SOURCE AND CHANNEL CODING RECIPES FOR MOBILE 3D TELEVISION

Atanas Gotchev[1], *Dominik Strohmeier*[4], *Karsten Mueller*[3], *Gozde Bozdagi Akar*[4], *Venceslav Petrov*[5]

[1]Tampere University of Technology, Finland; [2]Ilmenau University of Technology, Germany;
[3]Fraunhofer HHI, Germany; [4]Middle East Technical University, Turkey; [5]MM Solutions AD, Bulgaria

ABSTRACT

The paper overviews the transmission chain of 3D television for mobile services with special emphasis on stereoscopic video transmission. Specifically, methods for source and channel coding are overviewed for their suitability for such applications. They are analyzed in terms of performance and complexity as well as in terms of user acceptance and satisfaction. Furthermore, viewing conditions involving mobile auto-stereoscopic displays are analyzed emphasizing the limitations of those displays, such as accommodation-convergence rivalry and inter-view cross-talk along with corrective measures such as anti-aliasing filtering and comfort zone maintenance. Extensive subjective tests studying the importance of visual 3D cues on portable displays and visibility of artifacts are presented. Thus, the visualization conditions are put against the processing methods in order to prescribe practical optimal selection of source-channel coding methods.

Index Terms**—** Multi-view coding, forward error correction, DVB-H, auto-stereoscopic displays, Quality of Experience

1. INTRODUCTION

Technologies for creation, processing and visualization of 3D visual content, denoted as 3D Media technologies have become objective for an intensive research. These technologies have gathered strong attention due to the current success of 3D digital cinema and have become a focus of consumer electronics developments and a spectacular topic at commercial shows, such as IBC and CES. 3D standardization initiatives have emerged in bodies such as MPEG [1], SMPTE [2], ITU [3], and DVB [4].

Mobile3DTV, a three-year research project funded by European Commission, has addressed this trend from the point of view of the most dynamic that is, the *mobile*, user. The project has been targeting the development of core technological components of a mobile 3D television system over DVB-H channel. Among other objectives, the project has aimed at developing advanced methods for stereo-video content creation and coding and its error-resilient transmission over the wireless channel. Extensive objective and subjective tests have been conducted in order to identify the optimal source and channel coding methods in terms of compression efficiency, reduced complexity and perceived quality. This paper overviews the results of the experiments in a form of practical recipes for 3DTV delivery to handheld devices ensuring the added value of depth presence and quality appealing to the mobile user.

2. MOBILE 3DTV SYSTEM

A typical mobile 3DTV system is illustrated in Fig. 1.

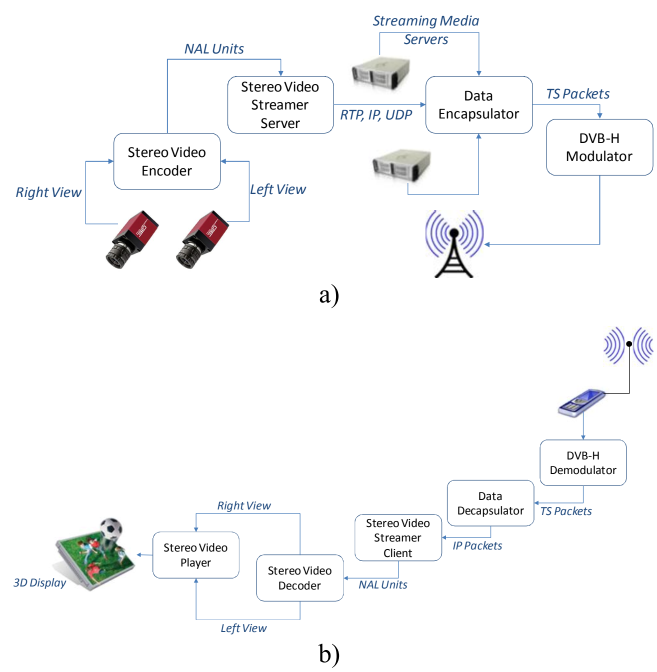


Fig. 1. Delivery of stereo video over DVB-H channel. a) Transmitter side; b) Receiver side.

It starts with stereo video content, which is repurposed and converted to the adopted representation format, suitable for effective encoding and transmission. The content is then compressed to obtain Network Abstraction Layer Units (NALUs) to be fed to the stereo video streamer. A packetizer encapsulates the NALUs into Real Time Transport Protocol (RTP) mono-compatible, User Datagram Protocol (UDP) and finally Internet Protocol (IP) datagrams for each view separately. The resulting IP datagrams are encapsulated in the DVB-H link layer where the Multi

