On comparing different codec profiles of coding methods for mobile 3D television and video

Dominik Strohmeier
Institute for Media Technology
Ilmenau University of Technology
Ilmenau, Germany
dominik.strohmeier@tu-ilmenau.de

Gerhard Tech
Heinrich-Hertz-Institute
Fraunhofer Institute for Telecommunications
Berlin, Germany
gerhard.tech@hhi.fraunhofer.de

***Abstract*—The paper describes the results of a large-scale user study on comparing simple and complex codec profiles for mobile 3D television and video. Four different coding methods – Simulcast, Multiview Video Coding (MVC), Mixed Resolution Stereo Coding (MRSC), and Video+Depth were evaluated at two codec profiles. Baseline profile, using IPPP and CAVLC, and high profile, enabling hierarchical B-frames and CABAC, were under assessment. Our results show that MVC and Video+Depth provide the best overall quality for both coding profiles. In overall, perceived quality of both codec profiles is at the same level. However, the high profile provides the same quality at a lower bit rate.**

Keywords - mobile 3DTV; stereo video coding; quality evaluation; coding methods; codec profiles

I. INTRODUCTION

Mobile 3DTV enables three-dimensional television on mobile devices. However, technological development of the core technology for mobile 3D television and video always has to take into account both system characteristics. Especially in the development of coding methods limited calculation power of the mobile device is a clear tradeoff. Stereo video coding methods have been under developed recently. However, studies on stereo video coding for mobile devices are still missing. Evaluation of 3DTV has focused on larger screen sizes and the impact of artifacts on perceived overall quality [1]. Together with our previous work [4], this paper presents first results on the subjective quality evaluation of coding methods for mobile 3D television and video.

In our approach to find an optimum coding method for mobile 3D television and video [4], four different coding methods were under assessment. However, the first study was limited to the use of Baseline codec profile. It uses IPPP structure and CAVLC. In [4], Baseline profile was chosen with respect to limited computational power of current mobile devices. Our results showed that Multiview Video Coding [5] and Video+Depth [6] provide the best perceptual overall quality under the given constraints. Although there were differences between these two methods content-by-content wise, they performed similar in overall. Both methods outperformed Simulcast [14] and Mixed Resolution approach [12].

But as currently computational power of mobile devices is increasing, prospective mobile devices will be able to process more complex coding structures. In this paper, we extend the existing results by applying High profile as a more complex codec structure in a large-scale quality evaluation. In contrast to Baseline profile, High profile uses hierarchical-B pictures and CABAC. The paper aims at extending the knowledge about an optimum coding method for mobile 3D television and video towards prospective possibilities in terms of computational power.

The paper is organized as follows. In section 2, we present the research method that we chose for our study. Section 3 includes the results of the High profile evaluation as well as a comparison of Baseline and High profile results. Section 4 discusses the results and concludes the paper.

II. RESEARCH METHOD

A. *Participants*

87 participants (age: 16-37, mean: 24) equally stratified in gender took part in this study. We collected a parents' consent for all underaged participants before the study. All participants were recruited according to the user requirements for mobile 3D television and system and classified as naïve participants. They were screened for normal or corrected to normal visual acuity (myopia and hyperopia, Snellen index: 20/30), color vision using Ishihara test, and stereo vision using Randot Stereo Test ($\leqslant$ 60 arcsec).

B. *Stimuli*

1) *Variables and their production*

In our study, we evaluated four different coding methods. Each coding method was optimized for two quality levels. Furthermore, two codec profiles were used.

Coding Approaches The left and right view are coded as independent streams using H.264/MPEG-4 AVC. This method does not need any pre- or post processing before coding and after decoding, the complexity on sender and receiver side is low. Redundancy between channels is not exploited. Optimization is carried out by jointly varying the quantization parameter (QP) for left and right view.

H.264/AVC MVC allows inter-view prediction. The left view is used as reference for the right view. Prediction has been enabled for anchor as well as for non-anchor frames. No pre- or post-processing is required on the sender or receiver side. Optimization is carried out by jointly varying the QP for left and right view.

