**Author(s)** Aflaki, Payman; Rusanovskyy, Dmytro; Utriainen, Timo; Pesonen, Emilia; Hannuksela, Miska M.; Jumisko-Pyykkö, Satu; Gabbouj, Moncef

**Title** Study of Asymmetric Quality Between Coded Views in Depth-Enhanced Multiview Video Coding

**Citation** Aflaki, Payman; Rusanovskyy, Dmytro; Utriainen, Timo; Pesonen, Emilia; Hannuksela, Miska M.; Jumisko-Pyykkö, Satu; Gabbouj, Moncef 2011. Study of Asymmetric Quality Between Coded Views in Depth-Enhanced Multiview Video Coding. International Conference on 3D Imaging 2011, IC3D 2011, December 7-9, 2011, Liège, Belgium. International Conference on 3D Imaging Piscataway, NJ, IEEE. 1-8.

**Year** 2011

**DOI** http://dx.doi.org/10.1109/IC3D.2011.6584381

**Version** Post-print

**URN** http://URN.fi/URN:NBN:fi:tty-201409231440

# STUDY OF ASYMMETRIC QUALITY BETWEEN CODED VIEWS IN DEPTH-ENHANCED MULTIVIEW VIDEO CODING

*Payman Aflaki*[a], *Dmytro Rusanovskyy*[b], *Timo Utriainen*[c], *Emilia Pesonen*[c], *Miska M. Hannuksela*[b], *Satu Jumisko-Pyykko*[c], *Moncef Gabbouj*[a]

[a]Department of Signal Processing, Tampere University of Technology, Tampere, Finland;
[b]Nokia Research Center, Tampere, Finland;
[c]Human-Centered Technology, Tampere University of Technology , Tampere, Finland;

**ABSTRACT**

Depth-enhanced multiview video formats, such as the multiview video plus depth (MVD) format, enable a natural 3D visual experience which cannot be brought by traditional 2D or stereo video services. In this paper we studied an asymmetric MVD technique for coding of three views that enabled rendering of the same bitstream on stereoscopic displays and multiview autostereoscopic displays. A larger share of bitrate was allocated to a central view, whereas two side views were coded at lower quality. The three decoded views were used by a Depth-Image-Based Rendering algorithm (DIBR) to produce virtual intermediate views. A stereopair at a suitable separation for viewing on a stereoscopic display was selected among the synthesized views. A large-scale subjective assessment of the selected synthesized stereopair was performed. A bitrate reduction of 20% on average and up to 22% was achieved with no penalties on subjective perceived quality. In addition, our analysis shows that a similar bitrate reduction gain with no difference in subjective quality can be achieved in multiview autostereoscopic display scenario

**Index Terms—** 3DV, MVC, asymmetric quality multi-view video, subjective assessment.

## 1. INTRODUCTION

3D video coding standardization in the Moving Picture Experts Group (MPEG) is a recent activity targeting at enabling a variety of display types and preferences including varying baseline to adjust the depth perception. Another important target of the MPEG 3DV standardization is the support for multiview autostereoscopic displays (ASDs), thus many high-quality views shall be available in decoder/display side prior to displaying. Due to the natural limitations of content production and content distribution technologies, there is no way that a large number of views can be delivered to user with existing video compression standards. Therefore, MPEG issued a Call for Proposals for 3D video coding (hereafter referred to as the 3DV CfP) [1] for a new standard which enables rendering of a selectable number of views without increasing the required bitrate.

One candidate for 3D video presentation is ASD, emitting more than one stereopairs at a time enabling glass-less 3D perception. However, the ASD technology ensures that subjects observe only one stereopair at a time and subjects can change their viewpoint and consequently observe different stereopairs of the same 3D scene. For this purpose many views should be available for the autostereoscopic display. A multiview video plus depth (MVD) format [2], where each video data pixel is associated with a corresponding depth map value, allows reducing the input data for the 3DV systems significantly, since most of the views will be rendered from the available decoded views and depth maps using a DIBR [3] algorithm. Autostereoscopic displays provide a larger viewing angle and as a result a wider camera separation is needed. Hence, as proposed by the 3DV CfP, a 3-view MVD coding scenario is suitable for creation of a wide range of required views for multiview ASD rendering while a suitable pair of synthesized views can also be used for rendering on a stereoscopic display.

