Mobile 3D television: Development of core technological elements and user-centered evaluation methods toward an optimized system

Atanas Gotchev*[a], Aljoscha Smolic[b], Satu Jumisko-Pyykkö[c],
Dominik Strohmeier[d], Gozde Bozdagi Akar[e], Philipp Merkle[b], Nikolai Daskalov[f]

[a]Dept. of Signal Processing, Tampere University of Technology,
P. O. Box 553, FI-33101 Tampere, Finland;

[b]Dept. of Image Processing, Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany;

[c]Unit of Human-Centered Technology, Tampere University of Technology,
P.O.Box 589, 33101 Tampere, Finland;

[d]Institute for Media Technology, Ilmenau University of Technology,
PF 10 05 65, 98684 Ilmenau, Germany;

[e]Multimedia Research Group, Middle East Technical University, Balgat, Ankara 06531 Turkey;

[f]Multi-Media Solutions Ltd., Tyntiava 15, Sofia, Bulgaria

ABSTRACT

A European consortium of six partners has been developing core technological components of a mobile 3D television system over DVB-H channel. In this overview paper, we present our current results on developing optimal methods for stereo-video content creation, coding and transmission and emphasize their significance for the power-constrained mobile platform, equipped with auto-stereoscopic display. We address the user requirements by applying modern user-centered approaches taking into account different user groups and usage contexts in contrast to the laboratory assessment methods which, though standardized, offer limited applicability to real applications. To this end, we have been aiming at developing a methodological framework for the whole system development process. One of our goals has been to further develop the user-centered approach towards experienced quality of critical system components. In this paper, we classify different research methods and technological solutions analyzing their pros and constraints. Based on this analysis we present the user-centered methodological framework being used throughout the whole development process of the system and aimed at achieving the best performance and quality appealing to the end user.

Keywords: stereo video, DVB-H, H.264, MPE-FEC, OMAP, user-centered design

1. INTRODUCTION

3DTV and mobile TV are two emerging technologies in the area of audio-video entertainment and multimedia. The former is expected to bring the realistic visualization of moving 3D scenes and to effectively replace the HDTV at users' homes. The latter is expected to appeal to mobile users by brining the most dynamic and fresh video content to 'cool' handheld gadgets.

The general concept of 3DTV assumes the content is to be viewed on large displays and simultaneously by multiple users [1]. Within this concept, glasses-enabled stereoscopic display technologies have been competing with glasses-free autostereoscopic displays [2]. Consumer electronics manufactures, such as Philips, Panasonic, Sony and Samsung have been supporting either of these display technologies. Experimental commercial broadcasts have started in Japan. On the research side, various aspects of 3DTV content creation, coding, delivery and system integration have been addressed by numerous projects and standardization activities [3][4][5].

As of mobile TV, standardization and legislation activities have lead to creation of similar yet continent- or country-specific standards [6]. In Europe, DVB-H has been identified as the single European standard for mobile TV [7]. First commercial DVB-H TV broadcasts have started in several European countries with a number of compatible handheld

terminals by Nokia, Samsung, Motorola and LG. In Korea, the counterpart T-DMB has already attracted several millions of users [8].

The current developments position 3DTV and mobile TV as two rather diverging technologies as the former relies on big screens and realistic visualization and the latter relies on portable displays. Would they converge to some new technology which would allow taking 3DTV into one's pocket? The symbiosis is rather attractive. 3D would benefit from being introduced also by the more dynamic and novel technology-receptive mobile tech market. Mobile TV and the corresponding broadcasting standards would benefit from the reach content and new business models offered by 3D. In order to make this happens, the research challenge is to adapt, modify and advance the 3D video technology, originally targeted for large screen experience, for the small displays of handhelds. This challenge has been addressed by the several projects, as the Korean 3D T-DMB [9], and the European projects 3DPhone [10] and Mobile3DTV [11]. The latter one specifically addresses the mobile 3DTV delivery over DVB-H system. This paper overviews the approach taken by the project partners and summarizes the first results obtained. A case is made of the user issue, as little is known about the user experience of 3D video content visualized on a portable screen.

The paper is structured as follows. Section 2 briefly presents our mobile 3DTV concept and overviews the system as a whole. Section 3 presents the content formats and coding approaches for 3D video and the modifications and adaptations specific for mobile 3DTV. Section 4 addresses the DVB-H channel. It describes the streaming application developed for it and gives insight on the DVB-H specific error protection schemes and their relevance for 3D video content. Section 5 briefly describes the handheld technology demonstrator and its capabilities for implementing the decoding, decapsulation and view-rendering on an auto-stereoscopic display. Section 6 deals with the user-centered design of subjective tests of critical parts of the system. Section 7 concludes the recent status and points to the future research.

2. SYSTEM OVERVIEW

Mobile 3DTV system is conceptualized by Fig. 1, i.e. stereo video content is captured, effectively encoded, and then robustly transmitted over DVB-H to be received, decoded and played by a DVB-H enabled handheld.

Capture ▸ Coding ▸ Resilience ▸ Transmission ▸ Decoding ▸ Visual optimization ▸ Display
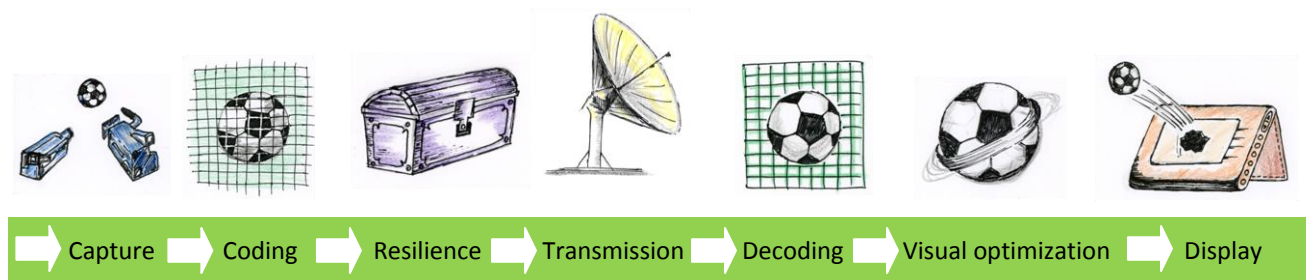
Fig. 1. Mobile 3DTV system.

An immediate remark is needed here. In our concept, we consider stereo video to be displayed by portable auto-stereoscopic display enabling two views only. In the auto-stereoscopic technology, the number of views can get higher for the price of decreasing the spatial resolution [2]. In the case of portable (i.e. 3'' to 4'') displays it would be unnecessarily complex to require more than two views for what is to be likely a single-user device. Multiple views would require also additional power for decoding and/or rendering the extra views, instead, for example, enabling higher frame rate. Therefore, we have adopted the stereo video framework as the sound compromise, which should provide comfortable 3D experience to the user with acceptable spatial resolution and frame rate.

At the stage of 3D content creation and coding, currently there is no single and generally adopted representation format for stereo video, taking specific mobile channel conditions into account. Most natural is to have two-channel stereo video. Capture of such video by synchronized cameras is relatively easy and the coding can be done efficiently, e.g. by the techniques of the emerging multi-view coding (MVC) amendment of the H.264 AVC standard [4]. There are mainly two problems with two-channel video targeted for mobile platforms. The first issue is that the video is to be displayed on a small-size display with relatively low spatial resolution. Usually, 3DTV content is considered to be captured by HD

cameras. Simple rescaling of HD stereo content would, in the general case, reduce the disparity between views making the perception of depth somehow limited. Hence, the baseline should be maintained properly e.g. by multi-camera capture and synthesis of views specifically suitable for the handheld display resolution. The second issue is to how to encode the stereo video. While MVC seems the direct choice, its syntax high-profile requirements might unnecessarily complicate the algorithms meant to be working at the power-constrained handheld device. From another side, simple stereo frame interleaving coding algorithms might not be appropriate for legacy devices.