Binocular suppression theory states that perceived image quality is dominated by the view with higher spatial resolution [1]. The mixed resolution approach utilizes this attribute of human perception. One view is decimated before transmission and up-scaled at the receiver side. For the coding experiments the right view was decimated by a factor of two in horizontal and vertical direction. The simple MRSC approach without interview prediction, optimized down-sampling and unsharp masking has been used. Optimization is carried out by independently varying the QP for left and right.

MPEG-C Part 3 defines a video plus depth representation of the stereo video content. One view and the associated depth signal are coded. At the receiver the second view is synthesized by depth image based rendering [2]. In most cases the depth signal can be coded at a fraction of the color bit rate. Nevertheless errors in depth estimation and interpolation at occurring disocclusions introduce artefacts to the rendered view. Optimization is carried out by independently varying the QP for video and depth.

Quality levels The coding approaches have been evaluated at a high and low quality. Note, that it is not useful to define a constant high and constant low bit rate for all sequences to achieve high and low qualities for all sequences. Reason for this is a variable compressibility of different sequences. A rate sufficient for a high quality for one sequence might produce a low quality for other sequences. To guarantee a comparable low and a comparable high quality for all sequences a low and a high rate point had to be determined for each sequence individually. The following approach was used to obtain these rate points:

To define a high and a low quality for all sequences of the coding test set the quantization parameters (QP) of the codec for simulcast coding was set to 30 for the high quality and 37 for the low quality. This results in a low and high bit rate for each sequence of the coding test set. Resulting bit rates are shown in TABLE I. and have been used as target rates for the other three approaches together with the baseline profile.

TABLE I. TARGET BIT RATES IN KBIT/S FOR HIGH AND LOW QUALITY

Profile	Quality	Bullinger	Butterfly	Car	Horse	Mountain	Soccer2
Base-line	Low	74	143	130	160	104	159
	High	160	318	378	450	367	452
High	Low	46	94	112	104	78	134
	High	99	212	323	284	208	381

Bit rates for the high profile are also shown in TABLE I. . They are the rates from the sequences coded with high profile and simulcast having the same PSNR as the sequences coded with the base profile and simulcast at QP 37 and QP 30. This guarantees a comparable objective quality for the baseline and high-profile sequences using simulcast. Hence it can be subjectively evaluated if the different GOP structures of the two profiles have an influence on the subjective quality which is not reflected by the PSNR.

Codec Profiles Coding has been carried out using two codec profiles. The simple baseline profile uses an IPPP structure and CAVLC. The complex high profile enables hierarchical B-Frames and CABAC. The Intra Period was set to 16. For the Simulcast, Mixed Resolution and V+D approach the AVC Reference Software JM 14.2 has been used. The MVC stimuli have been coded using the MVC reference Software JMVC 5.0.5.

2) Contents

Six different contents were selected. All contents matched the user requirements for mobile 3D television and video [ref]. The contents provided different spatial details, temporal resolution, and amount of depth (TABLE II.).

TABLE II. CHOSEN CONTENTS AND THEIR CHARACTERISTICS

Screenshot	Genre and their audiovisual characteristics
	**Videoconference – Bullinger** Spatial details: medium Temporal details: low Depth: medium Length: 7.7 sec Size in pixels: 432 x 240
	**Animation – Butterfly** Spatial details: high Temporal details: medium Depth: medium Length: 12 sec Size in pixels: 432 x 240
	**Action/Movie – Car** Spatial details: high Temporal details: high Depth: medium Length: 7.8 sec Size in pixels: 432 x 240
	**Nature/Documentary – Horse** Spatial details: high Temporal details: low Depth: high Length: 9.3 sec Size in pixels: 432 x 240
	**Nature/Documentary – Mountain** Spatial details: high Temporal details: low Depth: high Length: 8 sec Size in pixels: 320 x 240
	**Sports – Soccer2** Spatial details: medium Temporal details: high Depth: high Length: 13.3 sec Size in pixels: 320 x 240