The 3DV CfP, in addition to the data format, targets the development of new 3DV coding technologies. The MVD format can be considered as one of the most potential approaches for the 3DV CfP. The Multiview Video Coding extension of the Advanced Video Coding standard (H.264/MVC) [4] is the state-of-the-art standard in the field of multiview video coding. H.264/MVC can be applied for coding of MVD data, for example by coding the multiview texture video as one H.264/MVC bitstream and the respective multiview depth video as another H.264/MVC bitstream. Despite the high coding efficiency of H.264/MVC, the resulting bitrate of coded multiview data exceeds the bandwidth reserved for conventional 2D video services by a great margin. As a result, more research has been focused on possible approaches to reduce the bitrate of coded multiview video while preserving subjective quality of decoded views and preserving the compatibility with the H.264/MVC standard.

Asymmetric stereoscopic video has been researched as one of the possible solutions to reduce the bitrate and/or computational complexity. In asymmetric stereoscopic video, the two views have different visual quality. The usage of this technique is motivated by the binocular rivalry theory [5], which claims that the stereoscopic vision in human visual system (HVS) fuses the images of a stereopair so that the visual perceived quality is closer to that of higher quality view. Several subjective quality evaluation studies have been conducted to research the utilization of the binocular rivalry theory in stereoscopic video. For example, a set of subjective tests comparing symmetric, quality-asymmetric stereoscopic video coding were conducted in [6]. The presented results showed that subjective quality of symmetric and asymmetric stereoscopic videos provided similar quality under the same bitrate constraint.

In this paper, we study the applicability of asymmetric coding for the three-view (C3) test scenario of the 3DV CfP. We propose a 3-view coding arrangement, where the central view can be extracted for 2D viewing, a central stereopair can be derived for viewing on stereoscopic displays, and a multitude of views can be generated through DIBR for viewing on a multiview ASD. The side views are proposed to be coded at lower quality compared to the quality of the central view, referred to as "full quality". Consequently, the proposed method yields bitrates that are significantly lower than the bitrates of the corresponding symmetric full-quality bitstreams. A subjective assessment in a typical stereoscopic viewing environment was conducted to compare the proposed scheme with a conventional symmetric scheme. The results of the subjective evaluations confirmed that a significant decrease in bitrate (20% of bitrate reduction on average for best tested scheme) was achieved with no degradation in the subjection quality. Moreover, we objectively confirmed the applicability and efficiency of the proposed asymmetric scheme for ASD utilization.

The rest of the paper is organized as follows. Section 2 presents the utilized quality-asymmetric MVD coding scheme. The performed experiments are described in Section 3, while Section 4 provides the results. Finally, the paper concludes in Section 5.

## 2. ASYMMETRIC CODING FOR 3D MULTIVIEW VIDEO

This Section introduces the proposed three-view MVD coding method that utilizes asymmetric transform-domain quantization between views. In order to provide grounds for the proposed coding method, a review of asymmetric stereoscopic video coding and a description of rendering of 3D video on different types of displays are given in Sections 2.1 and 2.2, respectively. Then, we present the proposed coding method in Section 2.3.

### 2.1 Asymmetric Stereoscopic Video Coding

Asymmetric stereoscopic video coding includes a large variety of encoding schemes which provide a quality difference between two views. If different encodings are applied for left and right view, the coding artefacts of one method in the lower quality view can be masked by details presented in the higher quality view. 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. Asymmetry between the two views can be achieved by one or more of the following methods:

a) Mixed-resolution (MR) stereoscopic video coding, first introduced in [7], 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, i.e., represented by fewer pixels.
b) Mixed-resolution chroma sampling [8]. 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 [9] . 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 to 255 (i.e., 8 bits per sample) while the range may be scaled to the range of 0 to 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.
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.

It was found in [10] that the perceived quality of video clips produced using asymmetric transform-domain quantization was approximately equal to the average of the perceived qualities of the two views individually. The impact of the quantization of transform coefficients was verified in [11], 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 stereopairs. Furthermore, the same conclusion of the perceived quality of asymmetric transform-domain quantization was also reached in [6].

### 2.2 Rendering of 3D Video on Stereoscopic and Autostereoscopic Displays

More than two views are rendered simultaneously on a multiview autostereoscopic display. As stated earlier, many multiview autostereoscopic displays provide a wider separation of views as typical stereoscopic displays. At a given time a user sees two views, but by changing the head position the user is able to look at other stereopairs of the rendered views. For example, the display chosen for the 3DV CfP, Dimenco BDL5231V3D, renders 28 views. Hence, to address the required wider separation of views

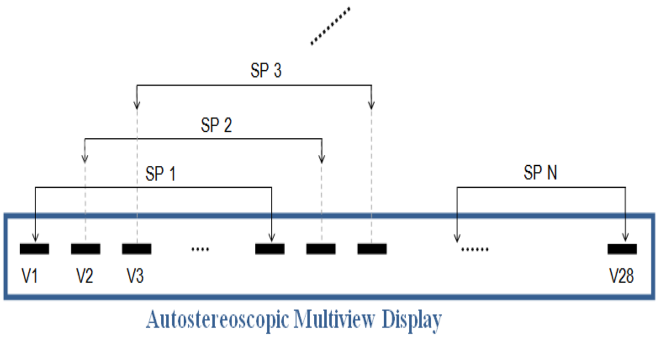



**Fig. 1**. Perceivable sliding stereopair for ASD

required by the Dimenco display and many other autostereoscopic displays, the 3DV CfP includes a 3-view scenario, which is suitable for creation of a wide range of required views.

