Mobile 3DTV Content Delivery Optimization over DVB-H System
Final Public Summary

Mobile 3DTV Content Delivery Optimization over DVB-H System (www.mobile3dtv.eu) is a three-year project partly funded by the European Union 7th RTD Framework Programme in the context of the Information & Communication Technology (ICT) Cooperation Theme and its Objective 1.5 Networked Media. The project started on 1 January 2008 and was carried out by a consortium of three universities (Tampere University of Technology, Technical University of Ilmenau, Middle East Technical University), one public research institute (Fraunhofer HHI), and two SMEs (Tamlink Ltd and MM Solutions Ltd).

The project ultimate goal is to develop and demonstrate the viability of the new technology of mobile 3DTV. By adopting an end-to-end system approach, the project targeted optimal methods for 3D video content creation and coding, error-resilient transmission over DVB-H, 3D video quality evaluation and enhancement, and its enabling on a prototype handheld device.

The project produced a large database of test 3D videos in various formats, including conventional stereo video, single video augmented with dense depth information (video-plus-depth), and mixed resolution stereo, to serve in coding and transmission experiments. Available coding standards for 3D video were optimized and evaluated for the specific conditions of mobile 3DTV. Based on extensive subjective tests, H.264/MVC was selected as the representation and coding format to be supported by the end-to-end system and terminal device.

The project addressed the specific error protection of 3D content over DVB-H channel. The results from extensive objective and subjective tests suggested using slice mode at the application layer while encoding the video by MVC, using equal protection scheme for FEC protection of the left and right channel and sending the channels in two separate bursts. Joint optimization mechanism of source and channel coding were proposed.

The project put the mobile user in focus and conducted extensive large-scale user studies. The studies went from studying user expectations and requirements for mobile stereo video through developing new methodologies for user-centered quality of experience evaluation to conducting psychometric studies to help in the optimization of the critical parts of the system. Specific subjective tests were conducted to help developing a new objective metric for stereo video quality evaluation. Two new methods, namely Open Profiling of Quality and Hybrid method for multimedia quality evaluation in the context of use were developed and proposed for standardization at ITU-R SG12.

The current auto-stereoscopic display technology was identified as the most critical part for the acceptance of the new mobile 3DTV technology. A methodology for measuring optical characteristics of 3D displays, and a model of the display as an image processing channel were developed to facilitate the development of visual-optimization methods. A patent application has been prepared on the topic. In addition, advanced signal processing methods were developed to properly enhance the 3D video quality.

A purposely designed backward-compatible handheld device equipped with a last-generation auto-stereoscopic display and capable of receiving, decoding and playing stereo video streams has been developed. The device serves as a terminal of an end-to-end system, set up in two forms – a full-scale DVB-H channel and a more flexible PC-based version suitable for further research.

The project produced half-hundred scientific papers and demonstrated the system at prestigious exhibitions. Sixteen master and doctor theses were supported by the project; two proposals were
submitted to a standardization body and one patent application is being prepared based on project results.

The project generated new knowledge about user expectations, and the acceptance and satisfaction of mobile 3D video content, which shall help in the development of most suitable and appealing novel 3D applications and services for mobile users.

**Project context and objectives**

The concept of providing television-like services to handheld devices is well known. Back in 2007, results from pilots on broadcast mobile TV services amongst consumers in Finland, the UK, Spain and France had revealed clear consumer demand for such services as well as important indications over future business models for commercial mobile TV services. Recognizing this high market potential, the European Commission called for a single European standard for mobile TV and identified DVB-H as the “strongest contender for future terrestrial Mobile TV deployment in Europe”. Later on in 2008, the standard was adopted for general mobile TV use in Europe.

At the same time, chipmakers and developers have been offering more and more powerful mobile platforms enriched with multimedia capabilities. Display producers have been continuously improving the quality of visual representation of scenes on portable devices. Greater realism appealing to the mobile user has been pursued by increasing the spatial resolution and utilizing an ever more realistic gamut of light and colour. Apple’s Retina display, Nokia’s ClearBlack AMOLED, Samsung’s SuperAMOLED and LG’s Nova LCD are some recent examples about how far the mobile display technology has gone in terms of brightness, vivid colours and small pixel size. During that very same time period, autostereoscopic displays have been introduced for laptops and handheld devices as a first step to introduce the 3rd dimension in presenting visual scenes to mobile users.

In parallel to the development of mobile multimedia, 3D media has emerged as a set of new technologies to deliver realistic presentation of third dimension of audio and video and to offer immersive experience to the users consuming such content. In general, 3D media notation assumes that the content is to be viewed on big screens and simultaneously by multiple users. Glasses-enabled stereoscopic display technologies have matured sufficiently to back the success of 3D cinema and have also been enabling the introduction of first generation 3DTV. Autostereoscopic displays have been developed as an alternative display technology offering glasses-free 3D experience for the next generation 3DTV. Challenging problems in 3D content creation, coding, delivery and system integration have been addressed by research projects and standardization activities of bodies such as MPEG, DVB, ITU-R, SMPTE.

The combination of opportune European Commission support, technology conditions and the availability of appropriate standards called for the development of the next generation of mobile 3D TV services. Thus, the context, within the project was set and was then conducted, can be characterized by rapidly developing mobile technologies, great interest toward 3D and business expectations toward the speedy take off of mobile television.

The project indentified a number of core elements of the future mobile 3DTV technology to be further developed within the context. The delivery channel was considered as a whole: from capture, through coding, transmission and display. The project scenario considered stereoscopic video content created for mobile use, then properly encoded, encapsulated and then broadcast over DVB-H channel to be received, decoded and played by a DVB-H enabled portable device equipped with an auto-stereoscopic display. The concept is illustrated in Figure 1.
Important unsolved questions and problems rise out of the end-to-end mobile 3DTV system concept. The first is about the optimal data format for mobile 3DTV content. The limited spatial resolution of mobile displays and the personal use scenario limit the display format to stereo video where coding and broadcasting format can be selected between full-resolution two-channel stereo and single-view plus depth. The two channel stereo-video is easy to display, but compression efficiency is questionable. The single-view video plus depth format offers good compressibility, but requires additional techniques for depth estimation at the content creation side. Depth image based rendering at the receiving side might be too computationally intensive for the low-power handheld. A new concept of mixed spatial resolution is expected to cope with the problems of fast rendering and efficient compression. All three above-mentioned data representations have been analysed, compared and optimized for mobile 3DTV utilization by the MOBILE3DTV project.

Very much related to the data representation format is the subsequent coding, where coding algorithms have to be optimized with respect to the associated content creation format, channel and terminal-device requirements, and user experience.

As a point-to-multipoint broadcast channel, DVB-H had been especially designed to maintain error protection by adequate tools. The MPE-FEC tool is to be emphasized as it offers flexible error protection related with the importance of the content to be transmitted. This specifies research hypotheses about the applicability of unequal error protection (UEP) concerning the error protection of stereo-video content over such a channel. Cross-layer error protection is also among the research direction of interest.

Mobile3DTV concept puts the mobile user in focus. Back in 2007, little was known about the user experience of 3D video content visualized on a portable screen. Degradations caused by channel errors or compression artefacts hamper the user experience at the receiver side and have to be quantified and qualified in terms of user acceptance of quality and acceptance and satisfaction scores. Furthermore, new perceptual quality metric is expected to facilitate the objective assessment of perceived quality of mobile3DTV content. Quality is an important issue for the acceptance of the new technology, so not also measurement tools but also enhancement tools are required. Problems, such as error concealment, image and video denoising and sharpening have to be tackled at the receiver side by efficient computational imagery algorithms so to obtain a superior visual quality at an acceptable computational effort.

A purposely designed, 3D-enabled and backward-compatible portable device is needed to receive, decode and play the stereoscopic video streams. ‘3D-enable’ means the device has to have the adequate computational power do deal with the higher amount of information to be processed. Backward-compatibility means it has to be able to operate in both 2D and 3D displaying modes by incorporating suitable display. This device is the terminal of an end-to-end mobile 3DTV system, where compressed stereo video streams are delivered to the terminal over a fully-functional DVB-H channel.
The project ultimate goal is to develop and demonstrate the viability of the new technology of mobile 3DTV, comprising the following features: suitable stereo-video content-creation techniques; efficient, scalable and flexible stereo-video encoders with error resilience and error-concealment capabilities, tailored for robust transmission over DVB-H; and also the corresponding stereo-video decoders and players working on a portable terminal device equipped with a stereoscopic display.

To achieve this goal the project consortium formulated and targeted the following objectives:

1. **Develop optimal appropriate formats for stereo video content creation for mobile 3DTV in terms of compressibility, efficient rendering and user satisfaction.**

The developed formats target the best compromise between compression ratios, rendering complexity and speed, and user acceptance and satisfaction rates.

2. **Develop optimal appropriate codecs for mobile 3DTV in terms of supported spatial and temporal resolution, compression efficiency and decoder complexity.**

The developed coding and decoding algorithms are evaluated by their performance: they meet the requirements of portable displays regarding spatial and temporal resolution and fit the bandwidth provided by the DVB-H channel; the decoder complexity fits the computational power of the portable device. Last but perhaps most important, the quality of the decoded stereo video meets high acceptance and satisfaction rates measured through subjective tests.

3. **Develop optimal tools for error-resilient transmission of mobile 3D video content over DVB-H.**

Optimal tools for 3DTV transmission over DVB-H are characterized by the effective compromise between compression efficiency and MPE-FEC code rates; robustness with respect to typical channel conditions, such as mobility speed and carrier to noise ratio. Again, the user acceptance and satisfaction rates measured through subjective tests are of primary importance to select the winning tools.

4. **Gather new knowledge about user experience in terms of acceptance of and satisfaction with mobile 3DTV content, relevant to the artefacts specific to mobile stereo-video compression and transmission and to the purpose for which the user will view such media.**

This objective effectively depends on the applied research methodology for organizing and carrying out subjective tests. This includes proper specifications of the test material database, experiment procedure, evaluation scales, and methods of analysis, number of tested users in different user groups and usage contexts and their evaluation task. The specifics of the research objective call for development of new research methodologies.

5. **Develop novel metric for objective assessment of quality of processed stereo-video streams relevant to the artefacts specific to mobile stereo-video compression and transmission.**
The achievement of this objective can be measured by how well the new metric correlates with the mean opinion scores on the quality of the test video sequences and how complex it. The new metric is considered as a compound metric composed of “2D quality” and “3D quality” components. Hence, the achievement of this objective relies on existing metrics for measuring the “2D quality” components and on the development of new methods for measuring the “3D quality” components and their proper pooling.