C. Stimulus presentation

The tests were conducted in controlled lab conditions [10]. An autostereoscopic 3.5” display with a resolution of 428px x 240px was used to present the videos. The display used HDDP technology [7]. That enables the mobile 3DTV display to provide equal resolution for monoscopic and autostereoscopic presentation. The display was connected to a Dell XPS 1330 laptop via DVI. The laptop served as a playback device. Items were presented in randomized orders. Each item was shown twice during the psychoperceptual evaluation.

D. Test procedure

A factorial, mixed design was used in this experiment. Within subject variables were content, coding method, and quality level. Codec profiles were used as the between subject variable (47 participants for Baseline and 40 participants for High profile). The test procedure was the same as in the first study [4]. Before the test, we collected demographic and psychographic data of the test participants. Following, an accommodation task was conducted in which test participants were introduced to use the autostereoscopic display.

In the next step, the training and anchoring, test participants watched a subset of test items. This subset contained the whole range of constructed quality. Items were chosen from Baseline and High profile. In the training, test participants practiced the evaluation task. We chose Absolute Category Rating [9] as evaluation method. In Absolute Category Rating, each test item is presented and rated independently one after another. The test participants rated acceptance of overall quality on a binary (yes/no) scale and satisfaction with overall quality on an 11-point unlabeled scale. The training was followed by the evaluation task. Test stimuli were rated twice in randomized order. Each session took about 90 minutes.

E. Method of Analysis

As no normal distribution was given for the ratings (Kolmogorov-Smirnov: $p<.05$), non-parametric test were applied. For the within-subject variables, Friedman test was used to measure effect of the independent variables on the ordinal dependent variables. Wilcoxon test was then applied to measure differences in pairwise comparison. To analyze the between subject variables, Mann-Whitney U test was used to measure differences between two unrelated, ordinal samples.

III. RESULTS

A. Results for High profile

Overall quality acceptance for high quality level was at least 75%. For low quality level, MVC and V+D reach an acceptance level of 55% and more. The analysis of the overall quality satisfaction scores shows that Multiview Video Coding (MVC) and Video+Depth (V+D) outperform Simulcast and MRSC. As in [4], MVC and V+D perform similar for High codec profile. Content was a determining factor in comparing the overall quality that is provided by different coding methods. Following, we present a detailed analysis of the results.

1) High quality level

The independent variables had significant effect on the test participants’ ratings for high quality level ($F_R = 192$, df=3, $p<.001$). The findings for high quality level (Figure 2) show that MVC and V+D outperform the other two coding methods averaged over content (all comparisons vs. MVC and V+D: $p<.001$). MVC and V+D get comparable mean satisfaction scores ($Z = -0.29$, $p>.05$, ns). MRSC gets the worst results (MRSC vs. Simulcast: $Z = -7.488$, $p>.001$). A content-by-content analysis shows that for each content coding method has significant impact on the dependent variable (Friedman: all comparisons $p<0.01$). For contents Car and Mountain, Video+Depth significantly gets the best mean satisfaction scores (all comparisons: $P <.05$). MVC performs significantly as the best coding method for contents Butterfly and Horse (all comparisons: $p<.01$). Except for Mountain and Bullinger, Simulcast always provides better satisfaction scores than MRSC (all comparisons: $P <.01$).

2) Low quality level

Again, coding methods had significant effect on the dependent variable ($F_R =422.46$, df=3, $p<.001$). At low quality level, V+D gets the highest mean satisfaction scores ($Z=-8.349$, $p<.001$). MRSC and Simulcast are the worst ($Z=.445$, $p>.05$, ns). Again a strong content dependency can be found. However, only for content Butterfly MVC performs better than V+D ($Z= -3.193$, $p=.001$). For all other contents, V+D gets significantly the best satisfaction score (all comparisons: $p<.05$). Again, MRSC gets the worst mean satisfaction scores (all comparisons: $p<.05$), except for content Bullinger.

