Three-Dimensional Media for Mobile Devices

This paper provides an overview of technologies to deliver 3-D media to next-generation mobile devices; the importance of efficient and robust transmission over error-prone channels is stressed.

By Atanas Gotchev, *Member IEEE*, Gozde Bozdagi Akar, *Senior Member IEEE*, Tolga Capin, Dominik Strohmeier, and Atanas Boev

ABSTRACT | This paper aims at providing an overview of the core technologies enabling the delivery of 3-D Media to next-generation mobile devices. To succeed in the design of the corresponding system, a profound knowledge about the human visual system and the visual cues that form the perception of depth, combined with understanding of the user requirements for designing user experience for mobile 3-D media, are required. These aspects are addressed first and related with the critical parts of the generic system within a novel user-centered research framework. Next-generation mobile devices are characterized through their portable 3-D displays, as those are considered critical for enabling a genuine 3-D experience on mobiles. Quality of 3-D content is emphasized as the most important factor for the adoption of the new technology. Quality is characterized through the most typical, 3-D-specific visual artifacts on portable 3-D displays and through subjective tests addressing the acceptance and satisfaction of different 3-D video representation, coding, and transmission methods. An emphasis is put on 3-D video broadcast over digital video broadcasting—handheld (DVB-H) in order to illustrate the importance of the joint source-channel optimization of 3-D video for its efficient compression and robust transmission over error-prone channels. The comparative results obtained identify the best coding and transmission approaches and enlighten the interaction between video quality and depth perception along with the influence of the context of media use. Finally, the paper speculates on the role and place of 3-D multimedia mobile devices in the future internet continuum involving the users in cocreation and refining of rich 3-D media content.

KEYWORDS | Autostereoscopic displays; graphical user interface; MPE-FEC; multiview coding; open profiling of quality; user-centric design; 3-D visual artifacts

I. INTRODUCTION

Three-dimensional media is an emerging set of technologies and related content in the area of audio–video entertainment and multimedia. It is expected to bring realistic presentation of third dimension of audio and video and to offer immersive experience to the users consuming such content. While emerging in areas such as 3-D cinema and 3-D television, 3-D media has also been actively researched for its delivery to mobile devices.

The general concept of 3-D media 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 3-D 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 3-D experience for the next generation 3DTV. Advanced light-field and holographic displays have been anticipated in the midterm future. On the research side, various aspects of 3-D content creation, coding, delivery, and system integration have been addressed by numerous projects and standardization activities [1]–[3]. At a first sight, these developments position 3-D media as a rather diverging technology with respect to mobile multimedia as the former relies on big screens and realistic visualization and the latter relies on portable

displays. Still, a symbiosis between 3-D and mobile media has been considered rather attractive. 3-D would benefit from being introduced also to the more dynamic and novel technology-receptive mobile tech market. Mobile TV and video and the corresponding broadcasting standards would benefit from the rich content leading to new business models. The research challenge of achieving this symbiosis is to adapt, modify, and advance the 3-D video technology, originally targeted for large screen experience, for the small displays of handhelds.

The introduction of 3-D media to handhelds is supported by the current trend of developing novel multicore processors as an effective way to reduce the power consumption while maintaining or increasing the performance [4]. Increasing the number of cores and thus offering parallel engines is perfectly suitable for 3-D data, which naturally call for parallel processing. New multicore platforms for mobile applications offer balanced architectures to support both data-dominated and control-dominated applications [5]. Examples are the Texas instruments' OMAP 4 [6], NXP's LH7A400 [7], Marvell's PXA320 [8], NVIDIA Tegra APX 2500/2600 Series, Next Generation NVIDIA Tegra [9], [10], Qualcomm Snapdragon Series [11], and ST Ericsson's U8500 [132]. The aim in designing such multicore processors has been to achieve high system clock rate, optimize the memory use and interconnections between cores, and provide functionality for new rich multimedia applications by more powerful graphical accelerators and digital signal processors. Support of 3-D graphics for 3-D user interfaces and 3-D gaming as well as existing and future multimedia encoders has been targeted. Specifically, 3-D rendering has been considered to be implemented primary on a dedicated hardware accelerator than on a general-purpose central processing unit (CPU), allowing both faster execution and lower power consumption, which are crucial for mobile devices. In addition, modern application programming interfaces, such as OpenGL ES 2.0, emphasize parallel processing design, making it also possible to support more advanced and data-intensive 3-D applications on a mobile device. One of the research challenges is to design efficient 3-D processing algorithms, which reduce the internal traffic between the processing elements and the memory, while maintaining low power consumption [12]. While modern multicore development platforms are available for integrating 3-D video decoding, processing, and playing algorithms, it is the new portable 3-D displays that should make the difference in delivering new user experience.

This paper analyses the process of bringing 3-D media to mobiles. Section I analyzes what is important to know *before* beginning the design of a 3-D media system for mobiles. The section starts with a brief overview of the basics of depth perception by the human visual system (HVS) and the relative importance of various 3-D visual cues. Along with psychophysical factors, novel user studies are presented that help to understand the user expectations and requirements

concerning 3-D media for mobiles. The introduction of new media requires also novel research approaches regarding users and new, user-centric, approaches in designing critical parts of the whole system. Those are presented next, just before the overview of the 3-D video delivery chain with its main blocks. Emphasizing 3-D video is important, as it illustrates the entertainment value of 3-D for mobile users. Optimal content formats and coding approaches, as well as streaming and channel coding approaches especially tailored to 3-D, are reviewed as to make a link to the other papers in this special issue. Thus, Section II connects the user with the system through psychophysical and psychological aspects and the ways those have to be investigated.

Section III is all devoted to portable 3-D displays, as the main part of the next-generation 3-D-enabled mobile devices playing a decisive role in the adoption of the new technology. Related display technologies are overviewed. Display optical parameters that determine the quality of 3-D perception are summarized and measurement results are presented to characterize and compare various displays.

The knowledge about portable 3-D displays forms the basics to proceed further with Section IV, where user experience of 3-D mobile media is explored in details. 3-D-specific artifacts are reviewed and put against the stages of the delivery chain being responsible for their generation and to the specifics of the human visual system. Furthermore, novel studies aimed at identifying best accepted 3-D video representation formats, and source and channel coding methods are presented. Objective comparisons are complemented by results from extensive subjective tests based on novel design methodologies. The studies on 3-D video are completed at the end of the section with an overview of recent advances in 3-D graphical user interfaces.

