campisi, P. Le Callet, E. Marini, "Stereoscopic images quality assessment", Proceedings of the 15th European Signal Processing Conference, 2007.

[18] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G.B. Akar, "Towards compound stereo-video quality metric: a specific encoder-based framework", IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 218–222, 2006.

[19] International Telecommunication Union, Rec. ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, Geneva, Switzerland, ITU, 2002.

[20] International Telecommunication Union, Rec. ITU-T P.910, Subjective Video Quality Assessment Methods for Multimedia Applications, Geneva, Switzerland, ITU, 2000.

[21] European Broadcasting Union, EBU-UER BNP 056, Technical Report, SAMVIQ – Subjective Assessment Methodology for Video Quality, Geneva, Switzerland, EBU, 2003.

[22] J.L. Blin, "New quality evaluation method suited to multimedia context SAMVIQ", International Workshop on Video Processing and Quality Metrics, 2006.

[23] A. Smolic, G. Tech, H. Brust, "Report on generation of stereo video database", Mobile3DTV Technical Report D2.1, 2010.

[24] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint Scalable Video Model JSVM-4", Joint Video Team, Doc. JVT-Q202, 2005.

[25] V. Baroncini, "New tendencies in subjective video quality evaluation", IEICE Transactions on Fundamentals of Electronics Communications, Vol. E89-A, pp. 2933–2937, 2006.

[26] D.M. Rouse, R. Pepion, P.L. Callet, S.S. Hemami, "Tradeoffs in subjective testing methods for image and video quality assessment", Human Vision and Electronic Imaging XV, Proceedings of the SPIE, Vol. 7527, pp. 75270F–75270F-11, 2010.

[27] N. Özbek, G. Ertan, O. Karakuş, "Interactive quality assessment for asymmetric coding of 3D video", 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2011.

[28] N. Özbek, "Trellis-based optimization of layer extraction for rate adaptation in real-time scalable stereo video coding", Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 20, pp. 557–567, 2012.