6. Develop optimal tools for stereo video quality enhancement so as to appeal to the mobile user.

The achievement of this objective can be measured by the performance of the suggested algorithms: quality of the reconstructed stereo-video measured by objective metrics and subjective tests and computational efficiency fitting the requirements of the portable device in terms of the number of operations.

7. Develop a backward-compatible prototype portable device capable of receiving and displaying 3D video streams.

The achievement of this objective can be measured by the performance of the prototype device: its capability to receive real DVB-H broadcast, to decode and display the delivered stereo-video in real-time meeting the expected frame rate and display resolution.

8. Build an end-to-end system enabling broadcasting of compressed and stored stereo-video content over DVB-H channel.

The achievement of this objective can be measured by the performance of the system: its capability to encapsulate compressed stereo-video streams and to transmit the encapsulated packets over a real DVB-H channel, and to deliver them to the terminal device, where they are received, decapsulated, and the stereo-video streams are decoded and played on a portable device. The end-to-end system is assembled from existing commercial hardware and software modules to form a complete DVB-H channel. The new components developed within the project are software extensions for error resilience of stereo-video compressed content.

In summary, the project aimed at specifying how mobile 3DTV content should be created, coded and transmitted over DVB-H in order to be visualized on a portable display with quality accepted by the user. The project outcomes are expected to assist content providers and software and hardware developers of stereoscopic video and DVB-H tools and devices in the development and optimisation of mobile 3DTV content.
Main S&T results and project foreground

User requirements
The introduction of novel mobile services needs to provide seamless and attractive user experience. A key requirement during their development processes is the understanding about the user requirements. In mobile 3DTV, two existing singular systems, mobile TV and 3DTV, are combined into one new system providing mobile 3D television and video services. The MOBILE3DTV project consortium adopted a user-centered system development approach. The knowledge about the user requirements for this new technology has built the basis for the development process of the end-system by offering input in defining functional and non-functional requirements as well as system constraints.

In our user-centred design study on user requirements of the system, we chose a methodological triangulation of survey, focus groups, and a probe study to get a broad view on the topic of user needs and expectations. The goal of the triangulation was the elicitation of explicit and implicit aspects of user requirements. The online survey was conducted with a total of 342 participants in Finland and Germany. The questionnaire which was available through the internet thereby represented the evaluation of the preferred features for the novel mobile 3D television based on an extensive literature review. In a second study, we conducted 8 focus groups in Germany (6) and Finland (2) with a total 46 participants from different user groups. While the online survey had exploratory characteristics on the user requirements based on the literature review, the focus groups targeted the development of new scenarios and requirements to broaden the view obtained from the online survey. Although online survey and focus groups already provide a wide viewpoint on user requirements, individual needs and requirements, but also concerns and threats are often disregarded in these methods. In our triangulation approach, we extended the studies with a creative probe study combining self-documentary tools and projective tasks. We aimed getting the users’ view on future usage scenarios especially through in-the-field observations that should help us to understand contextual requirements of Mobile 3DTV. A total of 10 participants took part in the probe study.

From the results of the three methods, we were able to derive design guidelines that had impact on the whole development process and served as guidance concerning users, content and context selection for these subsequent quality optimization experiments during the MOBILE3DTV project. With respect to the user characteristics, the results showed that the requirements of mobile TV are valid for mobile 3DTV, too. The user wants to kill time, to get informed or to just watch TV while being on-the-move. But the influence of 3D can be seen as entertainment becomes an additional motivation and also the information task profits from the three-dimensional representation of the content. Combining three-dimensional video presentation with the content also increases the realism and naturalness of the content and contributes to a raised feeling of being inside the content or presence at it is usually called in user experience research. The motivations of entertainment and information in combination with the advantages of three-dimensional representation can also be found in the expected content of services. The appropriate TV contents of mobile TV (e.g. news, documentaries) are still expected by the users in mobile 3DTV services. Hereby, the added values of 3D are strongly contributing to a higher interest in the content through raised realism, atmosphere and emotional identification. However, mobile 3D television will need to add new content like movies to its services which fulfil the entertainment requirements of mobile 3DTV. But not only television content is regarded to be interesting for mobile 3DTV services. Users expect also non-television content to be part of the services and this might be even more
interesting than TV content. Thereby navigation, interactive guidance, and games are the most appropriate services according to the user requirements.

Our studies also showed that the usage of mobile 3DTV depends on the contextual situations. Short entertainment and information program is popular within mobile TV context and it also seems to be very attractive for mobile 3D television services. Nevertheless, we can also identify long-time viewing periods while travelling where movies are becoming attractive for the new system which were usually related to 3DTV services. These patterns can also be found in case of private vs. shared viewing.

**Stereo-video content creation and coding**

**Representation formats**

Stereo video can be represented in different formats namely Conventional Stereo Video (CSV), Mixed Resolution Stereo (MRS) and Video plus Depth (V+D), as depicted in Figure 2.

![Figure 2. Representation formats and their processing to common display format: Conventional Stereo Coding left, Mixed Resolution Stereo middle and Video+Depth right.](image)

For the stereoscopic representation formats, different standardized coding methods exist. These coding methods are AVC Simulcast, AVC with Stereo SEI-Message, AVC Auxiliary Picture Syntax, MPEG-C part 3 with AVC and MVC. They can be applied to the representation formats. However, not each combination is practical. Reasonable combinations are listed in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Conventional Stereo Video</th>
<th>Mixed Resolution Stereo</th>
<th>Video + Depth Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulcast</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stereo SEI</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MVC</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>MPEG-C Part 3</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>AVC Aux. Pic. Syntax</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1. Reasonable combinations of representation formats and coding methods are marked with +

**Conventional Stereo Video:** Stereo video consists of a pair of sequences, showing the same scene for the right and the left eye view, as shown in Figure 2 left. Compared to conventional monoscopic video, stereo video has twice the amount of data to be stored or transmitted.
Especially for mobile video services with its bandwidth and memory limitations, very efficient compression of stereo video is required to realize 3D instead of conventional 2D video. Efficient compression of stereo video takes advantage of the fact that the left and the right view of a stereo pair show the same scene from slightly different perspectives and are therefore highly redundant. For CSV, the representation format equals the display format, such that no conversion processing is required.

**Mixed Resolution Stereo:** A reduction of the transmission rate can be achieved by exploiting the binocular suppression theory. Earlier studies had claimed that if the sharpness of left eye and right eye view differ (see Figure 2 middle), the perceived binocular quality of a stereoscopic sequence is expected to be rated close to the sharper view. In contrast, if both views exhibit different amounts of blocking artifacts, the binocular quality of a stereoscopic sequence would be rated close to the mean quality of both views. This leads to the assumption that a stereoscopic sequence, in which one view has a reduced resolution (mixed resolution representation, MR) the same subjective quality in comparison to the full resolution (FR) case is perceived. Thus, a lower bit rate at equal quality is expected for MR. For the conversion of MRS into the 2-view stereoscopic display format, post processing in the form of upsampling is required.

**Video plus Depth Representation:** The video plus depth format consist of a conventional monoscopic color video and an associated per pixel depth map (Figure 2 right), which can be regarded as a monochromatic, luminance-only video signal. The depth data is usually generated by depth/disparity estimation from a captured stereo pair. Such algorithms can be highly complex and are still error-prone. The advantage of this format is the possible baseline variation such that stereo pairs with baselines other than the original camera pair can be generated. This requires the most complex conversion method from representation to display format of the presented formats. Here, view synthesis is used to generate the second view from the V+D format. The major challenge is the visual quality of the synthesized view, as rendering artifacts may result in a wrong and thereby annoying 3D impression in case the left and right view are inconsistent.

**Creation of stereo video database**

For the Mobile3D TV project, a stereo video database with different formats was created and updated twice over the project lifetime. The database contains two channel stereo video, video plus depth and mixed resolution stereo video.

**Stereo video database:** The initial stereo video test data contained 28 stereo and multiview sequences. Content of various genres, different types of motion, texture, structure, have been included. The database was updated in the second project year by including computer animated sequences and a coding and a transmission test sets. The issue of original creation of content for mobiles has been studied with focus on camera baseline requirements for the production of sequences for small size displays and four animated stereo sequences from the “Big Buck Bunny movie project” have been rendered regarding baseline requirements of small displays. An additional outcome of the study on baseline requirements is that also down-sampled data from existing stereo content has been tested for their compliance with the baseline requirements for mobile devices. In total, 32 stereo and multi-view sequences were reported in the second version of database. A next update included 25 new sequences suitable for mobile displays. Some of the sequences have been produced by HHI with consumer camera, thus representing the case of typical user created content. Capturing has been carried out in cooperation with the EU project “3D Phone”.
**View plus depth database:** The video plus depth database was generated using stereo video sources and employing a depth estimation technique referred to as *Hybrid Recursive Matching* (HRM), previously developed at FHG-HHI. Depth maps related with left channels have been created along with rendered right channels. In general the depth maps are not perfect. However, the estimated depth maps are temporally and spatially smooth and stable. When watching stereo video including the original left view and the synthesized right view on a small display, artifacts due to imperfect depth are almost unnoticeable.

**Mixed-resolution stereo database:** The mixed-resolution stereo database was created to serve for the development of the Advanced Mixed Resolution Stereo Coding (AMRSC) approach. For the creation, one view has been sub-sampled by a factor of two in both dimensions. The resulting sequence has been up-sampled again to serve as uncoded reference view. Subjective test were carried out to investigate differences of Full and Mixed Resolution sequences. It has been found that the subjective quality of an uncoded Mixed Resolution stereo sequence is impaired compared to the Full Resolution stereo sequence (both views have base resolution). Nevertheless, the advantage of a Mixed Resolution sequence is that fewer pixels need to be coded and bit rate can be saved or spent for a higher coding quality. Further tests show that the subjectively perceived quality difference between Mixed Resolution and Full Resolution sequences is significant for large displays and small viewing distances, but decreases for small displays and large viewing distances. Moreover it was found that with increasing base resolution the subjective difference between Mixed and Full resolution sequences decreases.

**Coding Approaches**

Figure 3 summarizes the coding approaches studied within the Mobile3DTV project.