Section V presents a foreseeing of more futuristic usage scenarios of 3-D-enabled handhelds where 3-D media is not only *delivered* to users but also *co-created* by them using the tools as envisaged by Future Internet. Such concept poses even more challenging research questions addressing the way 3-D audio and video content is captured and processed by mobiles to contribute to a collaborative creation of rich 3-D media content and corresponding services.

II. INTERDISCIPLINARY ASPECTS OF 3-D MOBILE MEDIA SYSTEM DESIGN

A. Perception of Depth

The human visual system can be considered as a set of separate subsystems operating together in a unified manner. There are largely independent neural paths responsible for transmitting the spatial, color, and motion information to the brain [28]. On perceptual level there are separate visual mechanisms and neural paths, while on cognitive level there are separate depth cues contributing to the formation of 3-D spatial vision [28], [29]. These
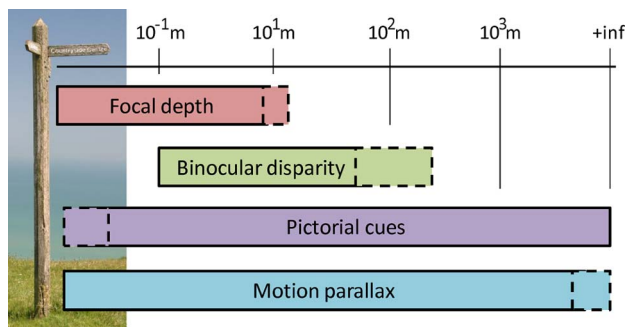
Fig. 1. *Depth perception as a set of separate visual "layers."*

depth cues are with varying importance for an individual observer [30]–[32]. The depth cues used for assessing the depth by different layers in human vision are shown in Fig. 1 and are as follows.

- Accommodation—This is the ability of the eye to optically focus on objects at various distances.
- Binocular depth cues—These result from the position of the two eyes observing the scene from slightly different angles. The eyes tend to take a position that minimizes the difference of the visual information projected in both retinae. The process is called *vergence* and can be characterized by the angle between the eyes used as a depth cue. With the eyes converged on a point, *stereopsis* is the subsequent process that uses the residual disparity of the surrounding area for depth estimation relative to the point of convergence.
- Pictorial cues—These include shadows, perspective lines, and texture scaling and can be perceived even with a single eye.
- Motion parallax—This is the process in which the changing parallax of a moving object is used for estimating its depth and 3-D shape. Similar mechanism has been observed to be used by insects and is commonly referred to as "insect navigation" [38].

A 3-D media system has to maintain adequate 3-D visual cues. Accommodation is the primary depth cue for very short distances, where an object is hardly visible with two eyes. Its importance decreases sharply with increasing the distance. HVS tends to combine accommodation with convergence, using the information from the latter to correct the refraction power and to ensure clear image of the object being tracked. In the real world, accommodation and convergence points coincide; however, on stereoscopic displays, they may differ as eyes focus on the screen and try to converge according to the binocular difference. This discrepancy leads to so-called "accommodation–convergence rivalry," which is a major limiting factor for such displays. Binocular depth cues have been the most used in "3-D cinema," and subsequently in 3DTV and 3-D for mobiles, by presenting different-perspective images to

the two eyes. Binocular vision is quite vulnerable to artefacts: an "unnatural" stereo pair presented to the eyes can lead to nausea and "simulator sickness," as the HVS is not prepared to handle such information [37]. About 5% of all people are "stereoscopically latent" and have difficulties assessing binocular depth cues [28], [29]. Such people perceive depth, relying only on depth cues coming from other visual layers. Pictorial cues work for longer distances, where binocular depth cues become less important. At medium distances, pictorial and binocular cues are combined and for such distance the perception can be ruined by missing subtle pictorial details, even if stereoscopy is well presented. It is said that the scene exhibits "puppet theater" or "cardboard effect" artifacts. The motion parallax depth cues might be affected primarily by artifacts appearing in temporal domain such as motion blur and display persistence.

An interesting suggestion is that binocular and monocular depth cues are independently perceived. It has been supported by both subjective experiments (e.g., the famous experiments with so-called "random dot stereograms" [33]) and anatomical findings. The latter have shown that first cells that react to a stimulus presented to either of the eyes (binocular cells) appear at a late stage of the visual pathways, more specifically in the V1 area of brain cortex. At this stage, only the information extracted separately for each eye is available to the brain for deduction of image disparity [28]. A practical implication of the above suggestion concerns the modeling, assessment, and mitigation of visual artifacts building on the hypothesis that "2-D" (monoscopic) and "3-D" (stereoscopic) artifacts would be perceived independently [34]. Planar "2-D" artifacts, such as noise, ringing, etc., are thoroughly studied in the literature [35], [36], while artifacts that affect stereoscopic perception have been addressed more recently [39]. We present more details on 3-D visual artifacts in Section IV, after presenting the main blocks of a 3-D media system and the specifics of portable 3-D displays.

B. User Issues at the Beginning of 3-D Media System Design

The perception of depth is an important aspect in the development of 3-D media on mobile devices. However, an optimized development of such systems must take into account further requirements. Like in every product development process, the goal is that the prospective end product as a whole will satisfy the end users. This satisfaction is a key requirement for the success of the product. To describe users' needs and expectations about the product under development, user requirements are commonly specified before and verified, and if necessary redefined, cyclically during the development process [105]. By definition, user requirements describe any externally visible function, constraint, or other property that a product must provide to reach user satisfaction [126]. However, this product-oriented definition is limited as it overlooks the

characteristics of the end users. User experience (UX) tries to understand end users' needs, concerns, and expectations more broadly. It has been defined as being about technology that fulfils more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex, and dynamic encounter [41]. According to Hassenzahl and Tractinsky [41], UX is "*a consequence of a user's internal state* [...], *characteristics of designed system* [...] *and the context* [...] *within the interaction occurs.*"

1) User Requirements for Designing User Experience for Mobile 3-D Media: In the development of 3-D media systems and services, the identification of user requirements plays a crucial role. Three-dimensional mobile media combines the technologies of 3-D media and mobile devices. Each of these technologies has its own user requirements that need to be fused into a new system providing a seamless UX. Mobile media research has identified three building blocks for UX. Roto [42] describes them as 1) user, 2) system and services, and 3) context. Following these building blocks of mobile UX, a large study of a methodological triangulation has been conducted to target the explicit and implicit user requirements for mobile 3-D video [103], [104]. In that study, an online survey, focus groups, and a probe study are combined to be able to holistically elicit user requirements. The survey has been used first to identify and verify needs and practices towards the new system. It has been then extended with the results of focus groups. The focus group studies have been conducted to overcome the weakness of online surveys to generate new ideas. More specifically, focus groups aimed at collecting possible use scenarios for mobile 3-D media as well as an imaginary design of the device and the relating services. However, both online survey and focus groups only cover the *explicit* user requirements. Especially focus groups do not take into account individual, implicit requirements as those are often overwhelmed by the group effect. To complete the user requirements, the probe study as the third method has been applied to collect those personal needs and concerns. In this probe study, test participants played with a probe package that contained a disposable camera, a small booklet, and material for a collage, as illustrated in Fig. 2. Their task was to log their thoughts and ideas about mobile 3-D media in different daily situations and therewith in different contexts with help of the diary and the disposable camera. At the end, test participants set up a collage in a reflective task about their own opinion on mobile 3-D video. Examples are shown in Fig. 3 [103], [104].