An alternative to two-channel stereo video is offered by the representation known as single-view video plus depth format, already standardized under the notion of MPEG-C, Part 3 [4]. Such representation leads to good compressibility, as the depth adds less than 20% to the bit-budget of a single video channel [12] but requires additional techniques for depth estimation at the content creation side and depth image based rendering at the receiving side [14]. Problems with occlusions might appear as well.

A new concept of mixed spatial resolution is expected to cope with the problems of fast rendering, efficient compression, and backward compatibility [4]. All three above-mentioned data representations and coding approaches are being analyzed, compared and optimized within the scope of the MOBILE3DTV project. More details on these approaches and preliminary comparative results are given in Section 3.

In our concept, DVB-H is considered to be the broadcast media of future mobile 3DTV. The standard has been a very successful development from the initial idea through comprehensive research and development to commercial services. The key issue is its flexibility. It is not just a tiny TV channel but rather general and powerful data broadcast technology. So far, DVB-H has been extensively studied for its capability to provide error protection related with the importance of the content to be transmitted (i.e. unequal error protection - UEP) [13]. Compression bit-rates and error-protection rates for conventional video have been tuned so-far in a way to meet the requirement of the decoding platform. With the invention of new media-reach platforms capable of decoding higher bit rates and supporting higher display spatial resolutions, the broadcast of stereo video over DVB-H seems quite feasible. What should be especially addressed is the specific error protection of stereo-video content over such a channel. It might be well protected by the current tools but it might also turn that it needs novel and more comprehensive UEP schemes. Section 4 shed more light on our first experiments with streaming stereo video over DVB-H and the problems with error resilience one might face in harsh mobile channel conditions.

Problems, such as error concealment, image and video deblocking and sharpening have to be addressed on the receiver side by efficient and highly-optimized algorithms so to obtain a superior visual quality at an acceptable computational effort. These shall be run on a handheld built on a platform with multimedia-reach capabilities and equipped with auto-stereoscopic display. Such platforms and displays have already matured quite a bit as also seen in our advance in developing the mobile technology demonstrator (Section 5).

As seen from the above brief overview, the components of the future technology are there. However, they should be put and optimized to work together. We have been working toward developing a new holistic evaluation method for experienced multimedia quality evaluation so to be able to effectively test and optimize critical parts of the system (Section 6).

3. CONTENT FORMATS AND CODING APPROACHES FOR MOBILE 3DTV

The ultimate goal of research in this area is to develop the best possible 3D video representation and coding for the specific application of transmission over DVB-H and mobile terminals. For that, different alternative approaches are developed, implemented, optimized and compared. This includes approaches beyond the current state-of-the-art as defined in available standards i.e. mixed resolution stereo video coding.

3.1 Considered approaches

As a starting point and reference for the further research, available coding standards for 3D video have been optimized and evaluated for the specific conditions in the Mobile3DTV project. Specifically, the optimizations are targeted for the demonstrator device as to be used in the project, i.e. how to use the different 3D video codecs in this specific context.

The following 3D video formats and codecs were evaluated [15]: H.264/AVC simulcast [16], H.264/AVC stereo SEI message [16], MVC [17], MPEG-C Part 3 using H.264/AVC for both video and depth [18].

The first three approaches use stereo video data only (V+V), while the last approach is based on the video plus depth format (V+D). This is illustrated in Fig. 2. A stereo image pair shows the same scene from slightly different viewpoints corresponding to the human eye positions. Such a stereo image pair can be directly visualized on a 3D display. In a V+D representation a video signal and a per pixel depth map is transmitted to the user. From the video and depth information, a stereo pair can be rendered by 3D warping at the decoder. Per pixel depth data as illustrated in Fig. 2 can be regarded as a monochromatic, luminance-only video signal. This can then be processed by any state-of-the-art video codec.

The ability to generate the stereo pair from V+D at the decoder is an extended functionality compared to V+V. It means that the stereo impression can be adjusted and customized after transmission. However, this advantage is paid by increased complexity for both sender side and receiver side. View synthesis has to be performed after decoding to generate the 2nd view of the stereo pair. Before encoding, the depth data have to be generated. This is usually done by depth/disparity estimation from a captured stereo pair. Such algorithms can be highly complex and are still error prone.

The concept of V+D is highly interesting due to the backward compatibility to classical 2D video and extended functionality. Moreover it is possible to use available video codecs. It is only necessary to specify high-level syntax that allows a decoder to interpret two incoming video streams correctly as color and depth. Therefore MPEG specified a corresponding container format "ISO/IEC 23002-3 Representation of Auxiliary Video and Supplemental Information", also known as MPEG-C Part 3, for video plus depth data [18] in early 2007. This standard already enables 3D video based on video plus depth. It was reported that in some cases such depth data can be efficiently compressed at 10-20% of the bit rate which is necessary to encode the color video [14][12], while still providing good quality of rendered views. A goal of this study was to prove these statements for a variety of different test data.

Fig. 2. Stereo image pair and depth map corresponding to the left image.

3.2 Experiments

Evaluation was done by simulations. Same test data were used in all experiments. Professional stereo sequences in 16:9 format (as the display in the demonstrator) kindly provided by KUK Filmproduktion were used. In total the 4 sequences *Horse*, *Car*, *Hands*, and *Snail* spanning a range of different types of content and complexity were used formatted to 480x270 (cf. Figs 2 and 3). The material was coded at different bitrates using optimum settings for each of those codecs under investigation. Quality was evaluated by means of PSNR versus bitrate and informal subjective expert viewing. For the case of V+V, PSNR was calculated over both (left and right) luminance channels. For the case V+D, PSNR was calculated for the luminance component of the rendered channel.

As for any type of video coding, the same amount of raw input data may lead to very different RD-performance. The required bitrate for achieving acceptable quality strongly depends on the properties of the sequence content, especially temporal variation and complexity of the scene. For instance, the very complex *Hands* sequence requires more than double bitrate for same PSNR compared to *Car*. The very simple *Snail* sequence on the other hand consumes less than 10% of this bitrate for same quality.

Fig. 3. Examples of test data used in the experiments.

3.3 Results for V+V coding

In this case coding operates on two related video sequences. H.264/AVC simulcast means simple independent coding of the two videos without exploiting inter-view redundancy. The other two approaches, namely H.264/AVC stereo SEI and MVC perform inter-view prediction between the stereo views in different ways. Significant coding gains can be achieved with hierarchical B pictures for temporal prediction. The gain from using hierarchical B pictures differs largely for individual sequences, depending on factors like scene content complexity and temporal variation. On the other hand hierarchical B pictures also mean increased complexity and memory requirements. It remains to be studied how far this can be implemented on a mobile terminal.