Fig. 1 illustrates the fact the user sees two views (stereopair SP x) at a time out of the 28 views of the Dimenco display and can choose his/her head position among the possible stereopairs. The difference of view numbers in a stereopair depends on several factors such as the interpupillary distance, the viewing distance, and the rendering parameters of the display. We assumed that views N and N+8 out of the 28 views displayed by the Dimenco display could be considered to form a stereopair in a typical multiview autostereoscopic viewing situation.

When a 3-view bitstream is adapted for comfortable viewing on a stereoscopic display, two views at a suitable separation have to be extracted. It can be assumed that content providers would appreciate that the mid-most stereopair is selected for such stereoscopic viewing, hence providing a "central" viewpoint to the content, rather than picking views from either side of the 3-view setting. Therefore the decoded views in a 3-view bitstream are not displayed as such for rendering on a stereoscopic display, but a suitable stereopair is synthesized from the decoded views.

### 2.3 Proposed Asymmetric Three-View MVD Coding

Inspired by the promising results of asymmetric stereoscopic video coding (see Section 2.1) we wanted to design an asymmetric coding scheme for 3-view MVD format. We started off with the following requirements and assumptions:

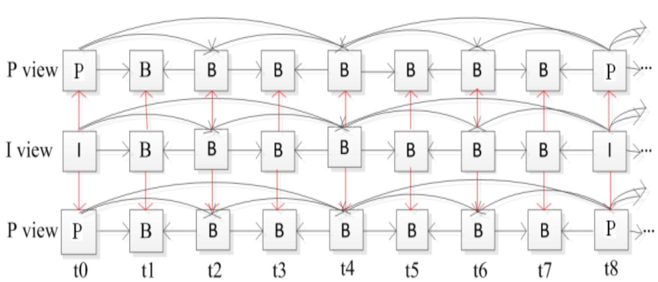



**Fig. 2.** PIP interview prediction structures

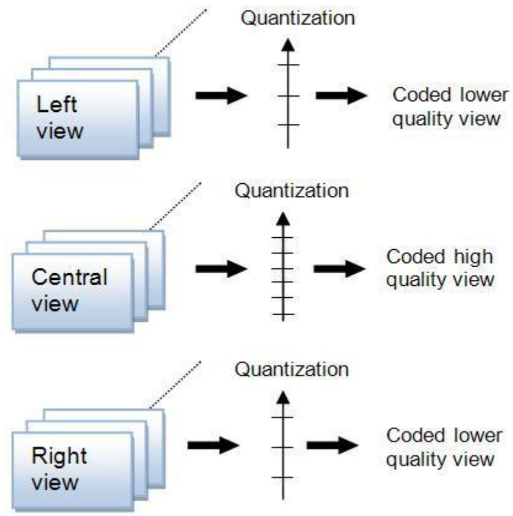
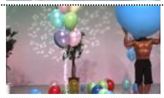



**Fig. 3.** Proposed asymmetric MVC scheme for 3DV coding

1. H.264/AVC decoders must be able to extract 2D video from the 3-view MVD bitstreams in order to obtain compatibility with existing decoders.
2. The extracted 2D video should be the central view of the content as it is likely to represent the captured 3D scene more appropriately than the side views.
3. No compromise on the 2D video quality should be made in the asymmetric coding, as for the time being 2D viewing is still more common than 3D viewing.
4. Two midmost views at a suitable separation should be generated for viewing on stereoscopic displays (see Section 2.2). The quality of the midmost views for stereoscopic viewing should not be compromised compared to the corresponding views obtained from symmetric 3-view MVD bitstreams.
5. Such number of intermediate views that suits the multiview ASD being used should be able to be created from the decoded 3-view MVD bitstream. The average quality of the perceived stereopairs on the multiview ASD should be similar to that obtained from symmetric 3-view MVD bitstreams.