The above referred studies [103], [104] have framed the user requirements for mobile 3-D video with respect to all three building blocks of UX: the user, the system and service, and the context. The results show that the prospective users of mobile 3-D television and video systems want to satisfy *entertainment* and *information* needs. Participants outline the importance of the added value given through

Fig. 2. *Probe package provided to participants during user requirement elicitation studies [104].*

increased realism and a *closer emotional relation* to the content. It is noteworthy that these expectations about added value differ from the common ideas about added value of 3-D. For large screens or immersive environments, the added value is commonly expressed as *presence*, the users' feeling of being there [106]. Related to system and services, users expect devices with a display of the size of 3–5″. The display must provide possibilities to shift content-dependently between monoscopic and stereoscopic presentation. The expected content relates to the entertainment and information needs. TV contents like sports, documentaries, or even news are mentioned by the test participants. However, the requirements also show that nontelevision content has high potential for the services. Applications like interactive navigation or games are of high interest for the users. To access the different services, users can image both on-demand and push services that will be paid by monthly payment or pay-per-view. The expected use (the context) is mainly in public transports, cafes, or waiting situations and in private viewing, when concentrating on the content. Especially young people have told also about a need for shared viewing. However, interaction with the context (as, e.g., defined in Section IV-C) or with other users on one display is not expected regularly. As mobile 3-D media is well suited for waiting situations and short transport trips, the expected viewing time is about 15 min. In exceptional cases like journeys also longer duration up to half an hour may occur.

2) A Holistic User-Centered Research Framework for Mobile 3-D Television and Video: The elicited user requirements for mobile 3-D video show what people expect from this new technology. A challenge during the development process is now how to include these requirements into the technology. The user-centered design process is defined in ISO 13407 [105] as a cyclic process within a product

Fig. 3. *Examples of two collages from probe study on user requirements for mobile 3DTV [104].*

development, as exemplified in Fig. 4. It is especially useful at an early stage of the development as it can show opportunities to improve the quality of the system related to the requirements of the prospective end users. However, user-centered design can be used during the whole development process.

Current work on mobile 3-D media has been conducted under the framework of user-centered quality of experience (UC-QoE) [93], [95]. In general, QoE is defined as "the overall acceptability of an application or service, as perceived subjectively by the end-user" [116]. QoE takes into account the cognitive processes of human perception that relate to interpretation of perceived stimuli with regard to emotions, knowledge, and motivation. More broadly, QoE can be regarded as a "multidimensional construct of user perception and behavior" [119]. The UC-QoE approach represents a holistic framework for subjective quality optimization of mobile 3-D video. It takes into account prospective users and their requirements, evaluation of system characteristics, and evaluation of quality in the actual context of use [95]. The framework provides a set of evaluation methods to be able to study the different aspects of QoE. Especially two challenges have been identified along with shortcomings of currently existing quality evaluation methods. Commonly, subjective quality is measured using psychoperceptual evaluation methods that are provided mainly in ITU recommendations [101], [102] (see [93] for a review). First, these methods target a quantitative analysis of the excellence of overall quality disregarding users' quality interpretations, descriptions, and evaluation criteria that underline a quantitative quality preference. Second, these methods have been designed for quality evaluations in controlled, homogenous environments. However, mobile applications are meant specifically for use in extremely heterogeneous environments as the user requirements show [96], [103]. To get a higher external validity of the results, these systems must be evaluated additionally in their actual context of use.

There has been a gap between quantitative evaluation of the user satisfaction with the overall quality and the underlying components of quality in multimedia quality evaluation [110]. To address this gap, an approach referred to as open profiling of quality (OPQ) has been developed and successfully applied in mobile 3-D media research [107], [108], [110]. OPQ is a mixed method that combines evaluation of quality preferences and the elicitation of individual experienced quality factors [110]. *Sensory profiling*, originally used in food sciences as a research method "*to evoke, measure, analyze and interpret reactions to those*
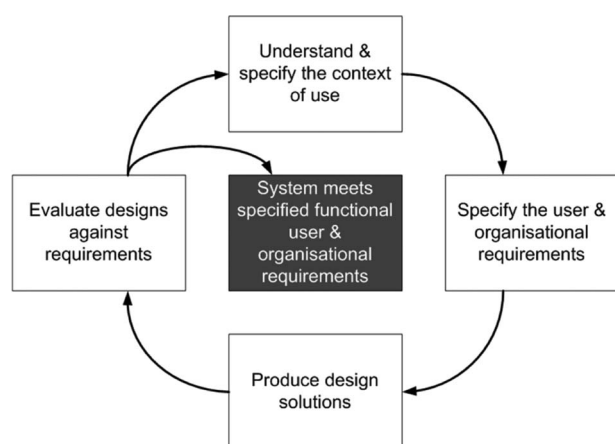

Fig. 4. *Cyclic process of user-centered design according to ISO 13407 [105].*

characteristics of food and materials as they are perceived by senses of light, smell, taste, touch and hearing..." [112] has been adapted for 3-D media studies. Final outcome of OPQ is a combination of quantitative and sensory data sets connecting users' quality preferences with perceptual quality factors. In its sensory profiling task, test participants develop their own idiosyncratic quality attributes. These attributes are then used to evaluate overall quality [109]. The sensory data can be analyzed using multivariate analysis methods [100], [117] and the results show a perceptual model of the experienced quality factors.

To overcome the limitations of a controlled laboratory environment, the second evaluation tool within the UC-QoE framework is a hybrid method for quality evaluation in the context of use [118]. Context of use is defined as the entity of physical and temporal contexts, task and social contexts as well as technical and informational contexts [94], [118]. The extension of quality evaluation to the context of use aims at extending the external validity of results gained in controlled environments. Concrete results of applying the two evaluation tools to characterize UC-QoE of mobile 3-D media are given in Section IV-B.