B. Comparison of Baseline and High profile

In overall, Baseline profile gets slightly better results at high quality level ($Z=-2.271$, $p<.05$) while no difference was identified for low quality ($Z=-0.123$, $p>.05$, ns). With regard to the selection of a coding method, a method-by-method analysis is interesting. At high quality level, Baseline and High profile get similar satisfaction scores for every method (all comparisons: $p>.05$) except MRSC ($Z=-2.559$, $p=.01$). For low quality level, Baseline profile is rated better for MVC ($Z=-2.047$, $p<.05$). High profile gets better scores at Video+Depth ($Z=-2.809$, $p<.01$). No difference was found for MRSC and Simulcast coding (all comparisons: $p>.05$).

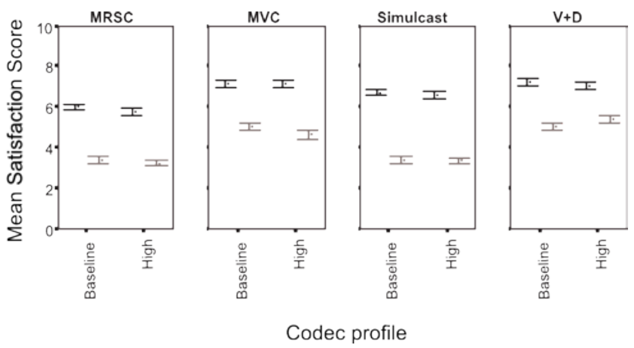

Figure 1. Comparison of Baseline and High profile results for each coding method. Results are shown for high quality level (black) and low quality level (gray). Error bars show 95% confidence interval.

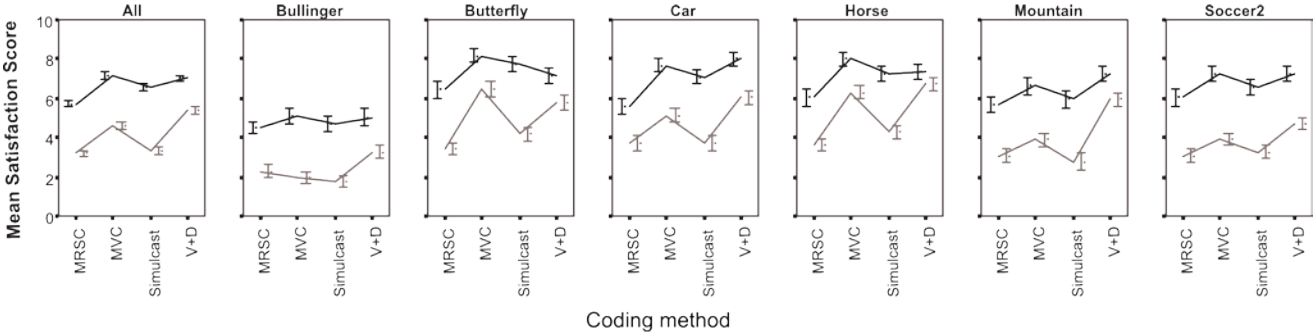

Figure 2. Mean satisfaction scores for High codec profile. Results are . Results are shown averaged over all contents (All) as well as content-by-content for high quality level (black) and low quality level (gray). Error bars show 95% confidence interval.

I. Discussion and Conclusion

The goal of this paper was to compare two codec structures for mobile3D television and video. The study extended our first study [4] to find an optimum coding method for mobile3DTV with the application of hierarchical-B pictures and CABAC.

The results presented in this paper confirm the findings of the Baseline profile evaluation [4]. MVC and Video+Depth also get the best satisfaction scores at High codec profile for both quality levels. Although there are differences for MVC and Video+Depth in the content-by-content analysis, the overall results are comparable. As with Baseline profile, they significantly outperform Simulcast and MRSC as coding methods in terms of satisfaction scores.