In order to respond to the requirements 1 to 3, we decided to use H.264/MVC independently for texture and depth views and select the PIP inter-view prediction structure for encoding (see Fig. 2). In this structure the central view is coded with H.264/AVC and utilized as reference for coding of the two side views. The transform coefficients of the central (base) view are quantized using a fine quantization parameter (QP0) while the side views are quantized more coarsely with a quantization parameter QP0 + deltaQP (Fig. 3). This will result in a lower quality of the side views compared to the quality of the base view and consequently brings a bitrate reduction comparing the symmetric quality case where all views are encoded using the same QP0. Alternatively, it is possible to encode an asymmetric MVD bitstream with the same bitrate as a

Table 1. Sequences and their characteristics

| Screenshot | Sequence | Resolution | Frames | Framerate |
|---|---|---|---|---|
|  | Poznan Hall2 | 1920x1080 | 250 | 25 |
|  | Undo Dancer | 1920x1080 | 250 | 25 |
|  | Balloons | 1024x768 | 300 | 30 |
|  | Newspaper | 1024x768 | 300 | 30 |

symmetric MVD bitstream by coding the central view with a QP value lower than the QP for the symmetric MVD.

Asymmetric transform-domain quantization was chosen because it enables the realization of the proposed coding scheme with the H.264/MVC standard as an encoding method without changes to the bitstream format or the decoder. In principle, a similar asymmetric MVD coding scheme could utilize also utilize the other types of asymmetry listed in Section 2.1.

In the rest of this paper, we study how the requirements 4 and 5 above are met with the proposed asymmetric MVD coding.

## 3. DESCRIPTION OF THE EXPERIMENTS

As introduced in the Section 2.3, we performed experiments to clarify whether the proposed asymmetric three-view MVD coding scheme is beneficial for stereoscopic viewing and multiview autostereoscopic viewing when compared to symmetric MVD coding. We decided to carry out a large-scale systematic subjective evaluation experiment for the stereoscopic viewing, because impact of view synthesis on the perceived quality has not been explored earlier in a similar viewing scenario as much as we are aware and the usage of objective metrics would have been therefore questionable. With the results obtained from the stereoscopic viewing test, we were able to make assumptions on the behavior of objective metrics with respect to perceived quality and hence generalize the findings of the stereoscopic viewing assessment for multiview autostereoscopic displays.

In Section 3.1 we introduce the test material, the compared coding scenarios, and the test stimuli preparation from the decoded bitstreams. The setup and procedure of the subjective viewing experience evaluation on a stereoscopic display are presented in Section 3.2.

Table 2. Test sequences, input views, and synthesized views

| Sequence | Input views | Synthesized views |
|---|---|---|
| Poznan Hall2 | 7-6-5 | 6.25-5.75 |
| Undo Dancer | 1-5-9 | 4-6 |
| Balloons | 1-3-5 | 2.5-3.5 |
| Newspaper | 2-4-6 | 3.5-4.5 |

Table 3. Tested rate points an corresponding QP settings (SS)

| Sequence | R2 | | R4 | |
|---|---|---|---|---|
| | **QP0** | Bitrate, Kbps | **QP0** | Bitrate, Kbps |
| Poznan Hall2 | **36** | 715.3 | **28** | 1747.1 |
| Undo Dancer | **40** | 1369.4 | **36** | 2296.8 |
| Balloons | **44** | 374.8 | **35** | 973.9 |
| Newspaper | **40** | 553.7 | **34** | 1049.3 |

### 3.1 Test stimuli

In this section the detailed steps of test material preparation is described. Four sequences, Undo Dancer, Newspaper, Poznan Hall2 [12], and Balloons, included in the 3DV CfP test set were used. Basic parameters of these sequences, such as resolution, and frame rates are given in Table 1.

The selected input and output (synthesized) views in our experiments are shown in Table 2. Note that input views for all tested sequences are the same as specified in the 3DV CfP for the C3 case. Based on the discussion presented in sub-section 2.3, we chose a stereopair for each sequence from the center of the three coded views such that the baseline difference of the chosen views suits viewing on a stereoscopic display. It is noted that the view separation of the stereopairs specified in the 3DV CfP for C3 is approximately half of a conventional stereo baseline. Such narrow view separation was chosen to mimic the viewing conditions on typical autostereoscopic multiview displays. As we targeted for a different use case, the synthesized views utilized in stereo viewing were selected differently from the MPEG 3DV CfP.

The proposed asymmetric MVC scheme was implemented on the top of JM 17.2 reference software [13]. The software was configured to produce test materials with three MVC schemes:

- Symmetric Scheme (SS): all views were coded with equal QP = QP0.
- Asymmetric Scheme 1 (AS1), the central view was coded at QP=QP0, and the side views were coded with QP=QP0+2
- Asymmetric Scheme 2 (AS2): the central view was coded at QP=QP0, and the side views were coded with QP=QP0+4

The SS bitstreams were encoded for two bitrate points referred to as R2 and R4 where the QP values were selected equal to those of the H.264/AVC anchor encoding of the 3DV CfP. The bitrates and the utilized QP values of these SS bitstreams are provided in Table 3. The other encoding settings were chosen to comply with the requirements of the 3DV CfP.