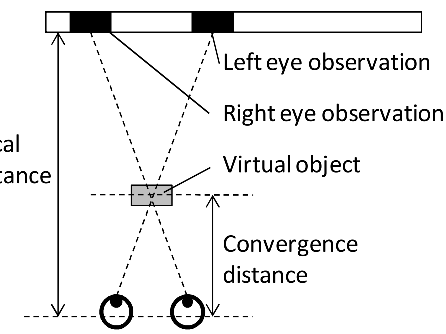
Protocol Encapsulation Forward Error Correction (MPE-FEC) and time slicing take place [7]. Through the MPE-FEC mechanism, IP datagrams are protected by adding additional bytes for a variable-length Reed-Salomon (RS) coding. Time slicing allows sending the packets into time slices (bursts) for efficient power consumption at the receiver site. The link layer outputs MPEG-2 Transport Stream (TS) packets, which modulate a radio-frequency signal at the physical layer of a DVB-T-compatible modulator. At the receiver side (Fig. 1b), the reverse operations take place. After demodulation, the received stream is decoded using the section erasure method, i.e. the MPE-FEC frame is filled with contents of the error-free MPE and MPE-FEC sections and the empty bytes in the frame are marked as erasures. RS decoding is performed to reconstruct the lost data, and finally, the received and correctly reconstructed IP datagrams are passed to the video decoder which generates stereo views to be played on a stereoscopic display.

3. VIEWING CONDITIONS

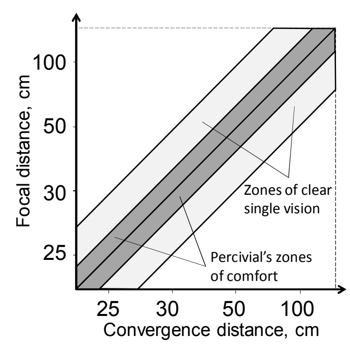
3.1. Portable auto-stereoscopic displays

The 3D portable display plays a decisive role in the acceptance of the mobile 3DTV technology. As in most stereo 3D systems, the 3D visual effect is achieved by showing two slightly different perspective images to the two eyes, which create binocular illusion of depth. However, wearing glasses for alternating proper views to the eyes is considered impractical for the case of mobile devices. Therefore, the use scenario for such devices considers the use of *auto-stereoscopic* displays. They work by beaming different image towards each eye. Additional optical filter mounted on top of a TFT-LCD screen makes the visibility of each TFT element a function of the observation angle. The optical filter can be either a *lenticular sheet*, i.e. an array of microlenses which redirects the light or *parallax barrier*, i.e. a mask, which partially blocks the light traveling in certain directions [1], [6].

Mobile autostereosopic displays usually support two (left and right) views only, as a tradeoff between the number of views and the spatial resolution of one view. This defines the use mode of 3D-enabled mobile devices as *personal*, rather than *multi-user*. The left and right views exhibit differences or *disparities* in the horizontal positions of objects in each view, which are responsible for the binocular illusion of depth [8]]. The illusory distance between the object and the display is referred to as *apparent depth*. What differentiates the mobile auto-stereoscopic displays from their bigger-size counterparts is the amount of minimum and maximum depth which is possible to achieve on such a limited-size display – the so-called display *comfort zone* [10]. It is driven by the effect of *accommodation-convergence* (A/C) *rivalry*. *Convergence* refers to the mechanism, in which both eyes perform inward or outward motion in order to bring the projection of the intended object to the foveae of both retinas. *Accommodation* refers to the ability of each eye to change its focal power, so the projection of the object is focused on the retina. On a stereoscopic display the convergence and focal distances to an object differ, as the distance to the converging point is influenced by object disparity, and the focal distance equals the viewing distance (see Fig. 2a). This difference creates the A/C rivalry. The zones of negligible A/C rivalry, referred to as Parceval's zones of comfort [11] are usually determined as being narrower and inside the zones of clear single vision [12]. They are plotted in graphs of "focal distance" versus "convergence distance" and can be used to suggest a comfort zone of a display with particular size and resolution. As seen in Fig. 2b, a smaller focal distance creates a more pronounced A/C rivalry which suggests that the range of "comfortable" disparities is more limited for handheld 3D displays than for displays allowing greater viewing distance.



a)



b)

Fig. 2. Accommodation-convergence rivalry: a) focal and apparent distance to of an object, b) zones of clear single vision and Persival's zones of comfort [13].

The angular dependence of the views and their discrete template as groups of (sub) pixels creates an inevitable effect of inter-view *cross-talk*, when part of the light intended for one eye is also visible by the other. It has been suggested that crosstalk of 25% is the maximum acceptable in stereoscopic imaging [9]. Correspondingly, the position, from which a view is seen over the whole surface of the display with cross-talk below 25% is defined as a *sweet spot* for that view. Optimal viewing conditions are achieved at the *optimal viewing distance* where the elements of the display are seen with a maximal brightness, and the visual separation between the views exhibits minimal cross-talk [13].