**AVC Simulcast:** An H.264/AVC encoder is applied to each of the two input sequences independently, resulting in two encoded bit- or transport-streams (BS/TS). After transmission over the channel the two streams are decoded independently, resulting in the distorted video sequences of the stereo pair. AVC Simulcast can be applied to all representation formats.

**AVC with Stereo SEI-Message:** The left and right sequences are interlaced line-by-line into one sequence, where the top field contains the left and the bottom field the right view. The H.264/AVC coder is applied to the interlaced sequence in field coding mode, resulting in one encoded bit- or transport-stream (BS/TS). After transmission over the channel this stream is decoded, resulting in the distorted interlaced sequence. For output this sequence is deinterlaced to the two individual view sequences.

**MVC:** The H.264/MVC encoder is applied to both sequences simultaneously for inter-view predictive coding. The resulting two dependent encoded bit-streams (BS) may contain the camera parameters as auxiliary information. For transmission both bit-streams are interleaved frame-by-frame in the multiplexer (MUX), resulting in one MVC-compliant transport-stream (TS). After transmission over the channel this stream is decoded and demultiplexed, resulting in the distorted video sequences of the stereo pair. Currently, MVC only supports two-view video with identical format and is therefore applicable to CSV only. Tests with non-standard conform MVC implementation have been also applied to achieve the coding of Mixed Resolution Stereo Video. MVC can also be applied to the V+D format in its current implementation; however the missing correlation between video and depth will hinder significant coding gains in comparison to separate coding approaches.
Figure 3. Coding approaches suitable for Mobile3DTV

**MPEG-C part 3 with AVC:** The video and the depth sequence are encoded independently, resulting in two bit-streams (BS). For transmission these two bit-streams are interleaved frame-by-frame in the multiplexer (MUX), resulting in one transport-stream (TS), that may contain additional depth map parameters as auxiliary information. After transmission over the channel the demultiplexer (DEMUX) separates this stream into the two bit-streams. These are decoded independently, resulting in the distorted video and the distorted depth sequence for one of the two views of a stereo pair. This coding method is applicable to the V+D representation format.

**AVC Auxiliary Picture Syntax:** The H.264/AVC codec is applied to both sequences simultaneously but independently (with the video being the primary coded picture and depth the auxiliary coded picture), resulting in one encoded bit- or transport-stream (BS/TS). After transmission over the channel this stream is decoded, again simultaneously but independently for primary and auxiliary coded pictures, resulting in the distorted video and the distorted depth sequence for one of the two views of a stereo pair.

**Evaluation and Advancement of Stereo Video Coding for Mobile3DTV**

**Subjective comparisons:** Combinations of representation format and coding approaches suitable for Mobile3DTV have been evaluated in small scale and large scale subjective studies. In the large scale subjective studies a total of 87 test participants took part in the study. Four different coding methods that had been adapted for 3-D mobile television and video were evaluated. H.264/AVC simulcast, H.264/MVC, and MRSC using H.264/AVC were chosen as coding methods for a video + video approach. Video plus depth coding using MPEG-C Part 3 and H.264/AVC as a video + depth approach completed the coding methods.
under assessment. Six different test contents were selected in accordance to the user requirements for MOBILE3DTV. Each video was encoded at two quality levels (fixed QPs) and at two codec profiles (baseline and high profile). The results show that quality acceptance for watching the content of mobile devices was at least 60% for all cases. **H.264/MVC and the Video plus Depth** coding methods provide significantly better results for both high and low quality level. In the comparison of the codec profiles, the baseline and high profiles were equally evaluated. However, the High profile, which employs hierarchical-B pictures and CABAC, provides comparable quality for lower bitrate. These findings are promising for the future when calculation power of mobile devices will increase as lower bit rates provide advantages for transmission in terms of better error resilience.

**Advanced Mixed Resolution Stereo Coding:** Evaluations carried out in small scale tests showed the potential of the mixed resolution approach which prompted the further research on Advanced Mixed Resolution Stereo Coding (AMRSC). The three main features of the AMRSC approach are optimized down sampling, interview prediction and view enhancement using unsharp masking at the receiver side. An evaluation and optimization of down-sampling filters has been published. Existing up- and down-sampling methods as well as an alternative simple FIR filter with variable cutoff frequency for mixed resolution stereo are presented and evaluated. Coding experiments demonstrate that the simple FIR filter with a cutoff frequency of approximately 0.6 outperforms the standard methods. PSNR gains up to 1dB at a constant bit rate can be achieved. Inter-view prediction for the Mixed Resolution representation can be achieved using Multiview Coding (MVC). Here, the optimum is the down-sampling of the decoded full view and its usage as predictor for the low resolution. The third feature of the AMRSC is an unsharp masking filter. This filter especially targets to overcome the shortcomings found for the NEC display. The filter increases the subjective sharpness of the sub-sampled view; hence the approach has the potential to achieve a sharper overall image, as well as the potential to reduce the sharpness differences between both views. An advantage of this approach is that the bit rate for transmission does not increase.

**Video plus Depth:** The research on V+D targeted manipulation of depth maps so to remove irrelevant depth features; optimization of rate allocation between video and depth and improving the view synthesis.

To adapt the depth data to the used renderer and the coded texture a method for removing irrelevant information from depth maps has been developed. The depth map is filtered in several iterations using a diffusional approach. In each iteration, smoothing is carried out in local sample neighborhoods considering the distortion introduced to a rendered view. To adapt the algorithm to coding and the view synthesis, the reference view is rendered from texture data. Smoothing is only applied when the rendered view is not affected. Therefore irrelevant edges and features in the depth map can be damped while the quality of the rendered view is retained. The processed depth maps can be coded at a reduced rate compared to unaltered data. Coding experiments show gains up to 0.5dB for the rendered view at the same bit rate.

Experiments with the V+D approach focused on the use of MPEG-C part 3 with AVC as coding method. This coding method is superior to AVC Auxiliary Picture Syntax coding method since it enables an independent bit rate allocation for the video and the depth data. An optimal rate allocation for video and depth was investigated together with the generation of test stimuli for subjective evaluation. It was found that about 10% to 30% of the total rate should be assigned to depth for a best overall quality. This low rate for depth enables the encoding of video with a high quality. Gains up to 2dB for the video can be found compared to the CVS Simulcast approach.
A low complexity view synthesis algorithm suitable for mobile devices has been developed and implemented. For fast processing each row of the synthesized view is rendered using data of a line of the corresponding video and depth frame sequentially. This minimizes the amount of needed memory as well as the number of memory accesses. The implemented renderer provides two different modes: The first mode enables very fast processing by rounding disparities to integer values to avoid interpolation. The second mode supports floating point disparities by interpolation at sub pixel positions. For both modes pre-processing filters for the depth and post-processing filters for the rendered view have been implemented. The implemented renderer supports different data formats for disparity, e.g. inverse depth data or scaled disparities. The second sub pixel precise warping mode is particularly for the rendering of mobile display size views of high importance. Depth maps retrieved by down sampling from high resolution data comprise very accurate depth data. However this accuracy cannot be exploited when using full pixel precision.

**MVC:** For the advancement of MVC, new tools have been integrated into the JMVC Reference Software. These tools are a Rate-Distortion optimization using a new video quality metric and modules for error resilience. Both tools are beneficial for all representation formats since not only CSV and Mixed Resolution Stereo can be coded using the modified JMVC Software but also Video plus Depth data. For this, the interview prediction is turned off. Since the Mobile3DTV terminal device employs an MVC decoder all evaluated representation formats encoded with the modified JMVC Software can be decoded as well.

A new video quality metric, the PSNR-HVS, has been integrated to the JMVC Software for Multiview coding. The distortion classes, the Rate-Distortion interface and the Macroblock-Encoding class have been modified. Coding experiments have been carried out to evaluate the gains achieved by the modified Rate Distortion process. Constant scaling factors for the Lagrange multiplier used in the rate-distortion optimization have been evaluated. A QP dependent correction factor for the Lagrange multiplier has been determined for the NVQM. With the optimized Lagrange multiplier the rate-distortion optimization process leads to gains up to 1.6dB at high bit rates using the new video quality metric compared to an encoder using the SSD for optimization. Since the new video quality metric has been designed to emulate the human visual system, a subjectively increased video quality can be assumed as well.

The implementation of slice encoding into the H.264/MVC reference software JMVC has been carried out. Frames are stored in smaller data packets that can still be decoded independently in case of losses. An evaluation of the new encoder has been carried out using an error-free and error-prone channel. Coding tests show that the additional bit rate needed for the error resilience can be neglected and video quality only decreases slightly for error-free channels. In case of error-prone channel it has been demonstrated that the new slice mode provides sufficient error resilience and leads to a high gain of video quality.

**Error-resilient transmission over DVB-H**

Wireless networks are often error prone due to factors such as multipath fading and interferences. In addition, the channel conditions of these networks are often non-stationary, such that the available bandwidth and channel error rates are changing over time with large variations. In order to maintain satisfactory QoS, a number of technologies have been proposed targeting different layers of the networks. Among them, error resilient video coding is a technique at application layer to tackle the errors introduced during transmission. For MOBILE3DTV project, we studied error resilient transmission over DVB-H. A detailed analysis of the effects of encoding prediction structure, layering, unequal protection strategies
and allocation of resources among video quality and error correction coding have been performed.