The coding gain from inter-view prediction varies largely. It leads to a significant reduction of bitrate for some sequences. However, in some cases the gain is negligible. In our experiments we achieved up to 35% bitrate savings from inter-view prediction compared to stereo simulcast. Inter-view prediction whether performed as SEI or MVC does not add substantial complexity. However, a standard conform implementation of MVC requires that the decoder supports the H.264/AVC High Profile since MVC extends that. A standard conform implementation of the Stereo SEI Message requires that the decoder supports the H.264/AVC interlaced tools. Fig. 4 shows an example of the results.
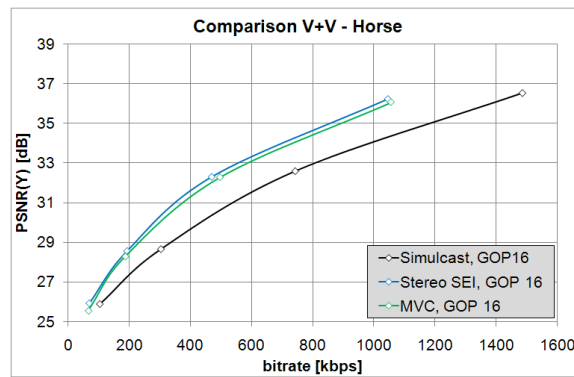

Fig. 4. RD-performance for V+V coding, gain from inter-view prediction over simulcast.

3.4 Results for V+D coding

MPEG-C Part 3 is suitable for encoding of video plus depth data. Basically video and depth are coded separately using e.g. H.264/AVC. Different bitrate ratios (and thereby qualities) for video and depth can be adjusted. The influence of such combinations on the RD-performance of the rendered right view has been evaluated. Fig. 5 shows the results for all 16 possible combinations of color and depth quality in our experiments using 4 qualities (i.e. QP settings, QP = 24, 30, 36, 42) for each. The curves combine points of constant color quality (C24, C30, …) and points of constant depth quality (D24, D30, …). Apparently curves of constant color quality are steeper. In most cases increasing depth bitrate has a stronger influence on overall quality than increasing color bitrate. At a certain bitrate the QP for depth should be chosen lower (i.e. better quality) than the QP for color to achieve best overall results, e.g. C30, D24. It can be concluded that good depth quality is essential for good overall quality, measured in terms of PSNR of the rendered channel. The numerical bitrate ratio between color and depth may vary largely depending on the sequence. We have found ratios between 1:1 and 6:1. In most cases for best overall quality a substantial portion of the bitrate has to be spent for depth, which is in contrast to previously reported results. Note, however, that our current experiments employed PSNR of the rendered channel as an objective quality metrics. Our previous study [12] has shown that PSNR is quite conservative and over-estimates the influence of depth map to the quality of the rendered channel. A quality metric, better adapted to the peculiarities of the human visual system is needed in order to provide numbers closer to what the mean observer will judge. In [12], we have used VSSIM [19], which while found better than PSNR was still not close enough to the mean opinion scored obtained by small-scale subjective tests.
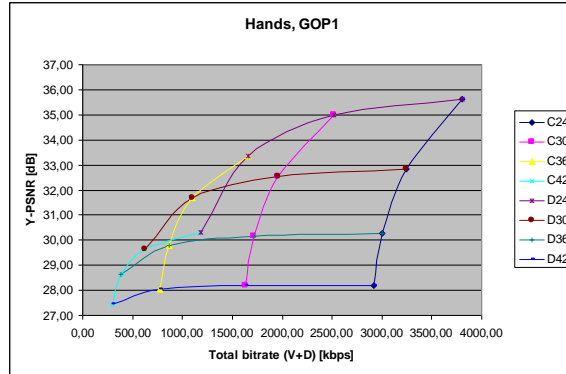
Fig. 5. PSNR vs. bitrate for rendered right view with different combinations of quality for color and depth.

3.5 Future work

A detailed comparison of V+V and V+D approaches is still to be done. This will include formal subjective tests and novel objective quality metrics especially tailored for stereo video. Further, the concept of mixed resolution stereo video coding will be studied in detail, as an extension of available 3D video formats. It is derived from the so called binocular suppression theory [20]. Subjective test have shown that to some degree, if one of the images of a stereo pair is low-pass filtered, the perceived overall quality of the stereo video will be dominated by the higher quality image, i.e. the perceived quality will be as if both images were not low-pass filtered. Based on that effect, mixed resolution stereo video coding can be derived. Instead of coding the right image in full resolution, it can be downsampled to half or quarter resolution. In theory this should give similar overall subjective stereo video quality, while significantly reducing the bitrate. Taking the bitrate for the left view as given for 2D video, the 3D video functionality could be added by an overhead of 25-30% for coding the right view at quarter resolution.

4. CHANNEL CONSIDERATIONS

4.1 DVB-H characteristics and DVB-H specific error protection

Wireless networks are often error prone due to factors such as multipath fading and interferences. 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 quality of service (QoS), a number of technologies have been proposed targeting different layers of the networks. DVB-H uses *forward error correction* (FEC) for error protection and comes with an optional FEC tool at the link-layer. This FEC uses Reed-Solomon (RS) codes encapsulated into *Multi-protocol* sections (MPE-FEC). The MPE-FEC was also introduced to provide additional robustness required for hand-held mobile terminals. MPE-FEC improves the carrier-to noise (C/N) and Doppler performance in the DVB-H channel while also providing improved tolerance of impulse interference. However, MPE-FEC might fail in the presence of very erroneous conditions.

In a DVB-H system, as in Fig. 6, the audiovisual content is passed to the link layer in Internet Protocol (IP) datagrams. The datagrams are encapsulated column-wise into an MPE-FEC frame, the size of which can be selected in a flexible manner. The encoding of the MPE-FEC frame using a Reed-Solomon (RS) code is performed row-wise, which results in an interleaving scheme referred to as virtual time interleaving [21]. By varying the amount of application data columns and RS data columns, different code rates can be achieved. For transmission, the MPE-FEC frame is divided into sections. An IP datagram forms the payload of an MPE section and an RS redundancy column forms the payload of an MPE-FEC section. The MPE sections are transmitted first, followed by the MPE-FEC sections. Both of them are transmitted in MPEG-2 transport stream (TS) format [21].
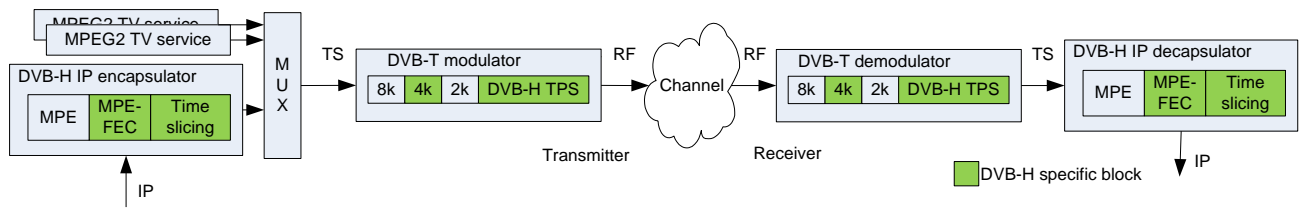
Fig. 6. DVB-H link and physical layers

Time-slicing is applied to enable power saving, so that one MPE-FEC frame is transmitted in one time-slice burst. However, if one burst is lost, the video stream is interrupted until the next burst is received. The degradation in video quality due to these loses depends on the amount of IP data transmitted in the burst and the data rate of the video stream. This cycle is different for the transmission of monoscopic and stereoscopic video because of the data rate. In addition, there are several link layer parameters such as the frame size, transmission bit rate and off-time between bursts and physical layer parameters such as, code rate, guard interval length, OFDM mode that effect the video bit rate thus the visual quality.

4.2 Channel measurements and simulations

With a large set of options as stated in the previous subsections, simulations are usually the most efficient way to find the optimal parameter combinations for robust transmission.