The tested schemes (SS, AS1 and AS2) produced identical bitrates for the coded central views, whereas the bitrates for the coded side views in AS1 and AS2 were significantly reduced compared to the SS. An average total

Table 4. Bitrate reduction of proposed asymmetric method compared to symmetric coding method

| **Sequence** | **AS1** (%) | **AS2** (%) |
|---|---|---|
| Poznan Hall2 | 12.9 | 22.1 |
| Undo Dancer | 10.1 | 16.8 |
| Balloons | 12.4 | 19.6 |
| Newspaper | 13.1 | 22.2 |
| **Average** | **12.6** | **20.2** |

bitrate reduction compared to SS was 12.6% for AS1 and 20.2% for AS2. The detailed bitrate reduction in a sequence basis is reported in Table 4.

Stereopairs, as specified in Table 2, were synthesized for each of tested schemes (SS, AS1, AS2). The view synthesis was performed with VSRS software, version 3.5 [14]. We utilized VSRS configuration files and camera parameters information, as they are specified in the MPEG 3DV CfP.

For the multiview autostereoscopic viewing we assumed the use of Dimenco BDL5231V3D autostereoscopic display as specified in the 3DV CfP. The 28 views required by the Dimenco display were produced as follows. First, we synthesized the views as described in the 3DV CfP, resulting into 49 and 33 coded or synthesized views for Newspaper and the rest of the sequences, respectively, and then we picked the 28 mid-most ones for rendering.

### 3.2 Subjective Quality Evaluation on Stereoscopic Display

25 subjects, (18 female, 7 male), aged between 19-29 years (mean: 23.9) participated in this experiment. A majority (84%) of them were considered naïve as they did not work or study in fields related to information technology, television or video processing.

The test session comprised three parts: 1) pre-test sensorial screening and demo-/psychographic data-collection, 2) actual voting using quantitative data-collection, and 3) post-test interview with qualitative data-collection. The candidates were subject to thorough vision screening. Candidates who did not pass the criterion of 20/40 (near and far vision, Landolt chart) visual acuity with each eye or color vision (Ishihara) were rejected. All participants had a stereoscopic acuity of 60 arc sec at the minimum.

The laboratory conditions were organized according to [15]. Hyundai 46-inch stereoscopic monitor model S465D with passive polarizing glasses was used, as suggested by the 3DV CfP. Furthermore, the viewing distance was four times the height of the image (2.29m for 1920×1080 and 1.63m for 1024×768 video sequences), as specified by the 3DV CfP.

The subjective test started with a combination of anchoring and training. The extremes of the quality range of the stimuli were shown to familiarize the participants with the test task, the test sequences, and the variation in quality they could expect in the actual tests that followed. The test clips were presented one at a time in random order and appeared twice in the test session following the ITU recommendation Double Stimulus Impairment Scale (DSIS) method [15]. A discrete unlabeled quality scale from 0 to 10 was used for the rating scale. The viewers were instructed that 0 stands for the lowest quality and 10 for the highest.

The post-test sessions contained a semi-structured interview that gathered the participant's impressions, experiences and descriptions of the visual quality to deepen the understanding behind the decisions of the participant [16]. The interview was constructed of main and supporting questions. The main question *"What kind* of *factors did you pay attention to while evaluating quality?"* was asked several times with slight variations during the interview. The moderator only used the terms introduced by the participant when asking the supporting questions to further clarify the answers to the main question: *"Please could you clarify if X was among the positive/negative factors or pleasant/unpleasant?"* and *"Which of the factors you mentioned was the most pleasant/unpleasant?"*.

The participant filled the Simulator Sickness Questionnaire (SSQ) [17] before and after the actual test. The questionnaire measures sickness symptoms using a weighted average of nausea, oculomotor and disorientation scores, while also calculating a composite score.

## 4. RESULTS

In this section we present the results and the analysis of the performed experiments. Section 4.1 presents the statistics of the quantitative viewing experience ratings on the stereoscopic display. The qualitative results obtained from the post-test interviews are summarized in Section 4.2. In Section 4.3 we analyze the obtained results through selected objective metrics and draw conclusions on the expected performance of the proposed method for multiview autostereoscopic rendering.

### 4.1 Quantitative quality evaluation

Fig. 4 shows the average and the 95% confidence interval (CI) of the subjective viewing experience ratings for all sequences in two different bitrates. The naming introduced in sub-section 3.2 is used and O stands for the original uncompressed sequences.