**Set up of real and simulated communication channels**

In order to demonstrate the viability of the results, an offline and an online simulator is developed (Fig. 4). In the online simulator a PC based transmitter setup that uses a DVB-H modulator card (Dektec DTA-115/DTA-110) for creating the DVB signal is used whereas in the offline simulator, the physical layer is simulated and the error traces are generated by a software DVB-H simulator, assuming different channel conditions. This software simulator consists of 3 main parts: The first part simulates the transmitter-side hardware operations. The second part properly models the RF signal transmission environment (reflection, multi-path, shadowing, fading effects etc), and finally, the last part simulates the receiver side operations. At the transmitter side, the link layer output is taken as the input to the simulator and the bitstream is processed (interleaving, error-control coding, modulation etc.) according to the blocks of the DVB-T physical layer. The channel simulator models the behaviour of the physical transmission by degrading the final bitstream provided by the first part of the simulator, according to the channel models such as Rayleigh or Rician Fading. Finally, the DVB-H receiver side processes the bitstream (operations such as error correction and synchronization are accomplished there) and provides the final output to link layer of the receiver. In both of the cases, a set of open-source software tools (FATCAPS, JustDVB-IT) is used for creating the transport stream. In the link layer, the IP datagrams are first read column-wise into the application data table part of the MPE-FEC frame, which consists of a total of 191 columns. The application data table is encoded row-wise using a systematic Reed-Solomon (RS) code; the resulting 64 correction bytes per row are then added to the RS data table part of the MPE-FEC frame. The MPE-FEC code rate (cr) can be adjusted by either zero-padding the application data table, or puncturing the RS data table. After the MPE-FEC frame is constructed, the application data table IP datagrams are encapsulated into MPE sections. In a similar fashion, the RS data table columns are encapsulated into MPE-FEC sections. A stream of MPE and MPE-FEC sections are then put into an Elementary Stream (ES), i.e., a stream of MPEG-2 TS packets with a particular program identifier (PID). The ESs are time sliced for transmission. In order to realize DVB-H link layer, we used FATCAPS: A Free, Linux-Based Open-Source DVBH IP-Encapsulator. The FATCAPS implementation builds on JustDVB-IT, which is another open source project allowing for establishing a low-cost, highly configurable environment for DVB-T playout as well as possessing a set of tools to create, transform and multiplex MPEG transport streams, and adds the DVB-H specific features. The final output can be forwarded to a hardware modulator to perform real broadcasting. In the implementation, we also modified the FATCAPS software for offline simulation purposes. Current implementation of the software captures IP datagrams from a streamer and encapsulates for transmission in real-time. In the modified implementation, stored IP datagrams can be fed to the software with a given rate and output TS packets can be stored offline. In this way, TS packet losses simulations can be performed.
Study existing error-resilient approaches and development new error-resilient tools

We have performed different sets of both compression and transmission tests in order to find an effective compromise between compression efficiency and robustness with respect to typical channel conditions using the above system.

We have first concentrated on three coding approaches, namely: Simulcast, MVC and Video + Depth. For all of these three coding methods, there are also encoding options that affect both compression efficiency and error resilience. We have focused on the effect of one of the most important error resilience tools of the encoder, namely the slice mode. To study its influence, we have implemented slice interleaving for error resilience and integrated it to the
JMVC reference software version 5.0.5. To study the effect of UEP/EEP on these videos, equal channel resources are provided for different test cases in order to make fair comparison. This is achieved by assigning the same "burst duration" value for the total of two channels each coding method has, so that the characteristics of each coding methods determines the protection advantage/disadvantage of it over the others. During the handling of distorted video, an appropriate error concealment tool is employed.

Extensive simulations have been performed to find an effective compromise between compression efficiency/MPE-FEC code rates and robustness with respect to typical channel conditions. For the slice modes comparison, we conclude that slice mode outperforms no-slice mode in most of the cases. No slice method outperforms the slice methods with a slight PSNR improvement at low loss cases. On the other hand, when a slice mode outperforms the no slice mode, the resulting PSNR difference is much more significant. In summary, we conclude that slice mode adds better error resilience to the 3D transmission system compared to no-slice case.

Secondly, the analysis of different coding methods shows that MVC outperforms Simulcast in most of the cases. Hence, simulations which will be performed in the future do not need to use simulcast method. Another important observation about MVC modes is the importance of allocation of FEC rate and payload for a given bitrate. When the channel SNR is greater than 17 dB, MVC (coded using better quantization parameter) is better than MVC2 (formed using more FEC). However, MVC2 has higher PSNR values when the packet loss rate increases. When we examine the performance of VD modes, we observe that due to rendered view, the right view based PSNR calculation may result in much higher or lower PSNR values compared to simulcast and MVC. The comparison results of VD and VD2 show similar trend as the one when comparing MVC and MVC2.

Thirdly, the analysis of different error protection methods shows that EEP and UEP1 (the least difference between left and right protection) have better quality than other FEC modes. Indeed, EEP outperforms others when the SNR is greater than 17 dB and it shows similar characteristic to the UEP1 when the packet loss rate is high. However, we do not conclude that EEP is always better UEP in general because the results we presented in this report correspond only to the coding and UEP structure we employed. UEP schemes which suit the coding structure better may perform better.

**Optimized methods and algorithms**

After extensive tests on encoding prediction structure, layering, unequal protection strategies and allocation of resources among video quality and error correction coding, an optimization framework has been introduced on both the video bitrate and the protection rate.

The evaluations are done both using PSNR and a novel quality metric. We have seen that under loss conditions better compression methods always lead to better transmission performance since they leave room for better error protection. Therefore, MVC is better than simulcast (video+depth is better due to smaller amount of data representing depth information) and full prediction is better than simplified prediction (without considering computation complexity and power). We have also experimented layered transmission in order to see the feasibility of data partitioning. We have seen that recovery probability decreases when we split the data into many smaller bursts rather than a single burst. Considering the backward compatibility and the effect of layering, we concluded to use two bursts for transmission, one for each view. Our unequal error protection strategies provided robustness for three layered transmissions, where the data regarding the left and right view frames are split into bursts according to their priorities. For two layered transmissions, where
frames of left view and right view are separated from each other, unequal error protection did not provide an improvement over equal error protection cases. But we have seen that the effect of video bitrate and FEC protection allocation is more significant on the received video quality. Using full prediction MVC with equal error protection, we have experimentally determined the optimum allocation of transmission bits among video quality and protection that results in minimum distortion in the received video. This process is similar for MVC and video+depth coding methods; hence we provided the results for MVC only.

In the second part of the study, we have modelled distortions present in the end-to-end system, in order to have a parametric tool to use in optimization. The modelling includes encoder distortion, decoder distortion, channel effect and error protection. We have provided the validity of the model by fitting the models result to exhaustive search. This modelling algorithm is exhibited using a small subset of the parameter set. Since it fits the actual exhaustive search results quite well (c.f. Figure 5), it is possible to extend the study and use the modelling for optimization for the rest of the system parameters.

![Figure 5 Simulation and modelling results for SNR 19](image)

**Developing UC-QoE Evaluation Methods**

The goal of multimedia quality studies is an optimization of critical quality factors produced under strict technical constraints or resources with as little negative perceptual effects as possible. Commonly, these studies have been conducted in accordance to standardized test methodologies provided by the ITU or other standardization bodies. These methods have provided a good basis for the hedonistic excellence of subjective quality. Recently, a broader view to quality has been taken by covering other aspects of active perception in the evaluations including in knowledge, different levels of human information processing, or even contextual behaviour. Multimodal quality evaluation studies have started to underline the characteristics of active and multilayered quality perception. Quality does not only derive from the characteristics of stimuli, but also from usage-, task-, and context-dependent factors. Within this track, a User-centered Quality of Experience (UC-QoE) framework was developed during the MOBILE3DTV project. User-Centered Quality of Experience is a framework evaluation of experienced multimodal quality (UC-QoE). We define it as an evaluation method which is a collection of factors and independent methods that relates the quality evaluation to the potential use of system or service. It takes into account 1) potential users as quality evaluators, 2) necessary system characteristics including its potential content and critical system components, 3) potential context of use resulting evaluation in the
controlled experimental and quasi-experimental settings 4) evaluation tasks are in relation to expected goals of viewing, they aims also at understanding the interpretation of quality and include ergonomic measures. Within this framework, two evaluation methods were developed, called 1) Open Profiling of Quality and 2) Hybrid method for quality evaluation in the context of use.

**Open Profiling of Quality (OPQ)** is a mixed methods evaluation approach that combines the evaluation of quality preferences and the elicitation of idiosyncratic experienced quality factors. OPQ is targeted for naive participants applicable to experiments with heterogeneous and multimodal stimulus material. The combination of quantitative evaluation and sensory profiling allows assessing the excellence overall quality and individual quality attributes in parallel. Our final goal is to construct a link between quality preferences and the underlying individual quality attributes.

**Hybrid method for quality evaluations in the context of use** tackles the problem of evaluating quality in the natural circumstances. Context of use is composed of several components, called physical, temporal, task, social, and technological and informational context and their properties which surround mobile human computer interaction. When compromising produced quality for consumer services used in dynamic and heterogeneous mobile contexts ecological validity of conventional quality evaluation methods can be questioned. To conduct quality evaluation study outside controlled laboratory conditions requires a shift in paradigm towards quasi-experimental evaluation and novel tools for evaluation procedure. The method is composed of 1) the process, including planning, data collection and analysis, 2) understanding the factors that surround the assessment in the context on a macro level (high-level features of a whole situation) and a micro level (situational, e.g. second by second) and 3) the use of several techniques over the study (e.g. characterization of context and relating it to user requirements, excellence evaluation, interview, observation, and task load analysis). The detailed presentation of the method provides also instructions for carrying out such experiences to minimize threats of validity.

Our extension of common quantitative methods towards a user-centered evaluation framework has provided deeper knowledge and understanding about perceived quality of mobile 3D television and video systems. We were able to explain the hedonic excellence and the impact of different quality parameters by understanding the relationship between quality and depth using sensory profiling. The descriptions of depth and error-freeness were attached to good quality when visual presentation mode, coding factors or transmission settings were varied. Bad perceived video quality is a limiting factor for taking benefit of the 3D perception. In addition, spatial quality dominated the temporal quality, i.e. blocking artifacts had higher impact on perceived quality than flicker or judder. Without descriptive data, the reasons beyond the quantitative excellence had been based on assumptions, while sensory data as a single method is not capable to show hedonic preferences. In addition, we were able to identify different preferences for auditory or visual perception in audiovisual environments between our test participants. This result drives the need for new tools of classifying test participants beyond the common visual and auditory screenings.
The results of quality evaluation experiments collected either in analogue or natural context of use underlined the contextual quality requirements and critical factors to actual use. Increase of freedom to hold device on hand with face-down position in relatively calm surroundings highlighted the importance of maintaining the sweet spot for viewing when using 3D with mobile device. This result was shown with two different display technologies used (parallax barrier, lenticular sheet). In the busy natural contexts of use (bus, station) the
viewing situation is characterised by surrounding noise, variable lightning conditions, interruptions, dynamic social context, and pseudo move of vehicle as well as active division of attention between the task on mobile device and surroundings. To maintain the optimal viewing conditions in these contexts more effort was required (expressed as increased actions to maintain the sweet spot) from users for 3D compared to 2D viewing. Future 3D display technologies on mobile devices need to respond to these challenges to provide pleasurable viewing conditions taking into account ability to maintain optimal viewing position and active glazing off from device as a part of mobile use. The results also highlighted the contextual quality requirements in two ways. At the low quality, people are more tolerant towards errors in the natural circumstances. In contrast, at the good perceived quality level needed system quality depends of the characteristics of context indicating possibilities for the contextual quality optimization. To summarise, quasi-experimental studies started to reveal the aspects of actual use and go beyond conventional quality evaluation.