The transmission simulation environment is based on the open-source Linux-based tools, Fatcaps and Decaps[22]-[24]. When a measured or simulated packet error pattern is applied to the transport stream, the system forms an end-to-end simulation of the DVB-H transmission chain. The simulator can be used for both online and offline simulations. In the online case, live IP streams are encapsulated in transport streams and sent back to IP network by decapsulator. In offline simulations, the phases can be separated from each other and it is not necessary to use live IP streams as input and output. For instance, one can first generate a transport stream that is then passed through various different channel conditions.

The building blocks of our system are given in Fig. 7. The input videos are first compressed with a 3D encoder. Resulting NALUs are fed to the 3D video streamer. The streamer encapsulates the NALUs into Real Time Transport Protocol (RTP), User Datagram Protocol (UDP) and finally Internet Protocol (IP) datagrams [25]. The resulting IP datagrams are encapsulated in the DVB-H link layer where the Multi Protocol Encapsulation Forward Error Correction (MPE-FEC) and time slicing occurs. The link layer output MPEG-2 Transport Stream (TS) packets are passed to physical layer where the transmission signal is generated with a DVB-T modulator. After the transmission over a wireless channel, the receiver receives distorted signal and possibly erroneous TS packets are generated by the DVB-T demodulator. Error-correction attempts are done in the link layer by the MPE-FEC functionality and the TS packets are decapsulated into IP datagrams. IP datagrams are handled in the 3D video streamer client and resulting NAL units are decoded with the 3D video decoder to generate right and left views. Finally, these views are put into an appropriate format to be displayed as 3D in the displayer.
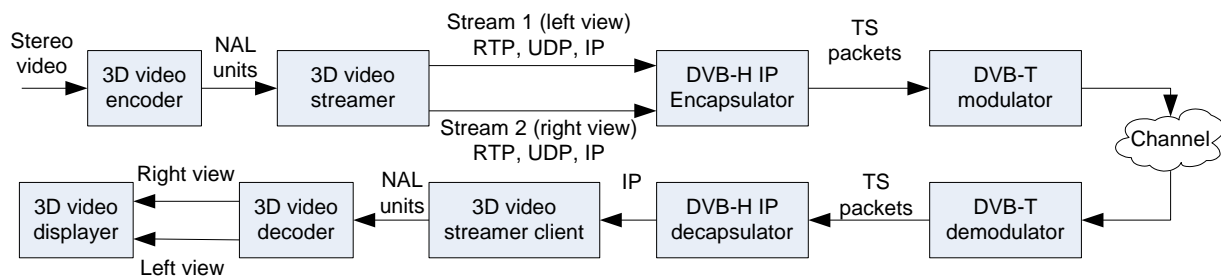

Fig. 7. Stereo video streaming over DVB-H system

For simulating the physical transmission channel, we have a MATLAB/Simulink tool that models the DVB-T/H modulation and demodulation processes. The channel is modeled as multipath Rayleigh fading channel with additive white Gaussian noise. Various commonly used channel models have been predefined in the simulator [26]. This tool can be used to collect reception statistics such as Bit Error Rate (BER), TS-Packet Error Rate (TS-PER), Average Burst

Error Length (ABEL) and Variance in Burst Error Length (VBEL) as well as to record packet error traces for video transmission simulations. The application layer performance is approximated by mapping the transport stream error trace to an IP packet error trace, under the assumption that all IP packets are of the same size which is also the number of rows in the MPE-FEC frame. These assumptions simplify simulation of the FEC decoding because both IP packets and RS data sections map directly to columns of the MPE-FEC frame and the so-called column erasure decoding method can be used. MPE-FEC code rate 3/4 was used in the simulation (191 application data columns, 64 RS columns).

To validate the performance of the simulation setup, we performed several transmission experiments where an MVC coded test sequence was transmitted over a simulated DVB-H channel. We simulated two transmission modes: 1) 16-QAM constellation, convolutional code rate of 2/3 and MPE-FEC rate of 3/4 over *typical urban channel* with 6 taps and 2) 16-QAM constellation, convolutional code rate of 2/3 and MPE-FEC rate of 3/4 over *residential outdoor channel*. The test sequence was encapsulated in a transport stream by putting each view in its own MPE stream.

We evaluated the impact of MPE-FEC by comparing two scenarios when extracting the H.264/MVC video from the transport stream: 1) Client receiving the FEC codes within the MPE-FEC frame, 2) Client receiving only the data part of the MPE-FEC frame where power consumption is reduced. This simulation process was repeated 50 times for each channel condition in order to obtain figures corresponding to the average behavior of the channel.

For simulations, the *Hands* video sequence was used. In Fig. 8, "No loss" corresponds to the scenario where there is no loss in the transmission, "FEC" corresponds to the scenario where MPE-FEC is decoded in the case of lossy transmission and "No FEC" corresponds to the scenario where the receiver discards the FEC bytes. The preliminary results show a clear improvement of quality when the received MPE-FEC data is used especially in low channel SNR cases. Although MPE-FEC provides the much needed data robustness for 3D video transmission in wireless channels, under very erroneous conditions it may fail. Using a-priori knowledge of the transmitted media and tuning the way MPE-FEC is applied across the media datagrams can provide better robustness. Also experimental results should be extended for different 3D representations and different channel parameters to find the optimum parameter set.
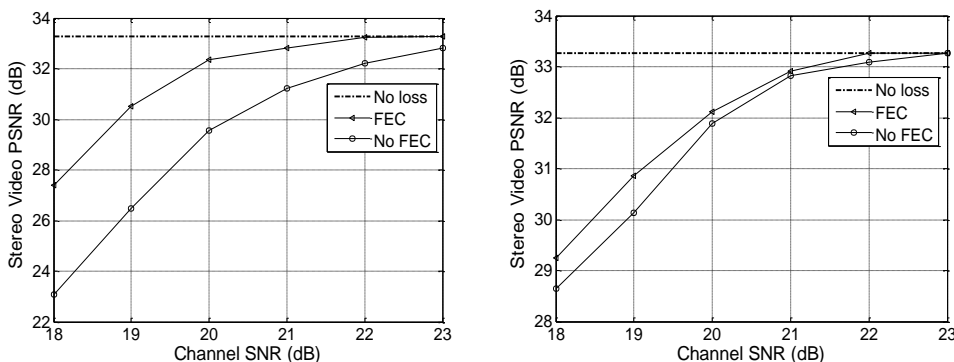

Fig. 8. Channel simulation results. Hand stereo sequence jointly coded with stereo GOP size of 16; 16-QAM constellation, convolutional code rage of 2/3, MPE-FEC rate of 3/4. Left: typical urban channel; right: residential outdoor channel.

5. TERMINAL DEVICE

5.1 Hardware platform

A number of objectives were targeted when the development platform for decoding and display of stereoscopic content was selected. We targeted platform being in the roadmap of Tier 1 mobile phone vendors. It would provide a means for connecting auto-stereoscopic display and DVB-H receiver as well as sufficient processing power to execute DVB-H stack, GUI and decoding. Decoding seemed to be the first decisive issue. In our further development, we would implement a custom type H.264 decoder as a base of multi-view or view+depth type of stereo video encoder. Therefore, we targeted platforms having sufficient processing power to decode simultaneously at least two H.264 BP simulcast coded streams (QWVGA@25fps) or H.264 BP multi-view coded stream (two views) plus up-sampling in vertical

direction by factor of two. Practical issues such as type of running high-level OS (Linux, Symbian or Windows Mobile), and debugging applications were also taken into account.