The views of most of mobile autostereoscopic displays are arranged column-wise and only half of the TFT elements are visible from an observation position. This is equivalent to horizontal down-sampling of the image by a factor of 2. Stereoscopic images are usually prepared for displays with square pixel aspect ratio and need to be suitably pre-filtered before visualized on a 3D display to avoid aliasing caused by the down-sampling at the optical layer. While interleaving the two views on the surface of the TFT-LCD, the step size used to interleave the TFT elements is a potential source of color bleeding artifacts as the color of the visible image is a function of the observation angle. For some auto-stereoscopic displays the amount of visible and partially occluded sub-pixels of certain color might prevail, which introduces color tint for some observation angles. Such displays have reduced sweet spot width, since the optimal observation angle of each color channel is slightly different, and the zone which is optimal for all three colors is narrower. It is possible to mitigate the effect of color bleeding by a non-ordinary yet quite effective rearrangement of pixels. An example of such arrangement is the 3D display with horizontal double-density pixel (HDDP) arrangement, where the color of the sub-pixels change along columns, but the view assignment of sub-pixels change along rows [15]. Displays with such topology do not exhibit color bleeding between the views. Additionally, due to the double pixel density in horizontal direction each view has square pixel aspect ratio which eliminates the most common reason for aliasing.

3.2. Importance of 3D visual cues and visibility of artifacts on portable auto-stereoscopic display

On cognitive level, the human-visual system (HVS) processes the 3D information as a set of 3D cues. Those are either binocular, i.e. perceived by the two eyes or monocular, i.e. perceived by a single eye. The binocular depth cues are a consequence of both eyes observing the scene at slightly different angles. In addition to the *convergence*, which aims at minimizing the visual information difference as projected in the two retinas, a stronger mechanism is *stereopsis*, which aims at using the disparity between the two eye's views for perception of depth relative to the point of convergence [8]. This point is strongly related with the *accommodation* – the ability of the eye to change the optical power of its lens in order to focus on the object of interest. Binocular depth cues are believed to be the most important for short distances and when combined with *motion parallax*, i.e. the relative motion of closer and further images on the retina, when the observer moves his/her head or eyeballs. For longer distances, the binocular cues are supported and then replaced by monocular cues such as shadows, perspective lines, texture scaling, and occlusions. It is important to characterize the effect of the 3D visual cues for different viewing conditions and on portable auto-stereoscopic displays and to access the visibility of artifacts caused by the processing stages of the mobile 3DTV system, such as compression and transmission over wireless channels.

The importance of 3D visual cues on portable auto-stereoscopic display has been addressed in a recent study [16]. Subjective tests have assessed the performance of the subjective depth estimation and evaluation of the subjective quality. The test stimuli included synthetic 3D scenes with controlled appearance of six different depth cues as follows: focal blur, shadows, texture, binocular views fitting entirely to the comfort zone of the auto-stereoscopic display used, binocular views optimal for the comfort zone of a large HD display rescaled to mobile resolution, and a combination of the first four aforementioned cues. These cues were used in images with three different JPEG compression levels: best quality with no compression, good quality with some slightly visible compression artefacts (90% JPEG compression), and low quality with noticeable blockiness (50% JPEG compression). Fig. 3 presents a stereopair of the test content.

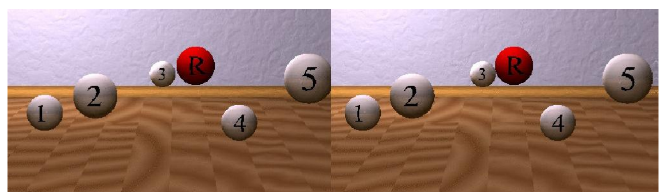
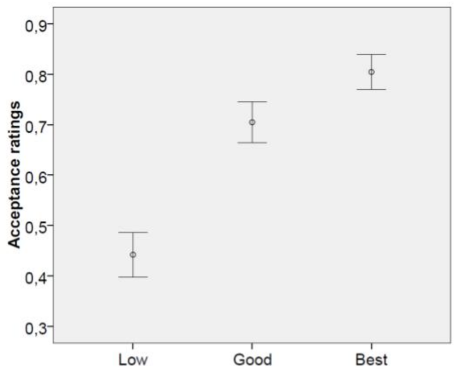
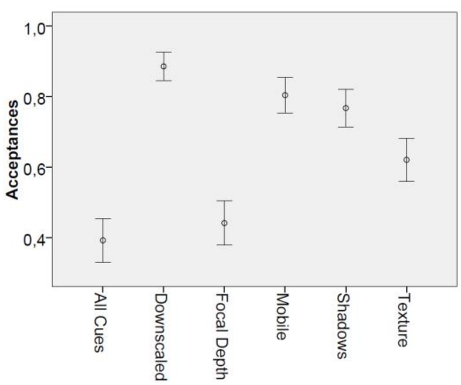


Fig. 3. Test stereo image. Disparity range spans the comfort zone of a portable display. 90% JPEG compression. Ball #4 is at the same distance from the viewer as the reference ball marked 'R'.