Descriptive model for 3D video on mobile device presents the perceptual characteristics for mobile 3D video. The model was composed of data-driven analysis of five different studies which utilised both OPQ and interview-based methods with over 90 naïve participants. The experiments contained an extensive and heterogeneous set of produced quality by varying content, level of depth, compression and transmission parameters, and audio and display factors for 3D. The descriptive quality of experience is constructed from four main components, 1) visual quality, 2) viewing experience, 3) content, and 4) quality of other modalities and their interactions. Later, the model has been operationalized to act as a descriptive data-collection tool for vocabulary-based evaluation.

<table>
<thead>
<tr>
<th>COMPONENTS (major and sub) - Bipolar impressions</th>
<th>DEFINITION (examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VISUAL QUALITY</td>
<td>Descriptions of quality of visual modality, divided into depth, spatial and motion quality</td>
</tr>
<tr>
<td>DEPTH</td>
<td>Descriptions of depth quality in video, characterized by perceivable depth, its natural impression, composition of foreground and background layers, and balance of their quality</td>
</tr>
<tr>
<td>Perceivable depth</td>
<td>Ability to detect depth or variable amount of depth as a part of presentation</td>
</tr>
<tr>
<td>Impression of depth</td>
<td>3D effect creates a natural, realistic and error-free impression instead of an artificial and erroneous impression (e.g. too much depth, double objects, shadows, seeing through objects)</td>
</tr>
<tr>
<td>Foreground-background layers</td>
<td>Depth is composed of foreground and background layers and the impression of the transitions between these layers can vary from smooth to distinguishable separate layers</td>
</tr>
<tr>
<td>Balance of foreground-background quality</td>
<td>Balance between the excellence of foreground and background of image quality (e.g. sharp foreground, blurry background or vice versa, or they are otherwise not in balance)</td>
</tr>
<tr>
<td>SPATIAL</td>
<td>Descriptions of spatial image quality of video, characterized by clarity, block-freeness, colors, brightness, contrast and ability to detect objects and edges</td>
</tr>
<tr>
<td>Clarity of image</td>
<td>Clarity of image in overall -- Clear (synonyms: sharpness, accuracy, visibility) vs. unclear (synonyms: blur, inaccurate, not sharp)</td>
</tr>
<tr>
<td>Block-free image</td>
<td>Existence of impairments with visible structure in image (e.g. blockiness, graininess, pixels)</td>
</tr>
<tr>
<td>Color, brightness and contrast</td>
<td>Excellence of colors, brightness and contrast</td>
</tr>
<tr>
<td>Accurate/inaccurate</td>
<td>Ability to detect necessary objects and details, their edges and outlines</td>
</tr>
<tr>
<td>MOTION</td>
<td>Descriptions of motion of video, characterized by fluency, clarity and nature of motion</td>
</tr>
<tr>
<td>Fluency of motion</td>
<td>Excellence of natural fluency of motion -- Fluent (dynamic, natural) vs. influent (cut-offs, stops, jerky)</td>
</tr>
<tr>
<td>Clarity of motion</td>
<td>Excellence of clarity of motion (e.g. accuracy under fast movement or movement out of screen) -- Clear, sharp vs. blurred, pixilated</td>
</tr>
<tr>
<td>Nature of motion</td>
<td>Nature of motion in the content or camera movements - Static (synonym: slow) vs. dynamic (synonym: fast)</td>
</tr>
<tr>
<td>VIEWING EXPERIENCE</td>
<td>Descriptions of viewing experience, characterized by ease and pleasantness of viewing, enhanced immersion in it, visual discomfort and impression of improved technology and overall quality</td>
</tr>
<tr>
<td>Ease of viewing</td>
<td>Easy to concentrate on viewing (e.g. free from extra effort and learning, viewing angle does not interrupt viewing)</td>
</tr>
<tr>
<td>Pleasantness of viewing</td>
<td>Pleasurable viewing experience, also for a longer period of time (e.g. 15min)</td>
</tr>
</tbody>
</table>
Enhanced immersion

**Enhanced/Not enhanced**
Feeling of enhanced immersion into the viewing experience (impression of becoming a part of the events in the content, involvement, fun and improved impression of naturality, like-likeness, tangibility and realism)

Visual discomfort

**Experienced/Not experienced**
Feeling of visual discomfort (eye-strain) and descriptions of related discomfort symptoms (headache, general discomfort)

Comparison to existing technology

**Improved/Not improved**
Impression that provided quality of new technology (3D) is higher than quality of comparable existing technology (e.g. 2D video on a mobile device)

Overall quality

**Good/Bad**
Impression of excellence of quality as a whole without emphasizing a certain factor (e.g. excellence over the time, relation between erroneous/error-free)

**CONTENT**
Descriptions of content, their content-dependency and interests in viewing content

**OTHER MODALITIES**
Descriptions of quality of audio modality and interaction between quality of audio and visual modalities

Audio
Audio and its excellence

Audiovisual
Bimodal audiovisual quality (synchronism and fitness between media) and its excellence

During activities towards standardization of the developed User-centered Quality of Experience evaluation framework, we additionally identified the need for a systematic comparison of mixed methods to be able to holistically compare the different research methods. Our comparison model was developed upon an extensive literature review about comparison criteria that are applied in different domains of user evaluations. We have presented a model of four major categories of economy, excellence, assessment, and implementation. The classes were formed by categorization of the identified comparison criteria. A first approach of operationalization of the comparison model for comparison OPQ and a sensory profiling approach based on fixed terminology has been undertaken.

![Comparison Model with the four categories and corresponding sub-criteria](image)

The two methods developed, Open Profiling of Quality and Hybrid method for quality evaluation in the context of use were submitted to the ITU-T SG12 as proposals for standardization on questions 13/12 “QoE, QoS and performance requirements and assessment methods for multimedia including IPTV”. The proposals were accepted for presentation in the ITU-T SG12 general meeting in January 2011.
Visual quality enhancement

Classification, modelling and simulation of stereoscopic artefacts

Estimation of the quality is the key factor in design and optimization of systems for stereoscopic 3D video content delivery. A first step towards objectively measuring and enhancing the visual quality is to identify the artefacts which could arise when dealing with stereoscopic content. MOBILE3DTV project has provided a detailed overview and categorization of stereoscopic artefacts which occur during delivery of mobile stereoscopic video followed by their modelling and simulation. As extensive Matlab tool has been designed capable of simulating a large group of artefacts and manipulating 3D videos in different formats.

The study considered the stages of mobile 3DTV systems as sources of artefacts and put them against the various depth cues – accommodation, binocular depth cues, pictorial cues and motion parallax, as illustrated in Figure 7. Thus, the classification listed potential sources of artefacts, and contemplated how they would be interpreted by the different visual subsystems – structural, colour, motion and binocular vision (see Figure 8). For each artefact we identified the causes for occurrence, illustrated the visual appearance and discuss the possibility of occurrence in our mobile 3DTV system.

The organization of the artefacts in groups allowed for their simulation in blocks in an artefact simulation system. Within the proposed simulation framework, the following artefacts have been simulated: 1) capture artefacts – content resizing (causing aliasing and improper disparity), blur, motion blur, barrel/pincushion distortion, keystone distortion, temporal mismatch, colour mismatch, and cardboard effect; 2) coding artefacts – blocking artefacts (as caused by harsh quantization), block-edge discontinuities, colour bleeding, staircase effect, cross distortion, depth bleeding and depth smoothing; format-conversion artefacts – disocclusion and temporal inconsistency; 4) transmission errors – error patterns of the DVB-H channel and channel losses; 5) reception and visualisation artefacts – temporal mismatch, content resizing, vertical banding and cross-talk. The introduced framework is modular and can be easily extended with additional types of artefacts while existing simulation blocks can be improved or replaced. A specific Matlab implementation includes algorithms for simulation of all artefacts, listed above, as well as a graphical user interface for setting parameters of each simulation run. The implementation has been made publicly available to be used by the community.

Figure 7. Depth perception as a set of separate visual “layers”
Design of objective perceptual quality metric for mobile stereo-video

The work on new objective quality metric begun with creating a model of the stereoscopic human visual system (HVS). Such model combines knowledge about HVS anatomy and physiology, psychophysical experiments, models of HVS features, an overview of existing visual quality metrics as well as results of specifically designed and conducted subjective tests. Specifically, two subjective studies have been carried out to complement the literature review on the HVS functioning and modelling. In the first study, subjective tests have been designed and conducted with the aim to investigate the impact of different depth cues on the perception of depth on portable auto-stereoscopic display. The aim of these tests was twofold: first, to estimate the relative importance of binocular depth versus various pictorial depth cues on the depth perception; and second, to assess the perceptual influence of video quality degradation over various cues and the ability to make depth judgements in degraded imagery conditions. The study 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. It also emphasized the effect of compression as the primary quality-influencing factor on both the accuracy and the speed of depth perception. Following these
findings, the second study addressed the problem of visual distortions introduced by compression versus varying depth presence. The aim of the study was to quantify the influence of the, possibly distorted, depth on the formation of the cyclopean view and the role of 2D compression artefacts on both the overall quality perception and the presence of depth. It aimed at generating a dense grid of mean opinion scores of differently distorted videos (five QP factors for compression, three stereo baselines – zero, narrow and wide) to be used for benchmarking the objective metric being developed. The study demonstrated that compression artifacts are a dominant factor primarily determining the quality of experience compared to varying depth range. 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. This finding coincided with the WP2 studies about stereo baseline confirming the possibility to directly convert high-resolution stereo video for mobile use.

Based on the developed HVS model, the research targeted a full-reference feature-based objective metric with feature-extracting blocks inspired by the stereoscopic visual perception and in accordance with the results of the supporting subjective tests. The proposed metric includes the following modules:

- **2D module** – it effectively estimates the quality degradation between 2D image segments by employing the effect of contrast sensitivity (CS) masking in DCT domain. The image segments are transformed in DCT domain where the coefficients against local cosine functions are properly masked to account for the effect of contrast masking. A modified version of MSE quantifies the quality difference.