The development platform selected was OMAP 3430 manufactured by Texas Instruments. We refer to its technical specifications in the TI bulletin [27] and relate some of them to our specific application. Demanding applications such as stereo-video decoding and playing, can be parallelized employing different cores with sufficient speed provided by the high clock rate. Specifically, the OMAP 3430 features a superscalar ARM Cortex-A8 RISC core, an image video and audio (IVA) accelerator enabling multi-standard (MPEG4, WMV9, RealVideo, H263, H264) encoding/decoding at D1 (720x480 pixels) 30 fps and an integrated image signal processor (ISP). Graphics accelerator with OpenGL CE support is integrated as well. The available XGA display support allows for choosing high-resolution LCD. Advanced 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 system architecture of OMAP 3430 is given in Fig. 9 (Figure provided by TI).

5.2 Decoding tools

As a first step, we have implemented decoding tools to process stereo video encoded in different formats (cf. Section 3). The OMAP 3430 SDP is bundled with video decoder test application, capable to decode MPEG4, H.264, MPEG2 and WM9 video streams with no audio synchronisation. The platform, denoted as OpenMAX (OMX), implements optimized decoder running on the DSP and using some vendor-specific hardware accelerators [28]. The block diagram of the decoder application and its components is shown in Fig.10. It allows decoding video-streams read from files and rendering on the LCD display.
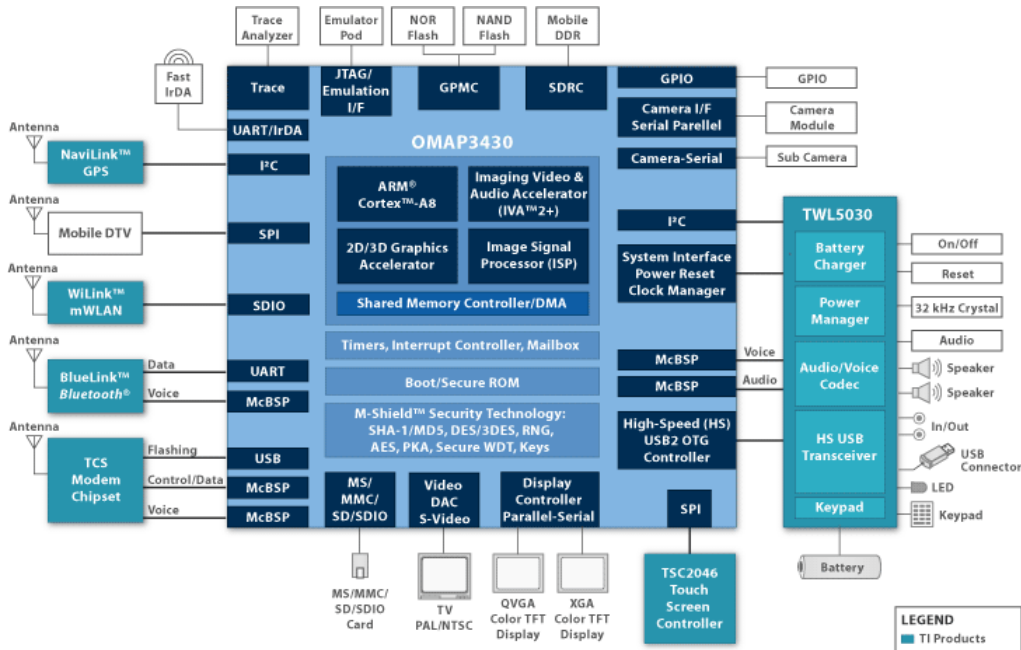

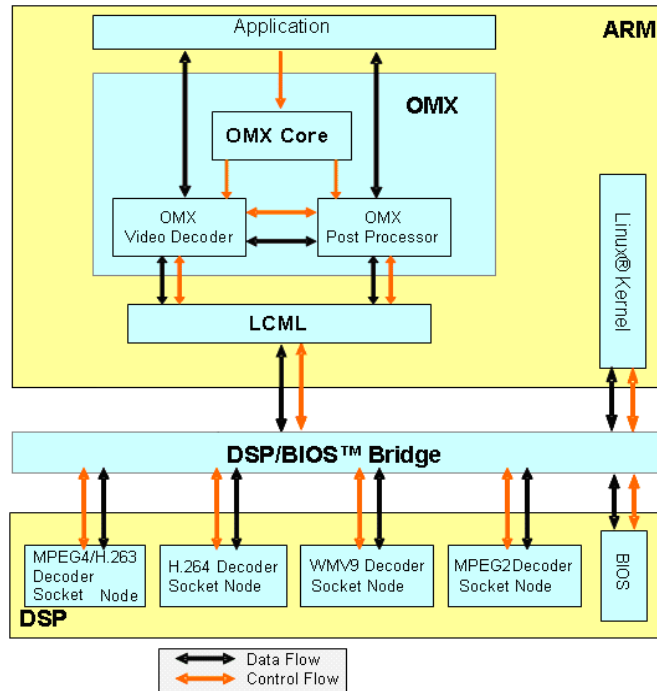Fig. 9.Development platform system diagram.

Fig. 10. Block diagram of video decode architecture in OMAP3430 SDP.

Based on the results reported in Section 3 and considering the capabilities of the chosen developing platform and the existing H.264 decoder, at that stage we have focused on two encoding approaches and respectively their decoding parts implementations, namely H.264 Simulcast and H.264 SEI Message. In the simulcast approach, we put left and right part of the stereo image side by side. The result is an H.264 stream with double frame size. Decoding this stream is equivalent to decoding the stream with double frame size. If we are able to double the existing decoder frame size, we can easily implement H.264 Simulcast. When using H.264 SEI message, left and right views are interlaced, resulting in a stream with double frame rate. From implementation point of view, decoding such stream is equivalent to decoding mono video stream with double frame rate, and de-interlacing each two frames in order to create stereoscopic views.

To implement H.264 Simulcast we needed having a decoder able to decode frames with double frame size, preserving frame rate while for implementing H.264 SEI message we needed a decoder able to decode with double frame rate, and preserving the frame size. Additional steps, such as decoding the SEI message, de-interlacing the left and right views and constructing interleaved frames for the display subsystem, were implemented as well.

At the moment, we are able to decode stereo sequences at HVGA@30fps, which is better than the minimum requirements defined (QWVGA@25fps). However, to ensure flexibility and support of different resolutions, we target HWGA@30fps. As the actual decoding is happening on the DSP of OMAP3430, the CPU load for decoding is quite small – average less than 25%. The re-formatting module is not consuming CPU resources because of a DMA based implementation.

5.3 Integration of auto-stereoscopic display

The limitations of the mobile device, such as screen size, CPU power and battery life limit the choice of a suitable 3D display technology to what is known as auto-stereoscopic displays. Such displays create 3D effect by using dedicated optical elements aligned on the surface of the screen so to ensure that the observer sees different images with each eye. Typically, auto-stereoscopic displays are capable of presenting multiple views to the observer, each one seen from a particular viewing angle along the horizontal direction. The number of views comes at the expense of lowered spatial resolution and brightness loss – and both are limited on a small screen, battery driven mobile device. As mobile devices are normally watched by only one observer, two independent views are sufficient for satisfactory 3D perception.

There are two common types of optical filters redirecting the light of a normal LCD – lenticular sheet, which works by refracting the light, and parallax barrier, which works by blocking the light in certain directions [2]. In both cases, the

intensity of the light rays passing through the filter changes as a function of the angle, as if the light is directionally projected. For our first prototype, we have selected a parallax-barrier based display as it is the much cheaper technology and provides backward compatibility (the ability to switch off, so to work in 2D mode). We have procured a 4.3"WVGA (800px x 480px) transmissive LCD display with barrier on top of it. It creates two different views resulting in resolution change, i.e. 800x480 in 2D mode, and 400x480 per view in landscape 3D mode.