The device used in the tests was a 3.1" NEC HDDP auto-stereoscopic display [15]. The test participants were asked to determine which ball is at the same distance as the reference one. The correctness and the speed of answers were measured. Then, the participants evaluated also the overall quality of the scenes. The results revealed that in terms of time for depth estimation, compression played no significant role, while the determining 3D cue was the binocular one. In terms of correctness of depth estimation, there was no significant difference between best quality (100% JPEG) and good quality (90% JPEG). However, the correctness dropped significantly in the case of 50% JPEG compression quality. As of the influence of different depth cues, the binocular one drove the highest amount of correct estimates. The overall quality evaluation revealed the strong influence of the compression artifacts. The low quality level (50% JPEG) was rated below 50% of acceptance and there was a significant difference between the overall quality of best and good compression (Fig. 4a). Surprisingly enough, the HD down-sampled content received the highest rank in quality evaluation followed by the binocular cue specifically adapted to the disparity range of the used display (Fig 4b). In overall, the tests in the referred study [16] confirmed that binocular disparity is the strongest depth cue, also on mobile auto-stereoscopic displays. Images with stereoscopic depth cues got significantly better ratings than the images with monocular cues when measuring the accuracy and speed of depth perception. Furthermore, it was found that the use of focal depth as depth cue lead to low rankings in all the tested areas: in accuracy, efficiency and in acceptance. It seems that depth of field does not work well as a separate depth cue on a mobile display observed from close distance. The present blur is perceived mostly as a “2D” artefact, and affects the cyclopean image of a 3D scene, rather than its binocular disparity.



a)



b)

Fig. 4. Acceptance ratings. a) Three different compression ratios: Low quality (50% JPEG compression), Good quality (90% JPEG compression), Best quality images (no compression); b) Six different depth cues.

The tests also emphasized the effect of compression level on both the accuracy and the speed of depth perception. The depth and the correctness of the estimation had a negative correlation, indicating that if the depth estimation was conducted fast, the result was also most likely correct. The quality evaluation was in accordance with a previous study by Strohmeier *et al.* [14]. It indicated that the depth effect has an added positive value to the perception of quality only in cases of negligible compression artefacts.

3.3. Spatial resolution and camera baseline

The issue of spatial resolution and the viewing mode (2D versus 3D) has been addressed in an extensive user-centered quality study [17]. The experiments in the study involved 30 naïve participants to test how quality is experienced in controlled laboratory and home contexts of use and to investigate the effect of coding parameters on experienced overall quality. The test material included four stereo videos of about 35 sec lengths spanning the genres of documentary, animation, series and user-created content. The videos were compressed as simulcast by H.264/AVC baseline with bitrates of 160, 320 and 768 kbps and framerates of 10 and 15 fps. In addition, extremely high bit rate of 1536 bps with 24 fps, was used as a hidden anchor along with the videos encoded in 2D mode (160 and 320 bps and 15 fps) as another hidden anchor. The tests were carried out on a mobile3DTV prototype device featuring a 4.3’’ transmissive autostereoscopic LCD display with 800x480 pixels native resolution and based on parallax barrier with two-view interleaving on pixel level. The parallax barrier technology allowed for changing between 2D and 3D mode by switching off and on the barrier. This however caused a change in the spatial resolution being twice higher in 2D mode than in 3D mode. The results revealed clear superiority of 2D mode versus 3D where only the two 2D coding modes went over the acceptance thresholds both for the lab and home contexts. In complementary qualitative tests, 2D mode was favorably described by positive terms (clear details, ease to view, error-free) while the 3D mode was characterized by mostly negative terms (double edges, color errors, lack of clarity, difficulty of viewing and change of viewing angle, eye strain) [17]. Thus, the study revealed the crucial importance of the display technology in general and the spatial resolution in particular to be used in mobile 3D video applications. It confirmed the impact of color artifacts generated by the parallax barrier at pixel-level interleaving and the importance of spatial resolution for high quality video perception on mobile displays.

In order to validate the findings about the display technology, spatial resolution and effects of depth presence, another study was run with the aim to examine the influence of depth and compression artifacts on quality of experience of 3D video for mobile devices [18]. The tests involved 30 naïve participants who evaluated content with varying stereo camera baselines and compression quality displayed on the lenticular-based autostereoscopic HDDP LCD display by NEC [15]. Three depth levels, i.e. mono, short and wide baseline, and five quantization parameters (QP=25, 30, 35, 40, 45) for different qualities within the video compression process were set. The selection of these parameters was driven by the aim to systematically study the (likely) positive influence of depth versus negative influence of artifacts and in accordance with the results obtained in [14]. There, the higher quality of stereoscopic compared to monoscopic video on small screen was confirmed for the same display technology and for the case of no compression artifacts [14]. In the setting, the short baseline corresponds to a scenario where a higher definition stereo video is downsized to mobile resolution, resulting into somehow shallower depth. The wide baseline corresponds to a scenario, where the depth budget is fully utilized to fit into the (theoretical) comfort zone of the portable stereoscopic displays. The results of the study [18] demonstrated that compression artifacts are a dominant factor primarily determining the quality of experience compared to varying depth range. More specifically, content with strong compression has been found unacceptable by the viewers. Only content compressed with QP≤30 has been found acceptable. For the cases of low compression artifacts, the 3D content provided higher quality of experience versus monoscopic content, while the perceived differences between short and wise baseline have been categorized as insignificant, thus making an interesting point about the way how a stereo video content has to be created/repurposed for mobile displays.