As can be judged from the average ratings and confidence intervals presented in Fig. 4, no significant differences were perceived between the encoding schemes for the same value of QP0. In other words, it can be observed from Fig. 4 that the described asymmetric scheme provided the same subjective quality as the symmetric scheme with the test material. However, the utilization of asymmetric coding is preferred since on average the bitrate was reduced by 12.6% and 20.2% for AS1 and AS2, compared to SS, respectively (see Table 4). The observation that there were no significant differences between the

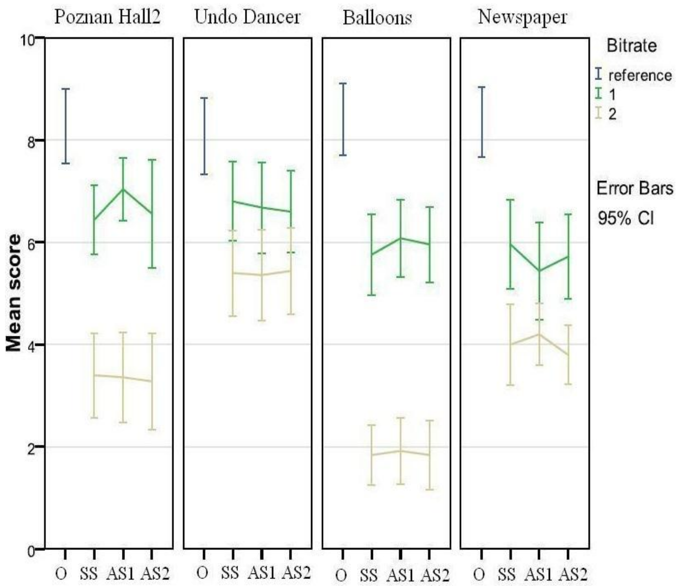



**Fig. 4.** Viewing experience ratings (O = original uncompressed sequence, SS = symmetric coded sequence, AS1, AS2 = asymmetric transform-domain quantization between coded views)

encoding schemes was further verified using statistical analysis as presented in the paragraphs below.

Non-parametric statistical analysis methods, Friedman's and Wilcoxon's tests, were used as the data did not reach normal distribution (Kolmogorov-Smirnov: $p<.05$). Friedman's test is applicable to measure differences between several and Wilcoxon's test between two related and ordinal data sets [18]. A significance level of $p < .05$ was used unless otherwise stated below.

The results of the Friedman's test verified that there were no significant differences between the encoding schemes in the test stimuli except that AS1 was rated higher than SS in bitrate 1 of Poznan Hall2 ($p<.05$). Furthermore, no significant difference ($p<0.001$) was observed for the subjective rating of different schemes for each sequence and for both proposed bitrates in the performed Wilcoxon pairwise comparisons either.

As explained earlier, the side views in the AS1 and AS2 coding scenarios had lower objective quality as the side views in the SS coding scenario. The objective quality as measured by the average luma Peak-Signal-to-Noise Ratio (PSNR) metric for the tested coding scenarios and sequences is reported in Table 5. The applied view synthesis algorithm utilized the decoded texture and depth views that are adjacent to the view being synthesized, hence the objective quality of the synthesized stereopair of AS1 and AS2 was lower than that of SS. This objective quality difference was analyzed by first deriving the average luma PSNR of each synthesized view of ratepoint R4 against the views synthesized from uncompressed data, so called Reference View Synthesis (RVS). Then, the the average of the PSNR values of the two views of the selected stereopair (denoted aPSNR) was taken and finally the difference (dPSNR) of aPSNR(ASx) to aPSNR(SS) was computed. The values of dPSNR are reported in Table 6. It can be seen that the dPSNR results for the synthesized stereopair of Dancer, Balloons, and Newspaper sequences contradicted with the obtained results of the subjective quality evaluation experience, hence giving an indication that two stereopairs having an average luma PSNR difference smaller than or similar to the ones reported in Table 6 may have an equal quality subjectively.

### 4.2 The results of the post-test interview

The analysis was based on grounded theory and its wide applications to visual and audiovisual quality [16, 19]. All recorded interviews were transcripted to the text as a pre-processing step of analysis. 30% of interviews (7 participants) were used as a base for open coding (read through, extraction of meaningful sentences and coding for creating the concepts and their properties). All concepts were organized into sub-categories and they were further organized under main categories. This phase was conducted by one researcher and reviewed by another researcher. The categorization created was used in the coding of the whole data. One mention per category per person was counted and frequencies in each category were determined by counting the number of participants that described the category.