- **Cyclopean formation module** – the cyclopean view is formed only implicitly to avoid the introduction of further (warping) distortions. At this stage, the effect of saccades and binocular disparity are taken into account by forming a 3D structure, to be jointly decorrelated by 3D-DCT. CS masking is applied in the resulting stack of coefficients and the modified MSE is calculated between the distorted and reference stereo pair.

- **3D module** – it accounts for local variations of disparity, calculated with respect to the size of foveal image related with the display size and resolution. Local disparity variations are emphasized for small 2D quality degradations and are less influencing for high degradations, thus following the trend indentified by the subjective tests.

The designed metric has shown good correlation with the MOS collected through subjective tests, as illustrated in Figure 9. It outperformed all available 3D metrics as well as 2D metrics modified for 3D use. However, at that stage it is simply a metric to be used with mobile stereo-videos as all assumptions and tunings have been done against mobile-resolution test content.
Research on this topic has gone beyond encoding and transmission to develop methods supporting the mobile 3DTV stages and leading to improved stereo-video quality. The following methods have been studied and further developed:

**Sampling rate conversion for stereo video** – such conversion is required when higher definition video is to be downscaled to mobile resolution. It also appears in the mixed resolution stereo representations schemes, where one of the channels is deliberately downscaled for the sake of more effective compression and then up-scaled back for visualization. Standard up- and down-sampling methods together with an alternative simple FIR filter for down-sampling with variable cutoff frequency have been compared. Coding experiments demonstrated that the simple FIR filter with a cutoff frequency of approx. 0.6 outperforms the standard methods. PSNR gains up to 1 dB at a constant bit rate or bit rate savings up to 30% at a constant PSNR can be achieved.

**Denoising of stereo video** - it might be needed when the content to be delivered to the mobile device has been created under low light conditions. Noisy channels are more problematic not only for creating pleasant stereo perception but also for compression, depth estimation and view synthesis. One of the most competitive video denoising methods, abbreviated as VBM3D (video block matching in 3D) has been evaluated for its applicability and performance for stereo video. Experiments demonstrates that by using this method, the denoised left and right video channels are restored with very high quality, where all 3D visual cues are well preserved and in fact even enhanced. From implementation point of view the results show an equal performance of the algorithm when applied independently to the two channels or jointly. Marginal improvement can be expected only for content with high amount of motion.

**Deblocking of depth maps** is perhaps one of the most important pre- and post-processing tasks for the representation format ‘view plus depth’ since practitioners tend to employ standard, i.e. block transform based, compression methods. For refinement of depth maps, five filtering approaches has been compared. The methods included Gaussian smoothing, standard H.264 deblocking, and three more sophisticated methods utilizing structural and colour constraint from the presented color video channel. The methods have been optimized...
with respect to the quantization parameter of the H.264 compression used. For the best performing method, we have suggested practical modifications leading to a faster and memory-efficient implementation and extended it to video and for more general types of depth impairments (e.g. resulting from fast depth estimation or noise). Our final approach yield highly time-consistent depth sequences adequately restoring the depth properties of the 3D scenes. It has been further optimized for real-time implementation on a GPU platform.

**Deblocking and sharpening of stereo video** is perhaps the most desirable post-processing operation as blocking artefacts have been identified as the most annoying and quality-deteriorating factors in stereo-video perception on mobile displays. A new algorithm for joint deblocking and sharpening of stereo video has been developed based on the modern approach of non-local collaborative filtering. This approach has been extended toward stereo video by suitable grouping of similar patches along temporal dimension and by using the disparity relations to form two inter-related trajectories of grouped patches. This comprehensive 4D structure is decorrelated by a hybrid efficient transforms, namely DCT along spatio-temporal directions and Haar wavelet transform between stereo frames. Artefact suppression is performed through transform-domain thresholding followed by a second-stage transform-domain Wiener filtering. An elegant stereo sharpening can be accomplished in the transform domain by alpha-rooting. It considerably improves the stereopsis in the stereo pairs. The procedure leads to visually pleasant results and is also compared favourably against its simplified versions, i.e. sliding DCT filtering and bilateral filtering.

**Disocclusion hole filling by non-local inpainting**: Disocclusion is an annoying effect encountered in view synthesis in V+D representation and has been targeted by studying different approaches. Three state-of-the-art methods based on dept preprocessing (smoothing), diffusion filtering and non-local inpainting have been compared. The latter operates by taking a weighted mean of the k most similar patches to fully recover a patch containing disoccluded areas. This approach has been found superior in terms of performance and has been further modified. Improvements have been suggested for speeding up the search for similar patches and improving the quality by proper prioritization and non-local mean weighting. In order to prove the feasibility of the method, objective and subjective tests have been performed, comparing the technique with the other techniques. The developed technique showed high and consistent ranks over various video contents with reduced computational cost.

**Error concealment of V+D and MVC encoded and transmitted video** Error concealment is required for restoring lost frames in any scenario of video streaming over error-prone channels. Both V+D and MVC formats have been considered. For the V+D representation, two schemes have been considered, namely concealment using complementary frame and concealment using complementary motion vectors. For the MVC, concealment using complementary motion vectors has been studied and experimental results analysed.

**Optimized visualization on portable auto-stereoscopic displays**

Display-specific artefacts have proved crucial for the positive user experience of 3D on portable displays. Therefore, the research on visual optimization has been directed toward understanding and mitigation of the display-specific artefact. We have developed a methodology for finding the minimal set of most important optical parameters, which can be used for quality characterization of a 3D display. The parameters are to characterize three major types of display-specific artefacts – viewpoint related, optical mask related and content related. For each type of artefact, a characterization is formulated about how it is perceived by the HVS in natural observation conditions. A set of measurements assess the perceptual
influence of each artefact. Based on the measurements and a model of the display as an image processing channel, model parameters are identified which can be later used to design image processing algorithms for artefact mitigation. Using the measurement methodology, comparisons of available portable 3D displays have been performed and the following artefact mitigation methods have been developed:

- Artefact mitigation for 3D displays
- Pseudoscopy correction
- Point-of-view optimization
- Extended head parallax
- Antialiasing for stereoscopic displays (subjectively assessed)

**Design of 3D-enabled terminal device**

**Selection of development platform, DVB-H front-end and auto-stereoscopic display**

The development platform was selected to meet criteria such as acceptance by Tier 1 mobile phone vendors, means for connecting stereoscopic LCD and DVB-H receiver, sufficient processing power to decode simultaneously at least two H.264 BP simulcast coded streams (QWVGA@25fps), means to implement custom type H.264 decoder as a base of multi-view or view+depth type of stereo video encoder, sufficient processing power to execute DVB-H stack, GUI and decoding. OMAP 3430 manufactured by Texas Instruments was selected as it proved capable of providing suitable features supporting the MOBILE3DTV project goals. Using the platform, demanding applications such as stereo-video decoding and playing can be parallelized by employing different available cores (ARM, IVA and ISP) with sufficient speed provided by the high clock rate. XGA display support allows for choosing high-resolution LCD. Power reduction technologies enable the implementation of power-demanding applications such as rendering stereo-video with increased backlight. Various interfaces allow for easy embedding extra modules, e.g. the DVB-H front-end. The support of high-level operating systems is also a key factor for rapid application development.

The DVB-H platform was selected to meet criteria such as acceptance by Tier 1 mobile phone vendors, low power consumption, mobile form factor, availability of a development kit for rapid system integration and software API for HLOS. Based on these, DIB9080H module from DiBcom was selected.

During the project lifetime, four different auto-stereoscopic displays were interfaced to the platform: 4.3”WVGA (800px x 480px) transmissive LCD display by MasterImage, employing parallax-barrier technology; 3.1” LCD with 427x240 px resolution both in 2D and 3D mode by NEC LCD, employing lenticular technology with horizontal double density pixel (HDDP) arrangement, ensuring the same resolution in 2D and 3D mode; 3.1” LCD with 400x480 px resolution in 2D and 3D modes by 3M employing 3D film, light-guiding and time-sequential view switching; and finally 3.1” 640x360 px display by NEC LCD, employing the same lenticular HDDP technology but at higher spatial resolution. The displays have been studied for their performance employing the methodology developed in WP5 and interfaced by both hardwire means and by developing required software drivers.

**Versions of terminal device**

The device was developed in stages and five versions were released.

**First version:** Linux was selected as platform-specific operating system and the corresponding file system, first version of stereo video decoder based on H.264 and
supporting simulcast was developed. Conversion between size-by-side stereo and interleaved stereo was implemented. MasterImage display was interfaced through a daughter board.

**Second version:** New hardware components were developed including a PCB board for interfacing the DVB-T/H module and the OMAP3430 EVM, also compatible with the new releases of the OMAP3430 SDP, namely Zoom-1 and Zoom-2. The NEC auto-stereoscopic display was integrated. Concerning software, a driver for the DVB-T/H module was developed and debugged; modifications were made on the SDK provided by the vendor, so to suit the development platform. Additional SPI test application, needed for measurements and validation of the interface, and additional flexible framework for lighter implementation of all project-related software components on the device were developed.

**Third version:** The version targeted the final form factor where all components such as the processing platform, auto-stereoscopic LCD, and DVB-H receiver have been coupled together and operate simultaneously. The version integrated the 3M auto stereoscopic display and the new version of the processing platform (OMAP3621) in then-considered final form factor. The housing for the device was selected, procured and used for integration. The version also featured new DVB-H low level drivers, new auto-stereoscopic LCD drivers and MVC decoder based on H.264 decoder.

**Fourth version:** The final form factor was improved. A new hardware platform was developed to reach the form factor and the functionalities of modern PMPs (Portable Media Player) enriched with Mobile 3DTV features such as a modern auto-stereoscopic LCD and DVB-H receiver. The new version of the auto stereoscopic LCD (NEC) module was integrated. The new version of the processing platform (OMAP3621) was developed in a new form factor. For the fourth version, the Android OS has been ported. A new version of the MVC decoder was designed based on a DSP implementation.