C. Three-Dimensional Media Delivery Chain for Mobiles

A system for delivery of 3-D media to mobile devices is conceptualized in Fig. 5. On a general level, its building blocks do not differ much from the blocks of a general 3DTV system. The system includes stages of content creation, format conversion to a compression- and delivery-friendly format, compression with subsequent transmission over some wireless channel, decoding, and displaying on a mobile terminal.

The specifics of this general system are determined by the foreseen mobile applications such as video conferencing, online interactive gaming, and mobile 3DTV; the characteristics of the wireless networks such as digital video broadcasting—handheld (DVB-H), digital multimedia broadcasting (DMB), MediaFlo, 3G, and the computational power of the terminal device. For real-time video communication such as video conferencing, real-time encoding and decoding is necessary simultaneously at both terminal devices with low delay. The transmission bandwidth is restricted to the capabilities of the mobile phone line that makes the bitrate for the 3-D video signal very limited. For mobile 3DTV, the decoding is only done at the receiver side with some possible buffering. However, in this case, rendering and display at full frame rate and with minimum artifacts is needed. In addition, due to the characteristics of the wireless channel, the quality cannot be guaranteed, which brings the necessity of robustness to channel errors. For online interactive gaming, again fluent decoding, rendering, and possible content adaptation is needed at the terminal devices with low delay. In addition to all these specific requirements and limitations, low power consumption and low complexity is a must for mobile video applications.

1) Three-Dimensional Video Representation and Coding: Considering the above limitations, the first issue to look at is the format to be used for the delivery of 3-D video and 3-D graphics. If the latter is to be transmitted as a polygon mesh, formed by collection of vertices and polygons to define the shape of an object in 3-D, then MPEG4 AFX is a well-known compression method to be used. Three-dimensional video offers more diverse alternatives for its representation and coding and we will concentrate on these other than the 3-D graphics. The first research attempts and related standardization efforts regard 3-D video represented either by single video channel augmented by depth information [view + depth (V + D)] or by parallel video streams coming from synchronous cameras. In the latter representation approach, the video streams can be compressed jointly (multiview) or independently (simulcast).

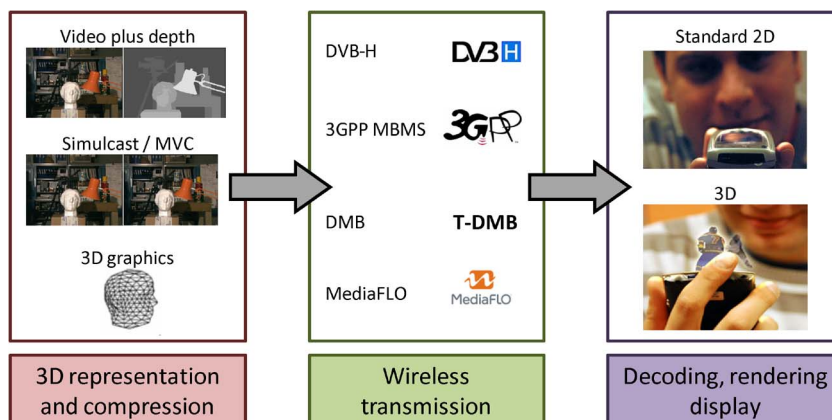*V + D Coding:* ISO/IEC 23002-3 Auxiliary Video Data Representations (MPEG-C part 3) is meant for

Fig. 5. *End-to-end 3-D video transmission chain.*

applications where 3-D video is represented in the format of single view plus associated depth (V + D), where the single channel video is augmented by the per-pixel depth attached as auxiliary data [122]. The presence of depth data allows for synthesizing desired views at the receiver side and adjusting the view parallax, which is beneficial for applications where the display size might vary, which is the case of mobile devices. V + D coding does not require any specific coding algorithms. It is only necessary to specify high-level syntax that allows a decoder to interpret two incoming video streams correctly as color and depth. Additionally, it is backward compatible and its compression efficiency is high as the side depth channel is represented by a gray-scale image sequence. Few studies have reported algorithms and prototypes for view synthesis based on V + D (ISO/IEC 23002-3) on mobile devices [129], [130]. Contrary to their compression efficiency, such systems have high complexity for both sender and receiver sides. Before encoding, the depth data have to be precisely generated. For real scenes, this is done by depth/disparity estimation from captured stereo or multicamera videos using extensive computer vision algorithms plus possibly involving range sensors. For synthetic scenes, this is done by converting the *z*-buffer data resulting from rendering based on 3-D models. V + D representation is only capable of rendering a limited depth range and additional tools are needed to handle occlusions. Recent advances to this approach suggest using so-called depth-enhanced stereo or multilayer depth [75], which successfully tackle the occlusion issue for the price of increased complexity. At the receiver side, view synthesis has to be performed after decoding to generate the stereo pair, which is not very trivial for mobile devices to achieve in real time especially for high resolutions.

Multiview Video Coding (MVC, ISO/IEC 14496-10:2008 Amendment 1 ITU-T H.264): It is an extension of the advanced video coding (AVC) standard [121]. It targets coding of video captured by multiple cameras. The video representation format is based on *N* views. MVC exploits temporal and inter-view redundancy by interleaving camera views and coding in a hierarchical manner. There are two profiles currently defined by MVC: multiview high profile and stereo high profile, which are both based on the ITU-T H.264 AVC with a few differences [77]. Stereo high profile is also chosen as the supported format for the 3-D Blu-Ray discs. The main prediction structure of MVC is quite complex introducing a lot of dependencies between images and views. In order to decrease the complexity, an alternative simplified structure is presented in [90] and shown to be very close to the main prediction structure in terms of overall coding efficiency. In this simplified prediction structure, the temporal prediction remains unchanged when compared to original MVC prediction structure, but spatial references are only limited to anchor frames, such that spatial references are only allowed at the beginning of a group of pictures (GOP) between *I* and *P*
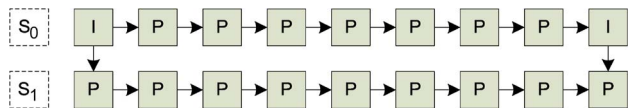

Fig. 6. *Simplified IPP… prediction structure of MVC codec with inter-view references in anchor frames.*

pictures. This simplified version is shown in Fig. 6 for stereoscopic video where only two views (left and right views—S_0 and S_1) exist.