The descriptive quality of experience was composed of five main components: 1) 3D quality, 2) spatial quality, 3) temporal quality, 4) viewing task and 5) content and quality variation. The most commonly mentioned negative quality factors were associated to spatial quality (inaccuracy in general, or inaccuracy of outlines of objects and details), visibility of impairments with detectable structure, impairments during motion, and hardness or unpleasurable viewing were mentioned by more than 48% of participants. In contrast, the most commonly described positive quality of experience factors were excellence of depth impression and fluency of motion (more than 48% of participants).

Table 5. PSNR of coded views

| | | dQP0 | | dQP2 | | dQP4 | |
|---|---|---|---|---|---|---|---|
| Sequence | View | R2 | R4 | R2 | R4 | R2 | R4 |
| Hall2 | 3 | 38.6 | 41.1 | 38.1 | 40.7 | 37.4 | 40.3 |
| | 4 | 38.6 | 41.1 | 38.6 | 41.1 | 38.6 | 41.1 |
| | 5 | 38.6 | 41.0 | 38.0 | 40.6 | 37.3 | 40.2 |
| Dancer | 1 | 30.4 | 32.2 | 30.1 | 31.8 | 29.7 | 31.4 |
| | 5 | 30.6 | 32.4 | 30.6 | 32.4 | 30.6 | 32.4 |
| | 9 | 30.6 | 32.4 | 30.2 | 32.0 | 29.8 | 31.5 |
| Balloons | 1 | 31.1 | 36.8 | 30.3 | 35.9 | 29.8 | 34.9 |
| | 3 | 31.5 | 37.3 | 31.5 | 37.3 | 31.5 | 37.3 |
| | 5 | 31.1 | 36.7 | 30.2 | 35.7 | 29.7 | 34.8 |
| Newspaper | 2 | 32.6 | 35.8 | 31.6 | 34.9 | 30.7 | 33.9 |
| | 4 | 33.1 | 36.2 | 33.1 | 36.2 | 33.1 | 36.2 |
| | 6 | 32.4 | 35.4 | 31.8 | 34.7 | 31.1 | 34.1 |

### 4.3 Objective analysis of results for ASD utilization

The results presented in Section 4.1 indicated that synthesized stereoscopic videos having a relatively small average luma PSNR difference between them appeared to have subjectively equal quality. In this section we compare the performance of different coding schemes (SS, AS1, AS2) using PSNR and show the efficiency of the proposed asymmetric MVD coding scheme for multiview ASD utilization objectively.

The PSNR analysis similar to that for the stereoscopic case was performed as follows. First, the average luma PSNR of each synthesized view of ratepoint R4 was computed against RVS. Second, the average luma PSNR values of each stereopair covering the full range of all available 28 views for the used ASD (see Fig. 1) were derived. Third, the average luma PSNR difference of the respective stereopair positions (dPSNR) between asymmetric schemes (AS1 and AS2) and SS were calculated. Table 6 provides the average dPSNR over all stereopairs visible in the Dimenco multiview ASD. It can be seen that dPSNR of the stereopair tested subjectively on the stereoscopic display is close to the mean dPSNR of the stereopairs perceivable on the ASD. This confirms that there is not likely to be a considerable difference between quality perception of 3D video in the multiview ASD presentation for the tested SS, AS1, and AS2 sequences. However, the utilization of the proposed asymmetric scheme is preferred, since a bitrate reduction up to 22% was achieved with the tested sequences. Moreover, it could be assumed that viewers tend to choose a position within a viewing cone that is more likely in the center of the cone than on either far side, because a center position allows more flexibility in head movement. Consequently, as the proposed asymmetric coding scheme favors the middle views coded or synthesized for the ASD, it could be even more favorably perceived than what an average quality measure over all possible stereopairs would indicate. Hence, we can assume that the conclusions for the stereoscopic case would hold for autostereocsopic rendering too.

We further investigated whether the PSNR calculation against RVS is a valid metric for our study. For this purpose, we considered the Undo dancer sequence for which we had six original views (out of the total 28 displayed ones) available. The same procedure as described above was performed to calculate the dPSNR values for the available stereopairs against RVS and original views. Results, reported in Table 7, show that there is no remarkable difference between dPSNR against RVS and original views. This confirms the validity of our conclusions utilizing PSNR calculations against RVS.