In order to provide proper functionality of the 3D LCD, we have designed a special daughter card to connect the display to the main development platform. It provides proper voltage to requested components, generates the supply voltage for the back-light, provides proper power on/off sequence of the 3D LCD, and level-shifts the signals from the platform to the main display. Fig. 11 shows the daughter-card hosting the 3D LCD as connected to the OMAP3430SDP. After integrating the display, optical parameters of the display, such as crosstalk and angular luminance profile were measured [29] utilizing the methodology developed in [30].

Currently, the platform is capable of decoding and playing stereo video. Integration of a DVB-H front-end is under development.

Fig. 11. Daughter card hosting 3D display.

6. TOWARD USER-CENTERED QUALITY OF EXPERIENCE (UC-QOE)

This section presents our approach toward testing critical elements of the mobile 3DTV technology. In general, subjective quality evaluation is conducted for signal or system development purposes. Its results are typically used in the optimization of system parameters, or, in the development of objective quality metrics. There are several synonyms for subjective quality, such as *perceptual*, *hedonic*, or *experienced quality*. The common thing for all of them is that evaluations of excellence of stimuli are based on human perceptual processes. This section firstly reviews three main streams of subjective quality evaluation methods. Secondly, it presents seven starting points for developing a new holistic evaluation method for experienced multimedia quality evaluation.

6.1 Current research methods in subjective quality assessments

Psichoperceptual approaches: Quantitative evaluation of subjective quality has been used for a long time in quality research. Already in 1974, the International Telecommunication Union (ITU) published its recommendation ITU-R BT.500 [32] "Methodology for the subjective assessment of the quality of television pictures". These methods are also applicable for the assessment of stereoscopic content as it can be found in recommendation ITU-R BT.1438 [31]. The goal of the methods proposed by the ITU is to be able to measure subjective quality according to human perception. Quality evaluation is either based on a pair-wise comparison of stimuli and judgment of perceived differences, or on single-stimulus methods in which quality of stimuli is rated independently [32][33]. All methods create preference orders of the stimuli under test which allows making conclusions about the impact of test parameters on subjective video quality. An overview of the existing methods of psychoperceptual quality evaluation and their application in the research field of stereoscopic video quality evaluation can be found in [34].

User-centered approaches: In contrast to psychoperceptual approaches, the methods within the user-centered approaches expand quality towards human behavioral level in the evaluation of subjective quality. Quality of Perception (QoP),

introduced by Ghinea and Thomas [35], is an evaluation framework which combines user's overall satisfaction with the content and his ability to understand the informational content of the stimuli. However, QoP does not offer any information about the acceptance threshold of quality. The evaluation of the acceptance threshold was first done by McCarthy et al. [36] according to Fechner's Method of Limits and has been extended by Jumisko-Pyykkö et al. [37]. In their Acceptance threshold method they propose a bi-dimensional measure combining retrospective overall quality acceptance and satisfaction. The analysis of the method results in an identification of the acceptance threshold and a preference analysis of the stimuli under test.

While recommendations and standards about quality evaluation methods always target evaluation in controlled laboratory environments, quality evaluation in the *context of use* has become relevant especially for mobile systems. These studies try to close the shortcoming of the artificial settings of laboratory environments which only provide limited or even unknown level of realism [38]. These so called quasi-experimental settings present a new way to conceptualize experimental intervention conducted without full control over potential causal events, control and validity [39]. A review of user-centered approaches and their application in mobile TV and stereoscopic video research is included in [34].

Descriptive quality methods: In contrast to psychoperceptual and user-centered methods, methods of descriptive quality try to understand quality factors and/or relate these factors to quality preferences. Subjective quality is not seen as the result of several predefined variables. Jumisko-Pyykkö et al. [40] state that human behavior in determining subjective quality must be seen as a challenge and, hence, test methods must be open to understand underlying quality attributes. The approaches are thereby different. While Jumisko-Pyykkö et al. used interviews to identify "experienced quality factors" [40], other approaches follow the methods of sensory profiling [41]. In visual quality research these methods have been adapted for still images by Radun et al. [42] and audiovisual, stereoscopic video content by Strohmeier in his Free Choice Profiling approach [43]. Detailed reviews of these descriptive quality methods can be found in [34].

In sum, there are three main streams of subjective quality evaluation methods with different emphasis. Highly controlled psychoperceptual experiments, following ITU recommendations [31][32][33], underline the quality preferences. The work in user-centered quality evaluation expands this concept of quality and relates the evaluations to the use of actual system or service in some degree. The work done in the descriptive quality has started to construct the factors of quality from viewer's point of view. In 3D quality research, additional challenges such as the concept of presence [44] influence and determine subjective quality. The challenge of modern quality research is to take into account all these aspects to build a holistic framework for quality evaluations which allows understanding quality as a whole according to users, system and context of use.

6.2 Seven starting points for developing user-centered evaluation method for measuring experienced quality in the product development purposes

1) Quality perception is an active process combining different levels of human information processing. Active perception integrates always low-level sensorial processing and high-level cognitive processing including personal knowledge, attitudes, expectations and emotions [45]. Each sensory modality has its special characteristics that depend on the physical dimensions of stimuli. In cognitive processing stimuli are interpreted and their personal meaning and relevance to intentions and goals are determined. For example, individual emotions, knowledge, expectations and schemas representing reality affect the weight that each sensory attribute is given and these factors enable human contextual behavior and active quality interpretation [40] [42]. In addition, multimodal perception has its special characteristics. The unified multimodal experience of audiovisual material is created when the information from audio and visual channels are combined. Different modalities can complement and modify the perceptual experience created by other perceptual channels and therefore multimodal experience is more than the simple sum of two different perceptual channels [46].

*2) Component user experience examines quality of critical system component by reflecting the factors of whole user experience***.** For understanding experienced quality of critical system component, we present the basic principles of holistic user-experience (UX) at first. *UX is about technology that fulfils more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex and dynamic encounter. UX is a consequence of a user's internal state --, characteristics of designed system -- and the context -- within the interaction occurs [47].* This definition not only underlines similar characteristics as human perceptual processes, but also summarizes the building blocks of user experience: **user, system** and **context**. User is defined as a person controlling or manipulating the system. She/he has the characteristics of needs, motivations, experiences, expectations, mental state and resources. System is defined as the

system required for the product under examination to work or to be useful. From the user's viewpoint the mobile system can contain a device, browser or player, connection and site or content. By content, we refer to any type of moving scenes. It contain components such as physical, temporal, social, task context, and dynamic changes within these are the essential in the mobile context (see [48] for more detailed descriptions and references).

Component user experience examines the quality of the certain system component taking into account the factors of holistic user experience (Fig. 12a). Critical system component refers to the part of the whole system which performance can impact negatively or prohibit the utility of the whole system from the user's point of view [37]. The aim is at ensuring that experienced qualities of components, developed in isolation from the other components of end-product, are not barriers for adoption, and therefore their acceptability should be studied in the duration of optimization process [37].

3) Quality evaluation experiments are part of human-centered design process. Broadly speaking, 'user-centered design is a design process that views knowledge about users and their involvement in the design process as a central concern' [49]. It is characterized by active user involvement to understand user's requirements, iterative design and evaluation and moreover, a multidisciplinary approach [50]. The gained benefits of user-centered design are listed in the terms of better end-user satisfaction and lower costs of system development. The high-level emphasis is given for the user requirement elicitation at the beginning of development process. For mobile 3D television and video the earliest requirement elicitation was based on three user studies, surveys, focus groups and a probe study and the results are expressed as user, system and service, and contextual requirements. When improving and evaluating quality of critical non-functional system components we apply gathered user requirements to simulate the characteristics of end-product. The requirements offer guidance concerning users, content and context selection for these quality evaluation experiments. Our aim is also to use early stage prototyping parallel to the critical component improvement and evaluation to iteratively validate understanding of user requirements in greater depth.