It should be emphasized that this coding is also backward compatible meaning that the only mono-capable receivers will still be able to decode and watch left view, which is nothing but a 2-D conventional video, and simply discard the other view, since left view is encoded independent of the right view.

Research on coding of multiview video and V + D has reached a good level of maturity and the related international standards are perfectly applicable for mobile 3-D video systems and services. However, there are inferior points that prompt for further research. While the approach based on coding of single view plus dense depth seems to be preferred for its scalability, it might be too computationally demanding for the terminal device as it requires view rendering and hence make the device less power efficient. MVC, i.e., compressing the two views by joint temporal and disparity prediction techniques is not always efficient for compression. Researchers have hypothesized that in a mobile device the stereo perception can be based on reduced cues and suggested approaches based on reduced spatial resolution, so-called mixed resolution stereo coding (MRSC) [114]. In this approach, one of the views is kept intact while the other is properly spatially decimated to a suitable resolution where the stereo is still well perceived [114]. Though subjective studies have not proved the MRSC coding hypothesis and such compression has been evaluated inferior to MVC and V + D [109], the approach bears a research potential especially when combined also with MVC type of motion/disparity prediction [115].

Simulcast Coding/Interleaved Coding: Another way to code 3-D video is to use existing video codecs to stereoscopic video with/without an interleaving approach. If no interleaving is used, one achieves *simulcast coding* that is not any different than coding a conventional 2-D video with a video encoder in the sense that both of the views are coded as two completely independent 2-D videos [76]. This method allocates the highest bitrate for a video compared to the other solutions, but is the least complex. On the other hand, *interleaving* [78] can be used as time multiplexing [Fig. 7(a)], spatial multiplexing as over/under [Fig. 7(b)], spatial multiplexing as side-by-side [Fig. 7(c)] [Fig. 7(b) and (c) is also called frame-compatible modes]. This method is currently used by the broadcasters doing
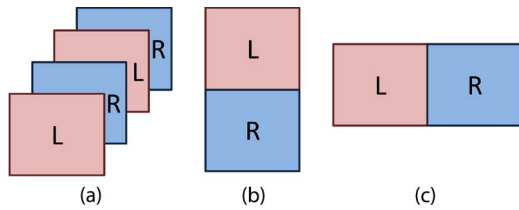
Fig. 7. *Interleaving of left and right channels (a) time multiplexing, (b) spatial multiplexing (up–down), and (c) spatial multiplexing (side-by-side).*

initial 3-D trials since both the encoding and the decoding can be done with any existing equipment. The losses of either temporal or spatial resolution as well as the reduced robustness to errors position this kind of representation as an inferior with respect to the other 3-D video representation approaches.

Recent activities of the 3DTV video group at MPEG have been focused on combining the benefits of V + D and MVC in a new 3-D video coding format so as to allow for efficient compression and rendering of multiple views on various autostereoscopic displays [131]. Extensions denoted as "*depth-enhanced stereo*" and "*multiview multi-depth*" have been considered (as also described in this special issue).

2) Wireless Channels: After the coding format selection, the next issue to investigate is the channels to be used for delivery of 3-D video to mobile devices. The delivery channels to be used depend heavily on the targeted application. Video on demand services, both for news and for entertainment applications, are already being offered over the Internet, which can be extended to 3-D. Also, 3G and 4G mobile network operators use IP successfully to offer wireless video services.

On the other hand, when the same video needs to be distributed to many users, collaboration between the users may significantly enhance the overall network performance. Peer-to-peer (P2P) streaming refers to methods where each user allocates some of its resources to forward received streams to other users; hence, each receiving user acts partly as a sending user.

At the same time, mobile TV has recently received a lot of attention worldwide with the advances in broadcasting technologies such as DMB, DVB-H, and MediaFLO [79] from one side and the 3GPP's multimedia broadcast and multicast services (MBMS) [128] from another.

Currently, there are a number of projects conducting research on transmitting 3-D video over such existing infrastructures such as the Korean 3-D T-DMB [80], the European 3-D Phone [81], Mobile3DTV [82] addressing the delivery of 3DTV to mobile users over DVB-H system, and DIOMEDES [83] addressing 3-D P2P distribution and broadcasting systems. Recently, DVB has also established

3DTV group (CM-3DTV) to identify "what kind of 3DTV solution does the market want and need, and how can DVB play an active part in the creation of that solution" [87].

As summarized in this section, there is a significant amount of work done in the various standards organizations in the area of representation, coding, and transmission of 3-D data. The most critical part is to find the optimized solution to deliver content with satisfactory quality and give the user a realistic 3-D viewing experience on a 3-D portable display. These issues will be addressed in the subsequent sections.

III. PORTABLE 3-D DISPLAYS

Three-dimensional display is the most critical part of a 3-D-enabled mobile device. It is expected to create lively and realistic 3-D sensation, meeting at the same time quite harsh limitations of screen size, spatial resolution, CPU power, and battery life. Among the wide range of state-of-the-art 3-D display technologies [13], [14], not all are appropriate for mobile use. For mobile phones or personal media players, wearing glasses or head-mounted displays to aid the 3-D perception would be rather inconvenient. Volumetric and holographic displays are far from mature for mobile use due to required size and power. Another important factor is backward compatibility—a mobile 3-D display should support both 2-D and 3-D modes and switch to the correct mode when the respective content is presented.

While selecting the enabling display technology suitable for 3-D media handhelds, autostereoscopic displays seem the most adequate choice. These displays create 3-D effect requiring no special glasses. Instead, additional optical elements are aligned on the surface of the screen (normally an LCD), to redirect the light rays and ensure that the observer sees different images with each eye [13], [15]. Typically, autostereoscopic displays present multiple views to the observer, each one seen from a particular viewing angle along the horizontal direction. The number of different views comes at the price of reduced spatial resolution and lowered brightness. In the case of small-screen, battery-driven mobile device, the tradeoff between number of views and spatial resolution is of critical importance. As mobile devices are normally watched by single observer only, two independent views are considered sufficient for satisfactory 3-D perception and good compromise with respect to spatial resolution.

A. An Overview of Portable Autostereoscopic Displays

Basically, an autostereoscopic display operates by "casting" different images towards each eye of the observer in order to create binocular cues through binocular disparity. This is done by a special optical layer, additionally mounted on the screen surface of a display formed either by liquid-crystal diodes (LED) or organic light-emitting diodes (OLED). The

additional layer controls the light passing through it by optically selecting different pixels of the conventional LCD or OLED behind it to be included in left or right view. A composite image combining the two views is rendered on the display pixels but only the (sub)pixels that belong to the correct view are visible to the corresponding eye. There are two common types of optical filters—lenticular sheet and parallax barrier.