3.4. Viewing conditions recipes

The display has a determinant role in the visual experience of 3D video for mobiles. There is a noticeable difference between the theoretical and the true comfort zone, the latter being most often narrower than the expected. This is mainly due to the combined effect of cross-talk and aliasing artifacts appearing on such displays. Their degrading effect has been confirmed by subjective tests, where down-scaled (repurposed) content has been better accepted than a '*mobile display-optimal*' content. This result is instructive for the need of proper simultaneous anti-aliasing and cross-talk mitigating filtering at the receiver side for achieving better 3D effect. In selecting mobile auto-stereoscopic display technology, special care has to be taken for the acceptable compromise between spatial resolution and 3D effect. Presence of color bleeding artifacts should be avoided.

Apart from display-specific artifacts, 3D visual experience is mostly influenced by the compression artifacts. Low bit-rate 3D video affects the efficiency of depth perception, as the binocular HVS has no enough time to create a plausible 3D effect for moving objects and the benefit gained by added depth is invalidated by the artefacts. This indicates that 3D perception is an extremely delicate mechanism and vulnerable to artefacts. Furthermore, there is an influence of 2D artefacts over the 3D perception. A scene with a good "3D quality", (i.e. which allows fast depth judgement), might be graded as having low overall quality due to 2D artefacts. Another interesting outcome is that scenes with the same amount of blockiness (same compression rate for each channel) but with different depth levels (i.e. 2D, "downscaled" 3D and "display-optimal" 3D) are graded as having different quality. This effect implies that a different amount of 3D depth cues (binocular disparity) influences the perceived quality.

4. SOURCE CODING

The formats considered for stereo video representation in mobile 3DTV applications include Conventional Stereo Video (CSV), Mixed Resolution Stereo (MRS) and Video plus Depth (V+D), as exemplified in Fig. 5. CSV consists of a pair of sequences, corresponding to left and right eye views. Ideally, this pair is already properly resized at the transmitter side and no prior size-conversion processing is required for the display. Through lowering the spatial resolution of one of the views, MRS aims at reducing the coding bitrate while expecting almost the same subjective quality in comparison to the full resolution case. This is based on the binocular suppression theory, where the human brain is able to mask some blurry artifacts in one view with the other [19]. Up-sampling to full resolution prior to display is required. V+D consist of a conventional monoscopic color video and an associated per-pixel depth map sequence. A reduced bitrate compared with CSV is achievable due to the monochromatic and piecewise smooth characteristics and therefore higher compression capability of depth data. The format offers scalable-baseline view synthesis at the price of additional rendering at the receiver side and comprehensive high-quality depth estimation at the transmitter side.

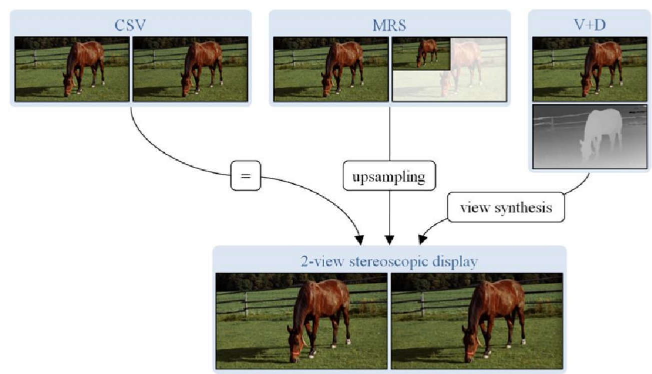


Fig. 5. Candidate stereo video representation formats for mobile 3DTV.

4.1. Objective comparison of coding methods

The corresponding coding approaches are mainly based on the H.264 AVC compression standard [20] and include AVC Simulcast, AVC with Stereo SEI-Message, AVC Auxiliary Picture Syntax, and MVC [22], denoted as *video+video* (V+V) approaches, and MPEG-C part 3 with AVC [21], denoted as an *video+depth* (V+D) approach. All candidate approaches have been optimized and objectively compared for their applicability in mobile 3DTV [23]. The V+V approaches have been compared under the conditions of mobile spatial resolution and different motion/disparity prediction structures. The V+D approach has been studied for the optimal allocation of bit budget between the color view and monochromatic depth. The conclusions from the coding experiments are the following: For V+V representations, coding methods which explore the interview correlation (i.e. Stereo SEI message and MVC) are superior to simulcast with higher gain achieved by efficient motion prediction structures (e.g. through hierarchical B-pictures). Fig. 6 illustrates a typical rate-distortion performance, in terms of averaged PSNR over the two luminance channels, of CSV coding methods for the *Horse* test sequence [23].

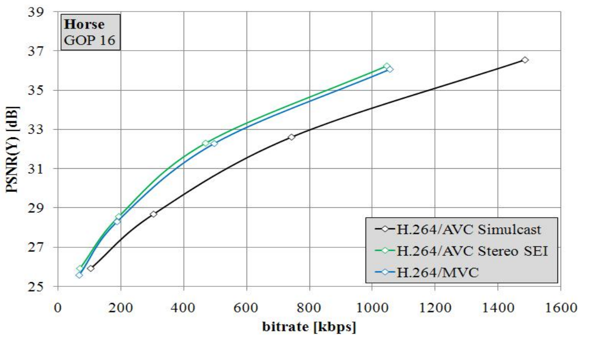


Fig. 6. Rate distortion curves for V+V encoding methods: H.264/AVC Simulcast, H.264/AVC Stereo SEI and H.264/MVC.