**Fifth version:** This is the final version, featuring all planned components. All remaining issues such as efficient power consumption, antenna support and display support have been addressed. Linux drivers for 2D graphics and 3D video for this display have been developed and tested. DVB-T/H module from Dibcomm has been integrated also and tested. Issues related with the MVC decoder such as dropping of frames, and slowing frame-rate were fixed through better memory-management solutions. Another decoder supporting V+D was integrated together with the accompanying view renderer. Thus, both stereo video and V+D formats are supported at the terminal device. A minimalistic user interface has been developed, allowing for selection of TV channels as well as playing 3D media content (videos and images) stored locally in the device.

*Figure 10* shows different versions of the device integrating different displays.
End-to-end mobile 3DTV system over DVB-H

A full-featured transmitter setup was installed at Tampere University of Technology (Figure 11). Equipped with 50W power amplifiers and rooftop antennas, it is capable of delivering DVB-H broadcasts within the university campus. The DVB-H playout server by Cardinal is used for creating timesliced transport streams of A/V content in MPEG-4 files or live IP streams. File delivery over FLUTE data carousels is also supported. Transmitting complete, pre-generated or recorded transport streams is also possible.

Figure 11 DVB-H Transmitter setup

The DVB-H exciter takes the transport stream as its input and creates an OFDM modulated DVB signal to be broadcasted by the power amplifier and rooftop antenna. Optionally, it is possible to utilize also a diversity unit that prepares the RF signal for two-antenna transmitter diversity broadcasting.
For laboratory experiments and demonstrations, TUT, METU, and MMS have all obtained a PC based transmitter setup that uses a DVB-H modulator card (Dektec DTA-115/DTA-110) for creating the DVB signal (Figure 12). The DVB-H link layer is implemented through a Free, Linux-Based Open-Source DVB-H IP-Encapsulator (FATCAPS). Within the project, we have modified the FATCAPS software for offline simulation purposes.

![Figure 12 DVB-H Transmitter setup](image)

For research and development work, within the project, we have developed a PC based receiver with core functionality based on the project-developed Decaps. Compared to previous solutions, it adds features such as MPE-FEC error correction and collection of error statistics. The software components of the Linux receiver setup are illustrated in Figure 13.

![Figure 13 Linux based DVB-H receiver](image)

The final end-to-end system was setup at TUT in March 2011. Videos for three TV channels featuring sport, documentary, and animation respectively were encoded by H.264/MVC following the results of format and coding methods studies. They were then encoded to MPEG-2 transport streams utilizing the error-protection methods suggested within the project and the transport streams were stored at the playout to be constantly broadcasted over the TUT campus. Subjective usability tests of the receiving terminal device and the related service were carried out. The goal of the tests was to gather knowledge about the perceived quality of end user experience involving a prototype mobile device and real mobile 3DTV service. A total of 41 participants, categorized as early adopters took part in the test. The overwhelming majority of users indicated that they had good or at least satisfactory opinion about the system in general, indicating the clear interest for this type of devices and services. In addition users see, and expect, possibilities to combine mobile 3DTV devices with autostereoscopic screens with other technologies. Video games and internet browsing were mentioned as possible applications in prospective communication systems. The elicited results about the final MOBILE3DTV system are very much in line with the user
requirements collected at the beginning of the project. The user-centered optimization of the system has provided a well-accepted end-to-end system for users and underlines the possibilities that next mobile devices and related services can offer applying 3D technology.

Potential impact
Mobile 3DTV, as an emerging technology is at the threshold of making a transition from an area of academic interest and technological R&D onto the mass market. 3D video and broadcasting technologies have already been successfully adopted as a part of modern consumer electronics, movie and home entertainment industries. Transition to mobile platforms is an imperative for many emerging technologies. Mobile 3DTV, thus constitutes the next logical step in expansion of 3D media, as a paradigmatic shift, expected by the consumers. First portable devices with the authostereoscopic 3D screens are already entering the market.

In modern socio-economic landscape the success of any emerging technology is measured primarily in terms of market penetration and user acceptance. The subjective perception of quality of user experience is fundamental factor that determines both the speed and the breadth of acceptance of any new technology. In Mobile 3DTV project we have taken so-called user centric design approach, in order to make a bridge between, hard to gauge subjective impressions of user experience and underlying technological properties of a Mobile 3DTV system.

The stated objectives of the project and results achieved targeting these objectives, are expected to have significant socio-economic impact. This impact is expected to be multi-fold, touching both the provider and the consumer end of modern marketing equation.

Societal impact of Mobile 3DTV project steams from two sources, the intrinsic nature of 3D technology as a new media, and from mobility of the content delivery device. In the same manner results related with the Mobile 3DTV project is expected to have an impact onto two large sections of the society, general public, as the future consumers of mobile 3DTV content, and the industry especially producers of devices, content providers and content broadcasters.

In what follows we elaborate in detail the expected societal impact of the project results.

3D as a Paradigm Shift
Ever since the invention of moving pictures, one of great technological and conceptual paradigm shifts in history, the notion of adding the 3rd dimension hallmark of visions of near technological future. Public has grown to expect the transition to 3D as the next logical step in the development media technologies. The years of persistent presence in media has created a strong public exportation about how a 3D video experience should look and feel. So much so, that every person has at least some preconception about the 3D technology. We stand now on the cusp, when 3D technology finally makes transition from research community onto the mass market. The emerging 3D enabled devices intersect with public expectation of 3D technology in several often contradictory ways. On one hand product developers try to satisfy the public expectation, in a way following the preconceptions built over the years by futuristic imagery from media. Quite often, the properties of such devices reinforce already established perception. On the other hand, like in the case of any new technology, they extend the paradigms, and opens new possibilities. Yet in many cases the current technological solutions still lag behind public expectations, restating the evident need for further research in this domain.
This established public expectation underlines the need for user-centered approach in design of future Mobile 3DTV systems. User studies conducted as a part of Mobile 3DTV project and elsewhere, show a great level of acceptance for 3D technology in general and Mobile 3DTV in particular. However, this research also shows a high level of expectation of the quality of provided 3D content. The true value of 3D technology lies in its potential to bring about an entirely new language of visual expression, and not only to serve as a marketing support for already existing video media.

User-centered design is fundamental in this respect. Mobile 3DTV project has taken precisely this approach when developing the proposed end-to-end Mobile 3DTV system. A large part of project activities have been focused on gathering knowledge about the user experience and perceived quality. This knowledge forms the foundation upon the technological solutions proposed in this project were built.

Every major paradigmatic shift necessarily requires significant technological effort. Industry, especially, device manufactures, content providers, entertainment industry, etc. have already made significant investments in 3D technology in one form or another. Mobile 3D technology is a part of this wider 3D revolution. As such it shares a part of its broader societal impact. In order to be successful, Mobile 3DTV needs to build upon foundations set by development for 3D technology in general. Any development in Mobile 3DTV needs to be coordinated with preexisting efforts of industry regarding 3D technology. Several stated goals of Mobile 3DTV project had in mind, explicitly to address the issue of compatibility of needed standards for Mobile 3DTV with existing technological framework.

**Mobility as an imperative for emerging technologies**

Mobility, i.e. the need to change one’s place within his physical environment, represents one of essential human needs. Indeed, it arises from most fundamental evolulational imperatives which shaped us as beings. It is reflected to every aspect of our lives. It is reflected fundamentally upon the way we interact with technology. 3D as a technological concept is already established in non-mobile setting in form of 3D TV in home entertainment and 3D movies in cinemas. User studies conducted within the MOBILVE3DTV project indicated that the public in general perceives the transition of 3D technology onto the mobile platform as imminent.

Streaming video content already constitutes a large part of traffic over mobile telecommunication networks. As 3D video content grows more prominent the users expect to be able to consume it using their mobile devices. The end-to-end system developed as a part of Mobile 3DTV project has provided a prototype of a dedicated Mobile 3DTV device, however the same framework could be easily adapted for applications in embedded systems, such as mobile phones, tablets, netbooks, etc. This observation is in line with the prevalent trend of device convergence, where functions of erstwhile different devices become fused in a single consumer product, i.e. phone, camera, media player, etc. Usability tests indicate that the future real-life applications of Mobile 3DTV systems would probably involve some level of integration with existing families of products. First mobile 3D devices with authostereoscopic screens are already reaching the market. Furthermore, as evident from user studies, users expect that Mobile 3DTV technology will be used in conjunction with other emerging technologies, including social networks, augmented reality, location aware services, etc. Mobile 3DTV project was focused principally at designing a proof of content system for streaming of stereo video to mobile devices. However, the results generated at this project are expected to have an implication onto possible integration of Mobile 3DTV with other technologies, as it provides a solid standardized foundation upon which such integrations can be developed.
Societal Implication of Project Results

User Experience

As stated earlier, the success of any emerging technology is measured by the rate of its acceptance by general public. The quality of user experience is crucial in this sense. This perception of quality is inherently subjective, and as such hard to gauge. Yet, it is of the grave importance for any successful mass market application to be able to establish the connection between subjective impressions that users might have and actual underlying technological parameters which determine this perceived quality and user experience, and which can be then altered accordingly in order to enhance this experience. As a direct result of MOBILE3DTV project we now have a much better understanding of what drives the user quality of experience. We have successfully developed new methodological approaches which can be used to evaluate this QoE and to enhance it for greater user satisfaction.

We have been able to map the subjective user impressions to the relevant features of the entire end-to-end Mobile 3DTV system, especially, the ones related to stereo-video compression and transmission. The methodological framework established in this project is expected to contribute to a baseline standard for the evaluation of QoE in multimedia.

Assessment and Enhancement of Quality of 3D Video

The knowledge gathered about the user experience has permitted us to draw several important conclusions about which aspects of the underlying end-to-end Mobile 3DTV system make the most significant impact on the acceptance of 3DTV content and user satisfaction. Based on this knowledge, we have been able to design a metric to objectively assess the expected quality. Furthermore, a set of processing tools has been developed to enhance this quality. This set of tools is expected to be adopted by the industry. Among other things, the framework for optical measurement of portable auto-stereoscopic displays and mitigation of display-specific artifacts is especially promising. The data gathered from subjective tests has clearly marked this part of the Mobile 3DTV system as potentially detrimental for the user satisfaction, as low quality screens can render useless any quality optimization done in previous parts of 3DTV content creation and broadcasting chain. As stated earlier perceived quality matters. Users faced with unimpressive performance of one particular device can easily be discouraged from the whole concept, alienating them from the entire emerging technology. This is of the especially great importance at the current stage of development of Mobile 3D video technology, as it is a relatively young technology that is yet to assert its place on the market. There are no clear market leaders yet. Large number of device manufacturers is entering the market at the same time, producing devices with screens based on different technologies. As there are no established standards the quality of these technological solutions varies widely. The number of similar products is expected to increase several fold in near future. Therefore, the need for a tool which can be used to objectively assess the quality of individual screens is evident, and we hope that our framework will meet with wide acceptance, thus permitting the establishment of relevant standards in this section of the market.