The comparison of Baseline and High codec profile shows that the satisfaction scores are comparable. Indeed, analysis shows small differences for Baseline and High codec profiles for some settings. An overall view of coding methods and quality levels shows no differences among the two profiles. However, it is remarkable that High profile provides the same experienced quality using lower bit rates (TABLE I.). Lower bit rates of encoded methods can bring advantages for the transmission of these sequences in terms of better error resilience [15].

Concluding, the results of the paper show that MVC and Video+Depth are currently the two preferable coding methods for mobile 3D television and video. The selection of coding methods was limited due to the calculation power of mobile devices. Further work will need to take into account more sophisticated coding approaches [13]. However, for a final selection of an optimum coding method, transmission and the related channel characteristics [15] need to be taken into account due to different impact of coding methods on error resilience methods.

References

[1] Meesters, L., IJsselsteijn, W., and Seuntiens, P., \A survey of perceptual evaluations and requirements of three-dimensional tv," Circuits and Systems for Video Technology, IEEE Transactions on 14, 381-391 (March 2004)

[2] Tam, W.J. "Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV", JVT-W094, San Jose, CA, April 2007.

[3] Merkle, P., Wang, Y., Müller, K., Smolic, A., and Wiegand, T. "Video plus Depth Compression for Mobile 3D Services", Proc. IEEE 3DTV Conference, Potsdam, Germany, Mai 2009.

[4] Jumisko-Pyykkö, S., Weitzel, M., and Strohmeier, D. "Designing for user experience: what to expect from mobile 3d tv and video?," in UXTV '08: Proceedings of the 1st international conference on Designing interactive user experiences for TV and video, 183-192, ACM, New York, NY, USA (2008).

[5] Strohmeier, D., Tech, G. "Sharp, bright, three-dimensional" - Open Profiling of Quality for mobile 3DTV coding methods, Proceedings of 'Multimedia on Mobile Devices', a part of the Electronic Imaging Symposium 2010, San Jose, California, USA, January 2010.

[6] ISO/IEC JTC1/SC29/WG11, Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding. Doc. N9978, Hannover, Germany (July 2008).

[7] ISO/IEC JTC1/SC29/WG11, ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information. Doc. N8259, Klagenfurt, Austria (July 2007).

[8] Uehara, S., Hiroya, T., Kusanagi, H., Shigemura, K., and Asada, H. "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," in *Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX, Vol. 6803*, San Jose, USA, Jan. 2008.

[9] Recommendation ITU-T P.910, Subjective video quality assessment methods for multimedia applications, Recommendation ITU-T P.910. ITU Telecom. Standardization Sector of ITU (September 1999).

[10] Recommendation ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of TelevisionPictures, Recommendation ITU-R BT.500-11. ITU Telecom. Standardization Sector of ITU (2002).

[11] Gotchev, A., Smolic, A., Jumisko-Pyykk□o, S., Strohmeier, D., Akar, G. B., Merkle, P., and Daskalov, N., "Mobile 3d television: development of core technological elements and user-centered evaluation methods toward an optimized system," Multimedia on Mobile Devices 2009 7256(1), 72560J (2009).

[12] Brust, H., Smolic, A., M□uller, K., Tech, G., and Wiegand, T., "Mixed resolution coding of stereoscopic video for mobile devices," 3DTV Conference (2009).

[13] Fehn,C., Kauff, P., Cho, S., Kwon, H., Hur, N., and Kim, J., "Asymmetric coding of stereoscopic video for transmission over t-dmb," 3DTV Conference (May 2007).

[14] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Advanced Video Coding for Generic Audiovisual Services (November 2007).

[15] Bugdayci, D., Bici, M.O., Aksay, A., Demirtas, M., Akar, G.B., Tikanmäki, A., Gotchev, A., "Stereo DVB-H broadcasting system with error resilient tools", Technical report D3.4, December 2009