For the V+D coding, the optimal bit allocations have been found highly dependent on the content. Fig. 7 shows the coding behavior of AVC applied both to the monoscopic view and its associated depth sequence for the *Horse* test sequence, where each point denotes a pair of view/depth QPs. The envelop curve determines the optimal combinations of coding parameters for the two modes.

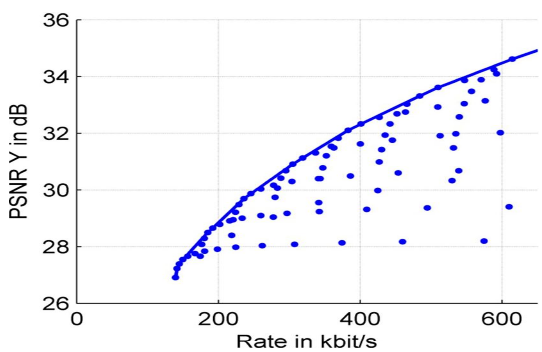


Fig. 7. Rate distortion points for different combinations of QPs for V+D. The envelop marks the optimal combination.

4.2. Subjective tests

The results from the objective tests were used to create a short list of formats, including MVC, V+D and MRS, to be compared against simulcast in a series of extensive subjective tests. While giving similar results, MVC was preferred against Stereo SEI message for its backward compatibility. The large-scale subjective tests included six video contents encoded in low and high quality levels, where the bitrates were anchored to the bitrates achieved by simulcast at QP=30 (high quality) and QP=37 (low quality) (Table 1). Two encoding profiles were used: base profile with simplified prediction structure of IPPP and CAVLC for entropy coding, and high profile with hierarchical B-frame prediction and CABAC for entropy coding [24]. Thus, the final test setting included 96 test sequences, tested by 87 naïve participants on the NEC autostereoscopic 3.1 inch display with resolution 428x240 pixels.

TABLE 1. TARGET BITRATES FOR TEST CONTENTS, ENCODING PROFILES AND ENCODING QUALITY

Profile	Quality	Bullinger	Butterfly	Car	Horse	Mountain	Soccer2
Baseline	Low	74	143	130	160	104	100
	High	160	318	378	450	367	452
High	low	46	94	112	104	78	134
	High	99	212	323	284	208	381

The results revealed the superiority of MVC and V+D versus simulcast and MRS, although content-dependent (e.g. V+D superior for *Car*, *Mountain*, and *Soccer2*; MVC - for *Butterfly*) [25]. The results showed equal subjective quality between baseline and high encoding profiles with, significantly lower bit rates for the high profile. The latter exploits more efficient, though more complex, coding structures [26].

4.3 Source coding recipes

Mobile video services impose bandwidth and memory limitations. Therefore, very efficient compression of stereo video is required to realize 3D instead of conventional 2D video. We favor the use of MVC with a simplified (IPPP) prediction structure as the best compromise between coding performance, user acceptance and computational resources at the terminal device. MVC has shown either highest or close to highest user acceptance and quality satisfaction while yielding additional coding gain compared to Simulcast. Furthermore, it is backward compatible to 2D video coding and requires no further processing at the receiver size. If the resources allow, one can opt for the high coding profile, which offers the same perceived subjective quality at lower bitrates, thus leaving some room for bit budget-rich error-resilient channel coding.

5. CHANNEL CODING

The channel coding has specifically addressed the transport of 3D video over DVB-H channel, including the candidate coding methods, error control and error correction schemes. Both objective comparisons and large-scale subjective tests have been conducted to indentify the optimal channel coding methods and corresponding parameters.

5.1. Objective comparisons

Three coding approaches, namely Simulcast, MVC and V+D have been considered to encode four test contents. At the application layer, the slice mode as an effective error-resilient tool has been studied. Slice interleaving for error resilience has been implemented and integrated to the multi-view coding reference software (JMVC version 5.0.5). At the link layer, two FEC error-protection schemes, denoted as equal error protection (EEP) and unequal error protection (UEP) have been studied. The UEP schemes have been implemented such that different fractions of the FEC rate for right-view protection have been transferred to the left view for higher protection [27]. For fair comparison, equal channel resources have been provided for the different test cases. This is achieved by assigning the same "burst duration" value for the total of two channels for each coding method. Channel conditions are simulated by adding additive white Gaussian noise (AWGN) with varying SNR values to the Typical Urban 6 Taps Mobile Channel model which provides a wide range of MFER values [27]. Table 2 summarizes the experimental setting for the objective tests.

TABLE 2. VARYING PARAMETERS FOR THE CHANNEL CODING OBJECTIVE EXPERIMENTS.