Lenticular sheets are composed by small lenses with special shape, which refract the light to different directions [15]. The shapes are formed as cylindrical or spherical in order to enable the proper light redirection. Parallax barrier is essentially a mask with openings and closings that blocks the light in certain directions [16]. 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. Each eye sees the display from different angle and thus sees only a fraction of all pixels, precisely those meant to convey the correct (left or right) view, otherwise combined in the rendered image. The two technologies are illustrated in Fig. 8.

Both technologies have certain limitations. The viewer should be placed within a restricted area, called a sweet spot, in order to perceive 3-D image. Moving outside this proper area, the user might catch the opposite views and experience so-called pseudoscopy. Nonideal separation between views creates inter-view crosstalk manifested in ghost-like images. This effect occurs especially if the viewer is not in the optimal viewing position. As different subpixels are responsible for different-perspective images, the spatial resolution is decreased and the discrete structure of views becomes more visible. Parallax barriers block part of the light and thus decrease the overall brightness. In order to compensate for this limitation, one needs extra bright backlight, which would decrease the battery life if used in a portable device. Nevertheless, autostereoscopic displays have been the main candidates for 3-D-enabled mobile devices. Amazingly enough, some of the drawbacks of autostereoscopic displays in bigger sizes, such as lack of continuous parallax, limited

Fig. 9. *Interleaving of image for stereoscopic display on pixel level (left) and subpixel level (right).*

number of different views, and inability to serve multiple users, are reduced in their mobile counterpart versions, since typical use scenario assumes single user and no multiple views. In addition, the user can easily adjust the device so to find the correct observation angle.

Thin-film transistor (TFT) displays recreate the full color range by emitting light through red, green, and blue colored components (subpixels). Subpixels are usually arranged in repetitive vertical stripes as seen in Fig. 9. Since subpixels appear displaced in respect to the optical filter, their light is redirected towards different positions. One group will provide the image for the left eye, and another for the right eye. In order to be shown on a stereoscopic display, the images intended for each eye should be spatially multiplexed. This process is referred to as *interleaving* [1] or *interzigging* [27] and depends on the parameters on the optical filter used. Two topologies are most commonly used. One interleaves on pixel level, where odd and even pixel columns belong to alternative views. The other interleaves on a subpixel level—where subpixel columns belong to alternative views. In the second case, different-color components of one pixel belong to different views.

The first display for a mobile phone was announced by Sharp Laboratories of Europe in 2002 [17]. Since then a few vendors announced prototypes of 3-D displays, targeted for mobile devices [18]–[20]. All of them are two-view, TFT-based autostereoscopic displays. The display produced by Sharp uses electronically switchable reconfigurable parallax barrier, working on subpixel basis [17]. The interzigging topology is similar to the one of Fig. 9(left). Each view is visible from multiple angles, and the angle of visibility of one view is rather narrow, making the visual quality of the 3-D scene quite sensitive to the observation angle.

Another 3-D-LCD module based on switchable parallax barrier technology has been produced by Masterimage [20]. It is 4.3″ WVGA autostereoscopic display that can operate in 2-D or 3-D mode. The parallax barrier of the 3-D LCD module can be switched between "3-D horizontal" and "3-D vertical" mode, allowing it to operate in landscape 3-D or portrait 3-D mode. The barrier operates on pixel level.

From the group of displays based on lenticular lenses, we refer to two prototypes, delivered by Ocuity Ltd. (2001–2010), Oxford, U.K. and NEC LCD Technologies Ltd., Kawasaki, Japan, respectively. The reconfigurable 2-D/3-D technology by Ocuity Ltd. uses a polarization activated
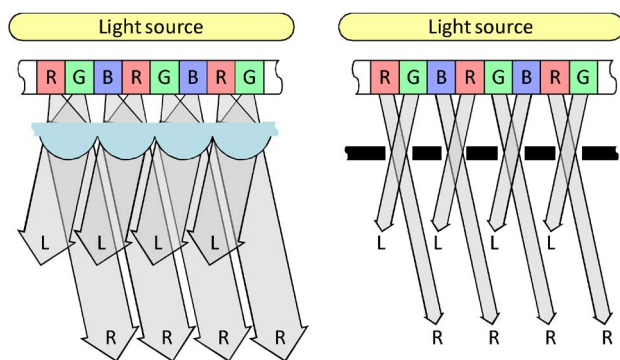

Fig. 8. *Light redirecting in autostereoscopic displays: lenticular sheet (left) and parallax barrier (right).*
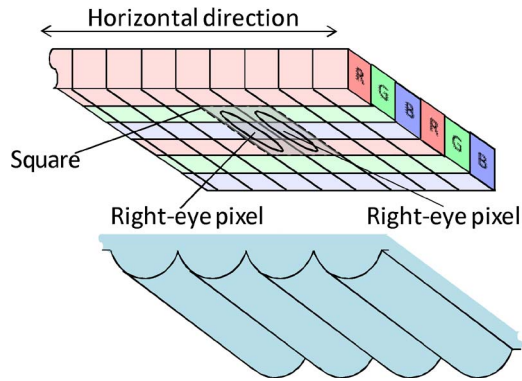
Fig. 10. *HDDP pixel arrangement.*

Fig. 11. *Three-dimensional film-based display.*

microlens array [19]. The microlens array is made from a birefringent material such that at the surface of the lens there is a refractive index step for only one of the polarizations.

The WVGA 3-D LCD module with horizontal double-density pixel (HDDP) structure as developed by NEC Central Research Laboratories uses NEC's proprietary pixel array for stereoscopic displays [18]. The HDDP structure is composed of horizontally striped RGB color subpixels; each pixel consists of three subpixels that are striped horizontally and split in half lengthwise. As a result, horizontal resolution is doubled compared to 3-D LCD modules constructed with vertically striped pixels, and 3-D images are produced through data for the right eye and data for the left eye being alternately displayed horizontally by pixel. Moreover, 2-D images may also be displayed when the same data are presented for adjacent pixels. Since the LCD module can display both 3-D and 2-D images at the same resolution, it can display a mixture of 2-D and 3-D images simultaneously on the same screen without causing discomfort to viewers. The pixel arrangement is illustrated in Fig. 10.