Table 6. dPSNR of stereopairs, AS1 and AS2 compared to SS.
"ASD, mean" = the mean dPSNR over all stereopairs of the used multiview ASD display.
"SD" = the selected stereopair for the stereoscopic display.

| | AS1-dPSNR (dB) | | AS2-dPSNR (dB) | |
|---|---|---|---|---|
| | ASD, mean | SD | ASD, mean | SD |
| Poznan Hall2 | 0.12 | 0.00 | 0.24 | 0.01 |
| Dancer | 0.15 | 0.11 | 0.29 | 0.21 |
| Balloons | 0.37 | 0.28 | 0.74 | 0.57 |
| Newspaper | 0.20 | 0.11 | 0.40 | 0.22 |

Table 7. Report on dPSNR against RVS and Original for available middle views

| | AS1-dPSNR (dB) | | AS2-dPSNR (dB) | |
|---|---|---|---|---|
| Dancer | ASD, mean | SD | ASD, mean | SD |
| Against Original | 0.11 | 0.09 | 0.21 | 0.16 |
| Against RVS | 0.14 | 0.11 | 0.27 | 0.21 |

## 5. CONCLUSIONS AND FUTURE WORK

In this paper we studied a quality-asymmetric multiview-video-plus-depth coding scheme for the 3-view test scenario specified in MPEG 3DV CfP. The asymmetric quality coding was implemented through coarser transform-domain quantization for the side views, whereas the central view was coded at high quality. Decoded three views were used by a Depth Image Based Rendering algorithm to produce virtual intermediate views that enabled viewing either on stereoscopic displays or multiview autostereoscopic displays. A large-scale subjective assessment of synthesized stereopair was performed on stereoscopic displays and the results showed that a bitrate reduction of 20%, on average, was achieved with no penalties on the perceived quality when compared to coding all the views at a symmetric quality. We also analyzed through objective quality metrics that the described asymmetric coding scheme is also likely to yield subjectively equal quality at the same bitrate reduction factor compared to coding views at symmetric quality when viewed on a multiview autostereoscopic display. As a future task, we plan to verify the conclusions for the multiview autostereoscopic displays with a systematic subjective quality evaluation study.

## 6. ACKNOWLEDGEMENT

The authors would like to thank M. Domañski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner for providing Poznan Hall2 sequence and Camera Parameters [12].

## 7. REFERENCES

[1] MPEG Video and Requirement Groups, "Call for Proposals on 3D Video Coding Technology", MPEG output document N12036, Geneva, Switzerland, March 2011

[2] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," Proc. of IEEE International Conference on Image Processing, vol. 1, pp. 201-204, Oct. 2007.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in Proc. SPIE Conf.

Stereoscopic Displays and Virtual Reality Systems XI, vol. 5291, CA, U.S.A., Jan. 2004, pp. 93–104.
[4] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.
[5] R. Blake, “Threshold conditions for binocular rivalry,” Journal of Experimental Psychology: Human Perception and Performance, vol. 3(2), pp. 251-257, 2001.
[6] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, M. Gabbouj, “Subjective Study on Compressed Asymmetric Stereoscopic Video,” Proc. of IEEE Int. Conf. on Image Processing (ICIP), Sep. 2010.
[7] M.G. Perkins, “Data compression of stereopairs,” IEEE Transactions on Communications, vol. 40, no. 4, pp. 684-696, Apr. 1992
[8] A. Aksay, C. Bilen, G. Bozdagi Akar, “Subjective evaluation of effects of spectral and spatial redundancy reduction on stereo images,” 13th European Signal Processing Conference, EUSIPCO-2005, Turkey, Sep. 2005.
[9] P. Aflaki, M. M. Hannuksela, J. Hakala, J. Häkkinen, and M. Gabbouj, “Joint adaptation of spatial resolution and sample value quantization for asymmetric stereoscopic video compression: a subjective study,” Proc. International Symposium on Image and Signal Processing and Analysis, Sep. 2011.
[10] W.J. Tam, “Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV,” Joint Video Team document JVT-W094, Apr. 2007
[11] P. Seuntiens, L. Meesters, and A. IJsselsteijn, “Perceived quality of compressed stereoscopic images: effects of symmetric and asymmetric JPEG coding and camera separation,” ACM Transactions on Applied Perception, vol. 3, no. 2, pp. 96-109, Apr. 2006
[12] M. Domañski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, "Poznan Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, October 2009.
[13] JM reference software: http://iphome.hhi.de/suehring/tml/download
[14] “View synthesis software manual,” MPEG ISO/IEC JTC1/SC29/WG11, Sept. 2009, release 3.5.
[15] ITU-R Recommendation BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” 2002.
[16] S. Jumisko-Pyykkö, “User-centered quality of experience and its evaluation methods for mobile television,” PhD thesis, Tampere University of Technology, 2011.
[17] R. Kennedy, N. Lane, K. Berbaum, and M. Lilienthal, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness,” Int. J. Aviation Psychology, 3(3), 203–220., 1993.
[18] H. Cooligan “Research methods and statistics in psychology,” (4th ed.). London: Arrowsmith., 2004.
[19] A. Strauss and J. Corbin, “Basics of qualitative research: techniques and procedures for developing grounded theory” (2nd ed.). Thousand Oaks, CA: Sage. 1998.