Contents, 4x	Heidelberg Alleys, Knights Quest, RollerBlade, RhineValleyMoving
Coding Methods, 5x	Simulcast, MVC, MVC2, VD,VD2
Slice Modes, 5x	OFF, ON (Fixed Slice Sizes of 1300, 1000, 750, 500 Bytes)
Protection Structures, 5x	EEP, UEP (4 different cases)
Video Bitrate Ranges, 5x	300 Kbps, 600 Kbps
Channel SNR Range, 4x	17 – 21 dB
# of Experiments, 100x	100 different error patterns for each transmission

In terms of coding methods, the comparisons have shown that MVC is the best performing method. The slice methods have performed better for the cases of high packet loses, e.g. at channel SNR=17 dB. The slice size is also found to be highly content-dependent [27]. In terms of FEC rate-based error protection schemes, EEP has performed better than UEP schemes in most cases and especially for channel SNR higher than 17 dB.

In a final experiment with the video sequence *RhineValley*, the whole set of parameters for both encoding and protection has been optimized. Given a total transmission bandwidth of 1024 Kbps, the test video was first encoded with MVC at varying qualities with QPs in the range 27-33 for the left and right channel yielding highest and lowest coding rates of 972.38 Kbps (QP_L=QP_R=27) and 396.45 Kbps (QP_L=QP_R=33). The remaining bit budget for each bitstream has been spent for error protection using several FEC ratios to fit the given transmission bandwidth. Varying channel characteristics in terms of SNR (18-21 dB) have been simulated and the optimal parameters have been found by an exhaustive search. The experiments have been repeated 50 times for each channel SNR and averaged [27]. Fig. 8 shows the results of varying FEC rates for the left and right channel for two different DVB-H channel conditions. It is clearly seen that more FEC protection is required when the channel SNR decreases.

5.2 Subjective tests

A large-scale study has targeted channel coding parameters similarly to the objective experiments, i.e. EEP and UEP schemes have been assessed at two frame error rates of 10% and 20% corresponding to low and high channel SNRs. MVC and V+D have been used as coding approaches for four video contents and tested by 77 naïve test participants [28].

At low error rates, the acceptance rate for all test items has reached at least 60%, which is promising for the selected coding and error protection methods based on standard approaches. Both MVC and V+D have outperformed Simulcast at all settings. MVC has been specifically rated better at higher error rates. For low error rates, the results have shown that MVC performs best at EEP, while V+D has been better at UEP. However, varying error protection had no impact on the perceived quality at high error rates [28].

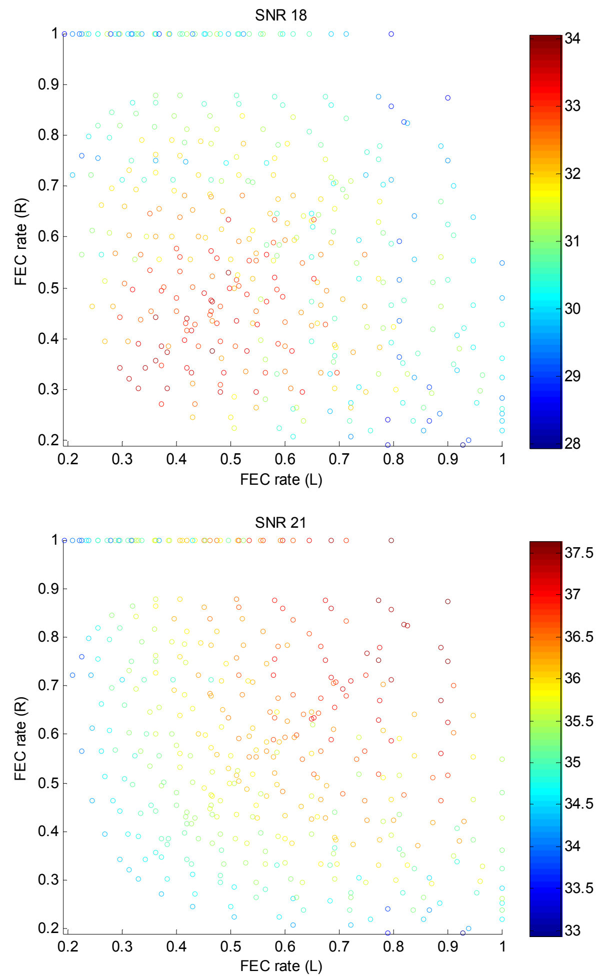


Fig. 8. Simulation results for different EEP and UEP schemes.

5.3 Channel coding recipes

We strongly favor the use of slice mode. It has outperformed the no-slice mode in most of the cases with significant difference. Only in very low-loss cases the no-slice mode have shown a marginal advantage.

MVC should be favoured as a coding approach because of its consistent behaviour. In opposite, V+D approach has shown a strong variability of performance compared to simulcast and MVC. That is attributed to PSNR as calculated on the rendered right channel and the rendering engine being also part of the approach.

EEP outperforms others protection schemes when the channel SNR is higher than 17 dB, which practically encompasses all typical cases. However, this finding is drawn in the case of relatively simple UEPs. More comprehensive UEP schemes and better tailored to the coding structure might lead to different results.

6. CONCLUSIONS

This paper has overviewed the results of objective and subjective tests leading to the selection of optimal source and channel coding methods for mobile 3D television. The viewing conditions utilizing portable auto-stereoscopic displays and corresponding artifacts have been reviewed first so to relate the processing methods with the peculiarities of the HVS.