Last display we overview is produced by 3M, St. Paul, MN. It is based on patterned retardation film, which distributes the light into two perspective views in a sequential manner. The display uses a standard TFT panel operating at 120 Hz with special type of backlight. It is composed of two sources of light: a lightguide and 3-D film between the LCD and the lightguide. The construction is shown in Fig. 11.

The two backlights are turned on and off in counter phase so that each backlight illuminates one view. The switching is synchronized with LCD, which displays different-perspective images at each backlit switch-on time. The role of the 3-D film is to direct the light coming from the activated backlight to the corresponding eye.

B. Optical Parameters of Portable Autostereoscopic Displays

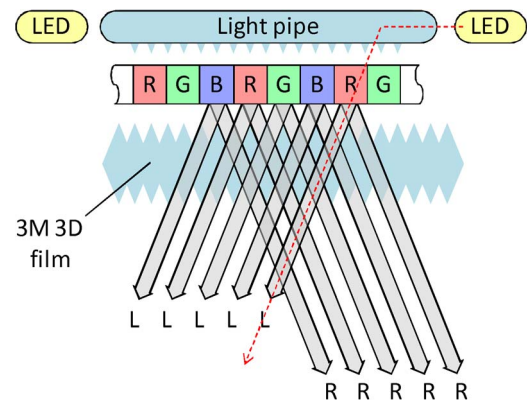Various optical parameters can be used for characterizing the quality of autostereoscopic 3-D displays. The set of parameters includes *angular luminance profile* [21], *3-D crosstalk and luminance uniformity* [22], *viewing freedom*, pixel "*blockiness*" and "*stripiness*" [23] as well as angular measurements in Fourier domain [24]. Visual appearance of a 3-D scene also depends on external factors, such as observation distance, ambient light, and scene content. Therefore, for comparing the visual quality of autostereoscopic displays, one should select the subset of perceptually important optical characteristics.

Crosstalk is perhaps the single most important parameter affecting the 3-D quality of autostereosopic displays. For autostereoscopic displays, crosstalk can be calculated as the ratio χ_{3D} of visibility of one view to the visibility to all other views [22]. A number of studies investigated how the level of crosstalk affects the perceptibility of stereoscopic 3-D scenes [25], [31], [40]. According to [25], crosstalk of less than 5% is undistinguishable and crosstalk over 25% severely reduces the perceptual quality. To characterize the influence of crosstalk, one can regard the visibility on the horizontal plane passing through the center of the display, the so-called *transverse plane* [24]. For autostereoscopic 3-D displays with no eye tracking, both the luminance of a view and crosstalk between views are functions of the observation angle with respect to that plane, as shown in Fig. 12(a). For each point on the display surface, there are certain observation angles, where the crosstalk is low enough to allow 3-D perception with sufficient quality. The positions at which one view is seen across the whole display surface have diamond-like shapes on the transverse plane and are called *viewing diamonds* [22], [23]. The areas inside the viewing diamonds where the crosstalk is sufficiently low are the *sweet spots* of the views [23]. In Fig. 12, areas marked with "I" and "III" are the sweet spots of the left and right views correspondingly. A crosstalk level $\chi_{3D} < 25\%$ can be used to define the sweet spots of the views.

A set of mobile 3-D displays is listed in Table 1. The HDDP device uses display with HDDP pixel arrangement [18]. The MI(P) and MI(L) devices use switchable parallax
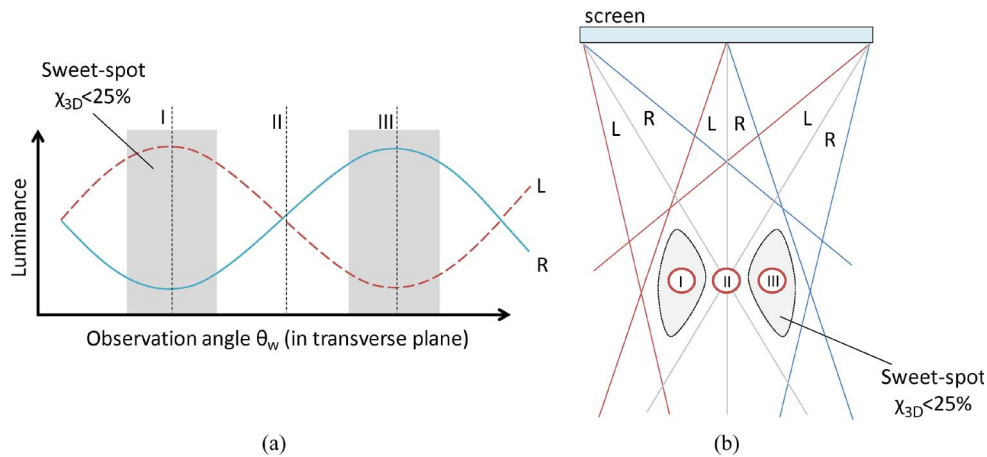
Fig. 12. *(a) Angular luminance profile of two-view autostereoscopic display and (b) its viewing diamonds.*

Table 1 Devices With 3-D Displays Used in the Measurement Tests

Label	Model	Type	OVD, cm
HDDP	NEC HDDP prototype	3.2" display based on the lenticular HDDP technology by NEC	40
MI (L)	MB403M0117135 – landscape mode	Mobile 3D display with parallax barrier, switchable between landscape and portrait mode	37
MI (P)	MB403M0117135 – portrait mode		
FF	FinePix REAL 3D V1	2D Photo Frame with parallax barrier	46
FC	FinePix REAL 3D W1	Consumer 3D Camera with 3D display	40
SL	Sharp AL3DU	3D laptop with parallax barrier	58
AL	Acer AS5738DG-6165	3D laptop with polarized glasses	60 (nominal)

barrier display interleaved on pixel level, operating in portrait and landscape modes correspondingly [20]. The FF [26] and SL [17] devices use switchable parallax barrier interleaved on subpixel level. The FinePix camera, designated as FC, uses time-sequential 3-D-film-based display [26]. As an alternative, measurement results for a row-interleaved, polarization-multiplexed 3-D display with glasses (AL) are presented in the last row of the table.

Due to imperfect display optics the views are never fully separated, and even in the sweet spots some residual crosstalk exists. This effect is referred to as *minimal crosstalk*, and its value determines the visual quality of the display for the optimal viewing angle and distance. The minimal crosstalk for all measured devices is given in Fig. 13. The HDDP display has the lowest crosstalk ($\chi_{3D} = 4\%$), and thus has the best overall quality among the compared displays. On the FinePix 3-D display (FC), the crosstalk measurements consistently reached over 30%, manifested in double edges visible at all times, though stereoscopic perception was still possible. Notably, the AL display performs better when watched with its original glasses ($\chi_{3D} = 24\%$) than when watched with another pair of general purpose polarized glasses ($\chi_{3D} = 29\%$).