Mobile 3DTV in Relation to Existing 3D Video Technologies

Mobile 3DTV, as an emerging technology, is embedded in wider technological and socio-economic landscape which is expected to have significant economic potential. In order to reach its full market potential it would need to attract a significant amount of investment by several sectors of industry. However, it is the most closely related to 3D video technology in wider sense, mobile TV and mobile telecommunications. Out of these three, arguably only
mobile telecommunications are an established technology. The other two are also just in initial stages of their mass market expansion.

3D Video technology has already been established in several areas, most importantly in form of 3D motion pictures in cinemas and as first-generation 3D TV content in home entertainment. Over the last couple years several important sectors of industry, namely, Movie industry, TV content providers, broadcasting companies, telecommunication companies etc. have made significant investments in 3D video technology, covering all aspects of the production chain, from capture, encoding and transmission to final delivery to end users. One of the most important questions at this point is which part of these investments in technology can be re-purposed for Mobile 3D video, and what is the amount of fresh investments will need to be made. The amount of investment, of course, being important for assessing the probable profitability of the platform. In this respect, the compatibility of Mobile 3DTV technology with already established 3D video technologies is crucial. In this sense the choice of optimal 3D video formats in particular is significant.

The data gathered during Mobile 3DTV project clearly shows the advantages of two-channel stereo encoded by MVC versus other formats and coding methods. Thus, we conclude that the existing MVC video format is a close approximation of optimal Mobile 3DTV format. Furthermore, our subjective quality analysis data clearly shows that there is no significant loss of quality if existing 3D video content is directly re-purposed for Mobile 3DTV platform. This conclusion has wider industrial repercussions as MVC the format of choice for Blu-ray 3D. In shot, this means that no significant change in technology for capture and encoding of 3D video especially for 3DTV needs to be made. Thus, no significant investment in this part of the process needs to be made, as existing content can be almost directly used. Furthermore, CSV encoded by MVC does not require subsequent rendering at the receiver end. This implies less processing power required to display content, which in turn implies reduction in manufacturing cost of actual devices. Lower production costs and simpler manufacturing procedures have implications on the final product price, influencing the profit margins for the industry and expanding the possible consumer base.

Lesser requirements in processing powers also imply lesser power demands from the device, extending battery life and operation time of the device. Battery life has been demonstrated to be one of the most significant factors determining the consumer acceptance of emerging mobile technologies. The MOBILE3DTV project contributed to determining the optimality of 3D video formats for their mobile use.

End-to-end Mobile 3DTV System

We have developed a proof-of-concept end-to-end Mobile 3DTV system. During the development of this system we had to address several open questions, such as optima data format, optima coding algorithms, optimal transmission over DVB-H, quality at the receiver side. This prototype system integrates the results from all other parts of MOBILE3DTV project in one coherent and functional chain. The key contribution of the proposed system is that it provides an optimal blueprint which any future real-life system will have to follow in a great degree. Although developed around DVB-H delivery channel, the proposed system is versatile, and can be easily modified and adapted to other delivery platforms such as mobile internet based services.

Mobile 3DTV as a Content Delivery Platform

The prototype system developed during this project represents a media delivery platform. Mobile video streaming services are a rapidly growing trend in modern mobile
communications. Recent forecasts have claimed that about 46% of the global IP traffic will be devoted to streaming video in 2014. Additional 27% will come from file sharing. The number of users accessing Internet from mobile devices is constantly rising and is predicted to overtake desktop access in year 2014. According to Cisco Visual Networking Index 2010 the Advanced Internet video traffic, a category which includes 3D and HD video content, will increase 23-fold between 2009 and 2014.

MOBILE3DTV user studies and the final usability tests have shown a clear interest of users in a device capable of delivering 3D video content on a mobile platform. Furthermore, we were able to identify several key issues regarding future user habits. These issues are related to situations in which 3D video content is going to be consumed, and type, length and genre of the content, as well as the form factor of the end delivery device.

Participants of the user studies have indicated that the main situation in which they would use a Mobile 3DTV device was to fill small amounts of idle time, such as during waiting or in public transportation, during travel, etc. This follows the general pattern of consumption of mobile video content, and other entertainment applications on mobile devices.

When it comes to the choice of genre of content there is no general consensus among users: movies, animated content, sport events and documentaries all mentioned as preferred types of content by large percentage of test participants. However, there seems to be only a limited interest in seeing news in 3D. The main purpose of Mobile 3DTV devices will thus be first to entertain, then to educate and finally to potentially inform.

According to experience with other type of streaming content, several conclusions can be made. The public expects that this content is delivered on-demand, rather than through broadcasting model. Data transfer fees need to be regulated on flat-rate basis while content itself should be either free or on pay-per-view basis.

The choice of time format, i.e. the length of single unit of content depends on the viewing context, the situation and the personal preferences of the users. In this sense, there are two diverging trends, one favoring movies and TV shows and another based around shorter video formats. Full length movies, and TV shows in 3D constitute the high end of video media offer. As demonstrated by the MOBILE3DTV project, the already existing 3D video content can be directly re-purposed for deployment on mobile devices.

However, the mobile aspect of devices also favours shorter forms as well, movie trailers and inserts, TV show snippets, shot animation sequences, news bulletins and short sporting event reports. Short video forms are also indicative of other possible area of application of such devices: the user-generated content. Video as user-generated concept is already an established concept. Youtube, the largest video sharing website streams over 2 billion video clips daily, and will have according to Citigroup estimates revenue of 1.2 billion dollars in 2011. Other major players in this category include Vimeo, and DailyMotion. Streamed Youtube videos account for a large percentage of data traffic of current Smart phone platforms. It is also a major driving force behind the adaptation of HTML5 web standard. User-generated 3D audio-video content is expected to follow this trend. Next important contribution from research community will be in area of editing tools. A set of efficient easy-to-use tools would enable the broad range of general public to unleash their creativity, developing perhaps new forms of expression. Especially, forms where 3D contributes significantly to expressive power of media, and is more than a novelty feature.

Device form factor is also one of the important factors in the future applicability of Mobile 3DTV. According to our studies, there is no universally preferred device form. Instead, several contradictory trends can be observed. Above all, many users have expressed a desire
to have 3DTV content integrated with outer services, and perhaps embedded within already existing devices. Confluence with mobile phones and other handheld systems is thus to be expected. When it comes to optimal screen size, two opposing opinions prevail. Some users clearly favor smaller devices which can easily fit into pocket. However, another group is more inclined towards larger screen sizes, citing better immersing effect achieved by 3D on such screens. One can extrapolate that the market for such devices will bifurcate in two distinct niches, one occupied by smaller devices, perhaps 3D-enabled smart phones, and other by bigger tablet like devices. In any case the system proposed in this project could form a flexible backbone which could facilitate content delivery in any of these situations.

Mobile 3D devices on the confluence of different emerging technologies

Each emerging technology is deeply rooted in existing technological concepts. The advent of Mobile 3DTV technology is based on the developments in information and communication technologies in recent decades, especially digital telecommunications and computationally powerful mobile devices. Several other emerging technologies are propelled by the same underlying socio-technological processes.

Most emerging technologies begin their commercial life-cycles in some form of non-mobile, i.e. desktop based setting. In most cases this is due to the fact that mobility poses much stricter design and performance constraints. However, in recent years it becomes evident that, transition to mobile platforms is an unavoidable path for any emerging technology.

Mobile 3DTV technology shares that same path with several other emerging technologies. As it shares the same platform with them, it is expected that mobile 3D technology will interact and cross-pollinate with these technologies.

The user studies that have been conducted within the project have indicated that the great majority of participants are ready to employ the proposed system in the envisioned scenario, i.e. as a device for streaming video content. However, in their feedback the participants have indicated the clear desire to see video 3D content used in conjunction with other possible applications. The most often mentioned ones include:

1. Teleconferencing applications,
2. Video games,
3. Access to social networks,
4. Ability to generate 3D content,
5. Advertisements.

We believe that the Mobile 3DTV technology can support and co-exist with other emerging technologies. Possible applications of Mobile 3DTV devices include the following:

1. Entertainment and leisure activities- streaming of 3D TV content, user-generated content, video games, social networks, adult entertainment.
2. Professional applications - teleconferencing, presentations, product advertisement, applications in design and architecture, etc.
3. Tourism.
4. Education and scientific research.

Main dissemination activities and exploitation of results

The MOBILE3DTV project actively disseminated the results of its research. The primary source of information is the project website at www.mobile3dtv.eu. It contains the project
summary and description of the consortium, and sections about technology and results. The latter includes public summaries, test data bases and downloadable software. All public deliverables are published there as well. The full collection of scientific publications generated during the project lifetime is available with free access.

Most excellent research results were reported at prestigious scientific conferences, such as the SPIE/IS&T Electronic Imaging Symposium, IEEE ICIP, 3DTV Conference, and published in peer-reviewed journals, such as the International Journal of Mobile Human Computer Interaction (IJMHCI), Journal of the Society of Information Display (JSID), IET Signal Processing, Advances in Multimedia, EURASIP Journal on Advances in Signal Processing. An invited paper in the Proceedings of the IEEE overviewed most of the project results at that time. Two conference papers on project results got the price of best paper award.

The project actively exhibited and demonstrated the technologies being developed. Project booths were set at the ICT Event 2008 and 2010, NEM Summit 2008 and 2009, Mobile World Congress 2009 and 2010, 3D Media workshop 2009 and 3DTV conference 2010, Dimension 3 Expo 2009. Two special sessions on 3D Media for Mobile Devices were organized jointly with ETRI, Korea at the Electronic Imaging Symposium.

Two proposals for standardization of user-centered QoE methods were submitted to ITU-R WG12 and few more results are bearing standardization potential for standardization bodies such as ICDM and MPEG. One patent application was prepared on a topic related with optimized visualization on auto-stereoscopic displays. Software packages on simulation of stereoscopic artefacts and simulation of DVB-H channel conditions were issued under general public license and few more software packages are copyrighted for potential licensing.