The suggested coding methods are scalable, flexible, and backward-compatible. They form the core of a system which can give users the freedom to choose and switch between 2D and 3D viewing mode, depending on their preferences and service availability. The channel and error-protection outcomes have set directions for optimal delivery of stereo video over DVB-H.

6. ACKNOWLEDGEMENT

This work was carried out within the MOBILE3DTV project, supported by EC within FP7. The authors would like to thank their collaborators Gerhard Tech, Heribert Brust Done Bugdayci, Anil Aksay, M. Oguz Bici, Satu Jumisko-Pyykkö, Timo Utriainen, Maija Mikkola and Atanas Boev.

7. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video", MPEG Doc. No.10357, Lausanne, Switzerland, Feb. 2009.

[2] T. Szypulski, "Stereoscopic TV Standards: SMPTE 10E40 Working Group on 3D Home Master", May 2010.

[3] ITU-R, SG6 - Survey on Three Dimensional Television (3D TV) Broadcasting Systems for the purpose of obtaining the views of the ITU-R Membership.

[4] Digital Video Broadcasting (DVB), Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV), DVB Document A154, Feb. 2011.

[5] S. Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

[6] N. Dodgson, "Autostereoscopic 3D Displays," *Computer* vol. 38, pp. 31- 36, 2005.

[7] ETSI, Digital Video Broadcasting (DVB): DVB-H implementation guidelines, TR 102 377 V1.3.1, 2009.

[8] C. Wheatstone, "Contributions to the Physiology of Vision.—Part the First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision." *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371 - 394. Received and Read June 21, 1838.

[9] F. Kooi and A. Toet "Visual comfort of binocular and 3D displays," *Displays,* vol. 25, pp. 99-108, 2004.

[10] D. Hoffman, A. Girshick, K. Akeley, and M. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue", *Journal of Vision* vol. 8, pp. 1–30, 2008.

[11] Howard, I. P. and Rogers, B. J., *Seeing in Depth*, University of Toronto Press, Toronto, 2002.

[12] Percival, A. S., *The Prescribing of Spectacles*, Bristol, J. Wright, 1920.

[13] A. Boev, A. Gotchev, "Comparative analysis of mobile 3D displays", in *Proc. SPIE*, vol. 7881, pp. 78810B, 2011.

[14] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "Open profiling of quality – A mixed method approach to understand multimodal quality perception," *Advances in Multimedia*, 28 pages, 2010.

[15] S. Uehara, T. Hiroya, H. Kusanagi, K. Shigemura, and H. Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," in *Proc. SPIE*, vol. 6803, pp. 68030O, 2008.

[16] M. Mikkola, A. Boev, A. Gotchev "Relative Importance of Depth Cues on Portable Autostereoscopic Display", *Proc. 3rd workshop on Mobile video delivery*, MoViD '10, pp. 63-68, 2010.

[17] S. Jumisko-Pyykkö and T. Utriainen, "Results of the user-centred quality evaluation experiments", *Mobile3DTV technical report D4.4*, Nov. 2009.

[18] S. Jumisko-Pyykkö, T. Haustola, A. Boev, and A. Gotchev, "Subjective evaluation of mobile 3D video content: depth range versus compression artifacts", in *Proc. SPIE*, vol. 7881, pp. 78810C, 2011.

[19] 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, NO. 2, March 2000.

[20] 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), Version 1: May 2003, Version 2: May 2004, Version 3: Mar. 2005 (including FRExt extension), Version 4: Sep. 2005, Version 5 and Version 6: June 2006, Version 7: Apr. 2007, Version 8: July 2007 (including SVC extension), Version 9: July 2009 (including MVC extension).

[21] ISO/IEC CD 23002-3: representation of auxiliary video and supplemental information (Doc. N8259), July 2007.

[22] ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding, Doc. N9978, July 2008.

[23] G. Tech, A. Smolic, H. Brust, P. Merkle, K. Dix, Y. Wang, K. Muller, and T. Wiegand, "Optimization and comparison of coding algorithms for mobile 3DTV," in *Proc. 3DTV Conference*, 4 pages, May 2009.

[24] T. Wiegand, G.J. Sulivan, G. Bjøntegaard,, & A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560-576, July 2003.

[25] D. Strohmeier, and G. Tech, "Sharp, bright, three-dimensional: open profiling of quality for mobile 3DTV coding methods," in *Proc. SPIE*, vol. 7542, pp. 75420T, 2010.

[26] D. Strohmeier, G. Tech, "On comparing different codec profiles of coding methods for mobile 3D television and video", in Proc. 2nd conf. 3D Systems and Applications, 3DSA, Tokyo, Japan, May 2010.

[27] D. Bugdayci, M. O. Bici, A. Aksay, M. Demirtas, G. B. Akar, A. Tikanmäki, and A. Gotchev, "Stereo DVB-H Broadcasting System with Error Resilient Tools", *Mobile3DTV Tech. Report D3.4*, November 2009.

[28] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, G. Tech, D. Bugdayci, and M. O. Bici, "Results of quality attributes of coding, transmission and their combinations," Mobile3DTV Tech. Report D4.3, Jan. 2010.