4) Quality evaluation for system optimization in product development has special characteristics. Quality evaluation stages can be divided into psychophysical assessment, quality optimization assessments and usability testing, and they can be related into product readiness, respectively. In the product development, the emphasis is given for two latest. The usability tests require the highest degree of product readiness to be extensive. Some of the typical features can already be applied earlier stages referring to the optimization studies. In usability testing stage, potential users, contents and usage contexts (especially in the field tests) can be known in the applicable degree as well as the measurement tasks. Quality optimization assessment is also a part of typical testing in product development. However, from the viewpoint of final product its product readiness is not as high as in the usability tests. To reach the acceptable quality level as early as possible in product development, the same factors as in usability testing should be in focus in quality optimization stage.

5) Towards positive quality – good enough for use. In past, the focus of quality evaluation studies has been in examining the negative aspects of constructed 3D quality. For example, cyber sickness, examined using simulator sickness questionnaire (SSQ) has been widely adopted in the studies of 3D display technologies [43] [52]. While these studies underline the failures of the technology, they do not relate quality evaluation to actual action in its context. To put it in a positive way, YouTube with relatively poor provided quality has become successful. The provided quality is good enough for enabling the actual user activities. Positive aspects, like user's interests, can override the negative aspects of technology [36][51][52].

6) Low produced quality, noticeable impairments and heterogeneous stimuli give requires guaranteeing the level of minimum acceptable quality. We use a term low quality to refer to presentation with perceived noticeable impairments, and to make a distinction to perceptually impairment-free high qualities (e.g. top-end multichannel audio or high-definition visual presentations). This distinction is made because the evaluation methods of low and high qualities can be different requiring for example different type of evaluation tasks, experimental procedure or evaluators (naive vs. professional) [32][33][37]. The noticeable impairments, or artifacts, can be resulted from single or combined factors throughout the whole value chain (from content production and packaging, its delivery and transmission and reception including device and its display due to the required high-level of optimization of the system. This can result very heterogeneous stimuli material for the experiments. For low quality and for these circumstances, the measurements of acceptance threshold indicating minimum useful quality is needed Fig. 12b [37]). The goal is to ensure that the low produced quality is set in a way that constitutes no obstacle to the wide audience acceptance of a product or service [37].

7) Overall quality evaluation approach is suitable for user-oriented quality evaluation. In overall or global quality evaluations, participants assess stimuli as whole. This approach is suitable for naive participants and under the usage of heterogeneous stimuli with different type of impairments and multimodal presentation. This approach also assumes that

human information processing is integrative in nature including low-level sensorial and high-level cognitive processing. In contrast to overall evaluations, perceptual evaluations of certain attribute (e.g. jerkiness) can be conducted with trained assessors.
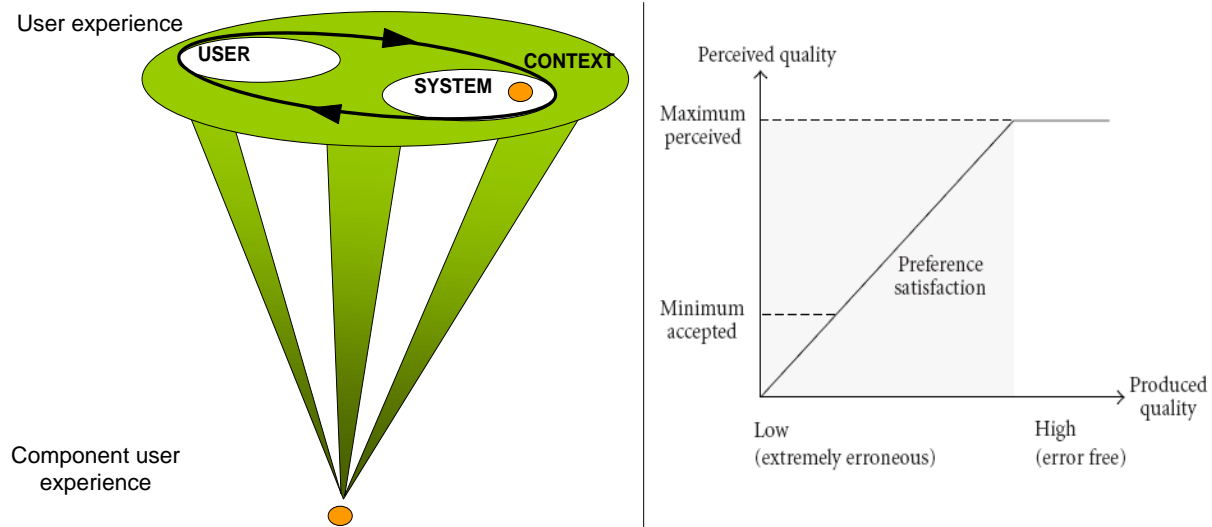

Fig. 12. a) User experience and component user experience; b) Levels of produced and perceived quality. In low produced qualities, perceptually minimum accepted quality level is a threshold for the useful quality. [37].

Taken together the seven principles presented: we have developed a holistic framework for user-centered quality evaluation method for measuring 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. The method is referred to as user-centered quality evaluation method if any of the four listed factors relating the evaluation to potential use is taken into account in the evaluation research.

7. CONCLUSIONS

In this paper, we overviewed the current stage of a project developing core elements of what we call mobile 3DTV technology. Some of the methods, such as coding and IP-streaming of stereo video are in a good level of maturity. However, they should be jointly optimized in order to deliver the best possible quality to the user given the media-rich yet power-constrained mobile platform. Addressing this optimization, we have been employing a modern user-centered approach to developing and testing the system components. We have been further advancing this approach to a novel methodological framework to be used throughout the whole development process of the system. In the future research, extensive subjective tests shall help to identify the most perceived artifacts created during the processing stages and experienced on the portable auto-stereoscopic display. The best formats and techniques for coding and error protection will be also selected after careful QoE testing.

REFERENCES

[1] Discussion forum "3D in the Home: How Close are We?", Stereoscopic Displays and Applications XVIII, San Jose, California, USA, 2007, www.stereoscopic.org .

[2] Pastoor S., 3D Displays, in *3D Video Communications*, (O. Schreer, P. Kauff, T. Sikora, eds.), John Wiley and Sons, 235-260 (2005).

[3] Alatan, A., Yemez, Y., Gudukbay, U., Zabulis, X., Muller, K., Erdem, C., Weigel, C.2007. Scene Representation Technologies for 3DTV—A Survey, IEEE T Circ Syst VT 17 (11),(Nov. 2007),1587-1605).

[4] Smolic, A., Mueller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G.B., Triantafyllidis, G., Koz, A. 2007. Coding Algorithms for 3DTV—A Survey, IEEE T Circ Syst VT 17 (11), (Nov. 2007), 1606-1621.

[5] Akar, G.B. Tekalp, A.M. Fehn, C. Civanlar, M.R. 2008. Transport Methods in 3DTV—A Survey, IEEE T Circ Syst VT 17 (11), (Nov. 2007), 1622-1630.

[6] EMBC (2007e) *Technical Workstream*, http://www.ebu.ch/CMSimages/en/tec_embc_technical_report_tcm6-50235.pdf.

[7] Communication on Strengthening the Internal Market for Mobile TV, COM (2007), 409 of 18 July 2007.

[8] World DMB, http://www.worlddab.org/country_information/south_korea

[9] Lee H., Cho S., Yun K., Hur N., Kim J., A Backward-compatible, Mobile, Personalized 3DTV Broadcasting System Based on T-DMB, in *Three-Dimensional Television* (H. Ozaktas and L. Onural, eds.), Springer, 11-28 (2008).

[10] www.3dphone.org

[11] www.mobile3dtv.eu

[12] Tikanmäki, A., Smolic, A., Mueller, K., Gotchev, A. 2008. Quality Assessment of 3D Video in Rate Allocation Experiments, in IEEE Int. Symposium on Consumer Electronics (14-16 April, Algarve, Portugal).

[13] Hannuksela, M., Kumar Malamal Vadakital, V., Jumisko-Pyykkö, S. 2007. Unequal Error Protection Method for Audio-Video Broadcast over DVB-H. EURASIP J Adv Sig P, Article ID 71801, 12 pages.

[14] Fehn, C., Kauff, P., de Beeck, M. Op, Ernst, F., IJsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., and Sexton, I. 2002. An Evolutionary and Optimised Approach on 3D-TV, Proceedings of Int. Broadcast Conference (Amsterdam, The Netherlands, September 2002), 357-365.

[15] Merkle P., Brust H., Dix K., Wang Y., Smolic A., Adaptation and optimization of coding algorithms for mobile 3DTV, Technical report available at www.mobile3dtv.eu/results

[16] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", November 2007.

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

[18] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information", Doc. N8768, Marrakech, Morocco, January 2007.

[19] Z. Wang, L. Lu, and A.C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.

[20] L. Stelmach, W.J. Tam; D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," IEEE Trans. Circuits and Systems for Video Technology, Vol. 10, No. 2, pp. 188-193, March 2000.

[21] G. Faria , J. A. Henriksson , E. Stare and P. Talmola "DVB-H: digital broadcast services to handheld devices," Proc. IEEE, vol. 94, pp. 194, Jan. 2006

[22] FATCAPS: A Free, Linux-Based Open-Source DVB-H IP-Encapsulator [Online]. Available at: http://amuse.ftw.at/downloads/encapsulator

[23] JustDvb-It [Online]. Available: http://www.cineca.tv/labs/mhplab/JustDVb-It.html

[24] Decaps software [Online] Available. http://www.mobile3dtv.eu/

[25] E. Kurutepe, A. Aksay, C. Bilen, C. G. Gurler, T. Sikora, G. Bozdagi Akar, A. M. Tekalp, "A Standards-Based, Flexible, End-to-End Multi-View Video Streaming Architecture," Packet Video Workshop 2007, Lausanne, Switzerland.

[26] R. Calonati, "Techniche a diversita compatiili con il livello fisico dello standard DVB-T/H", Laurea in Ingengeria thesis, Universita degli studi di Firenze, 2007.

[27] OMAP™ 3 family of multimedia applications processors, Product bulletin, available at http://focus.ti.com/pdfs/wtbu/ti_omap3family.pdf

[28] OpenMAX Overview, http://www.khronos.org/openmax/

[29] Boev A., Gotchev A., and Egiazarian K., Stereoscopic artifacts on portable auto-stereoscopic displays: what matters?, In Proc. Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 09, Jan. 15-16, 2009, Scottsdale, Arizona, U.S.A.

[30] Boev, A., Gotchev A. and Egiazarian K., "Crosstalk measurement methodology for auto-stereoscopic screens", Proc. of 3DTV-CON, Kos, Greece, 2007.

[31] ITU-R BT.1438, Subjective assessment of stereoscopic television pictures, Recommendation ITU-R BT.1438, ITU Telecom. Sector of ITU, 2000

[32] ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-11, ITU Telecom. Sector of ITU, 2002

[33] ITU-T P.911 Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia application," International Telecommunication Union - Telecommunication sector, 1998.

[34] Jumisko-Pyykkö, S., Strohmeier, D. Report on research methodologies for the experiments, Technical report of the MOBILE3DTV project, December 2008, available at www.mobile3dtv.eu/results

[35] Ghinea, G. and Thomas, J. P., "QoS impact on user perception and understanding of multimedia video clips," in Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA '98), pp. 49–54, Bristol, UK, September 1998.

[36] McCarthy, J. D., Sasse, M. A., and Miras, D. 2004. Sharp or smooth?: comparing the effects of quantization vs. frame rate for streamed video. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '04. ACM, New York, NY, 535-542

[37] Jumisko-Pyykkö, S. Kumar Malamal Vadakital, V., Hannuksela, M.M. "Acceptance Threshold: Bidimensional Research Method for User-Oriented Quality Evaluation Studies. ". International Journal of Digital Multimedia Broadcasting, 2008.

[38] Wynekoop, J. L., Russo, N., L., Studying system development methodologies: an examination of research Information Systems Journal. 7.47- 6. 1997

[39] Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

[40] Jumisko-Pyykkö, S. Häkkinen, J., Nyman, G. Experienced Quality Factors - Qualitative Evaluation Approach to Audiovisual Quality. Proceedings of IST/SPIE conference Electronic Imaging, Multimedia on Mobile Devices 2007

[41] Stone, H & Sidel, J. Sensory Evaluation Practises. 3rd ed. Food Science and Technology, International Series. 2004.

[42] Radun, J., Leisti, T., Häkkinen, J., Ojanen, H., Olives, J.-L., Vuori, T., and Nyman, G. 2008. Content and quality: Interpretation-based estimation of image quality. ACM Trans. Appl. Percpt. 4, 4, Article 21 (January 2008)

[43] Strohmeier, D: Wahrnehmungsuntersuchung von 2D vs. 3D Displays in A/V-Applikationen mittels einer kombinierten Analysemethodik, DiplomarbeitDiploma Thesis, Ilmenau University of Technology, Germany, 2007

[44] Minsky, M. (1980, June). Telepresence. Omni, 45-51.

[45] Neisser, U. 1976. Cognition and Reality, Principles and Implications of Cognitive Psychology, San Francisco: W.H. Freeman and Company.

[46] Hands, D. S., "A Basic Multimedia Quality Model," IEEE Transactions on Multimedia,Vol.6,No.6, 806-816(2004).

[47] Hassenzahl, M., Tractinsky, N. 2006, User Experience – a Research Agenda. Behaviour and Information Technology,Vol. 25, No. 2, March-April 2006, pp. 91-97.

[48] Jumisko-Pyykkö, S., Weitzel, M., Strohmeier, D. "Designing for User Experience: What to Expect from Mobile 3D TV and Video? " . Proceedings of the First International Conference on Designing Interactive User Experiences for TV and Video. October 22 - 24, 2008, Silicon Valley, California, USA.

[49] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T. Human-Computer Interaction. Wokingham, UK: Addison-Wesley, 1994.

[50] ISO 13407. 1999. Human-centred design processes for interactive systems. International Standard, the International Organization for Standardization.

[51] Jumisko, S., Ilvonen, V., Väänänen-Vainio-Mattila, K. The Effect of TV Content in Subjective Assessment of Video Quality on Mobile Devices. Proceedings of SPIE, volume 5684. Multimedia on Mobile Devices, Reiner Creutzburg, Jarmo H. Takala (Eds), March 2005, pp. 243-254.

[52] Häkkinen, J., Liinasuo, M., Takatalo, J., and Nyman, G. 2006. Visual comfort with mobile stereoscopic gaming. Proc. of SPIE. Vol. 6055.