For most autostereoscopic 3-D displays the stereoscopic effect can be seen within a limited range of observation distances. The *visibility range* of a 3-D display is defined as the range, for which both eyes of the observer would fall into view sweet spot simultaneously. It is limited by the
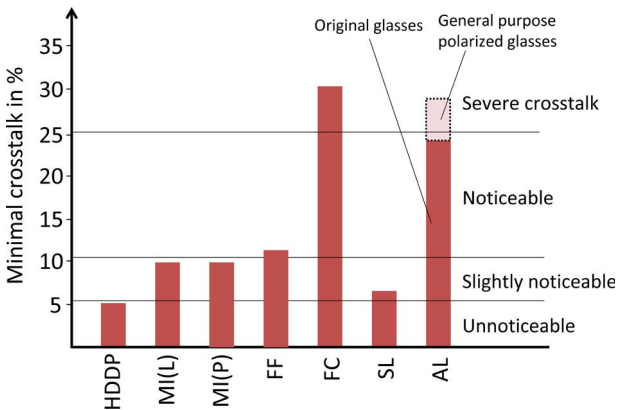

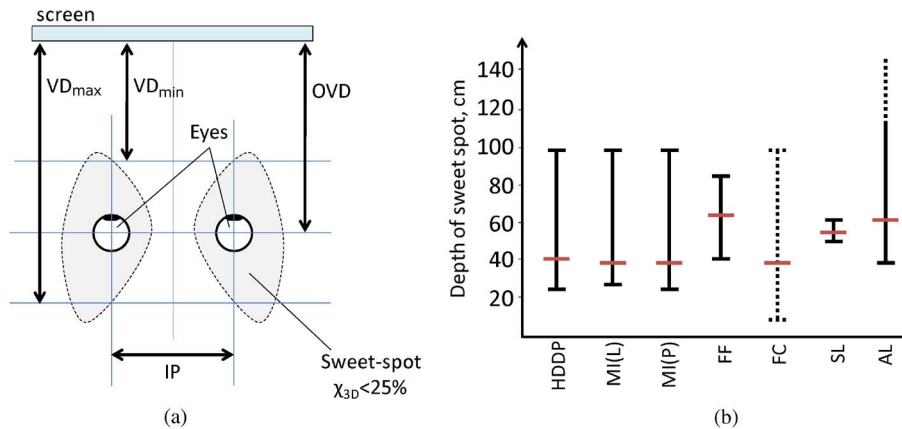**Fig. 13.** *Minimal crosstalk for various mobile 3-D displays.*

Fig. 14. *(a) Definition of OVD, VD_{min}, and VD_{max} values. (b) Measured values for various 3-D displays.*

minimum and maximum viewing distances VD_{min} and VD_{max} [cf. Fig. 14(a)] while at the optimal viewing distance (OVD) the sweet spot has typically the largest width. Usually at this distance the display has the lowest overall crosstalk as well. Since the sweetspots have nonsymmetric shape, the interpupilar distance (IPD) of the observer affects the VD_{min} and VD_{max} values. Comparative results for IPD = 65 mm and $\chi_{3D} < 25\%$ are given in Fig. 14 (see also the measured OVD values in Table 1). Since the minimal crosstalk of FC display is always over 30%, from herein it is represented with dashed line, for distances where 30% $< \chi_{3D} < 50\%$. The AL display does not have either optimal or maximal viewing distance in terms of crosstalk. For that display, the OVD is the nominal observation distance as suggested in the display manual.

We define the *width of sweet spot* as all angles on the transversal plane, where each eye of the observer perceives the correct view (i.e., not reverse stereo) with crosstalk $\chi_{3D} < 25\%$. The lateral sweet spot width can be measured in distances, as in [22] and [23]. However, assuming that the

observer is always at the optimal distance from the center of the display, the ranges can be measured also in angles, as illustrated in Fig. 15(a). This is done as it is more likely that the user of a mobile display is holding it at a constant distance, and is turning it in order to get the best view. Typical results for IPD = 65 mm are given in Fig. 15(b). Among all autostereoscopic displays tested, HDDP has the widest sweet spots, which makes it the easiest for the user to find a correct observation angle. On the contrary, the MI display has narrow sweet spots and users must hold it at a precise angle to be able to perceive stereoscopic effect. The AL display used with glasses delivers continuous 3-D effect over a wide range of observation angles.

The *sweet spot height* is measured as the range of observation angles in the plane passing through the center of the display (also known as sagittal plane), where observers' eyes perceive correct stereo with $\chi_{3D} < 25\%$. The user is assumed to be at the display's OVD, as shown in Fig. 16(a). The measurement results for IPD = 65 mm are given in Fig. 16(b). Most autosteoscopic displays have vertical observation range of $-30°$ to $30°$. Interestingly enough, the
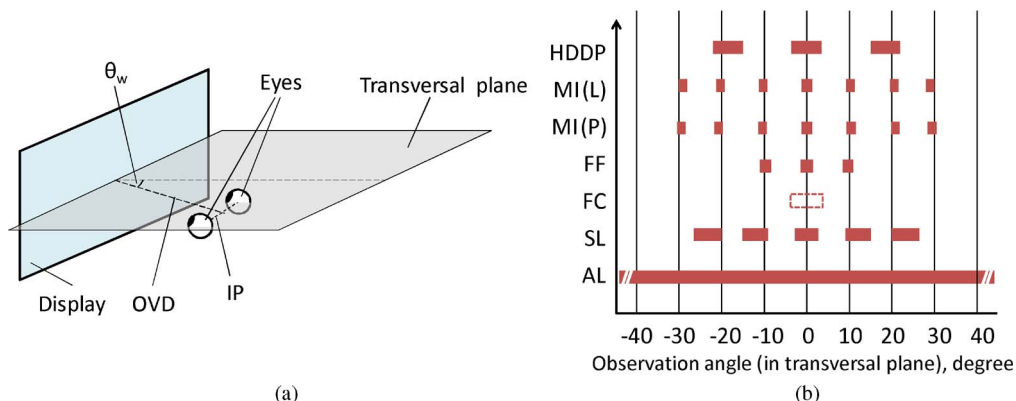

Fig. 15. *(a) Measurement of sweet spot width. (b) Sweet spot widths for various mobile 3-D displays.*
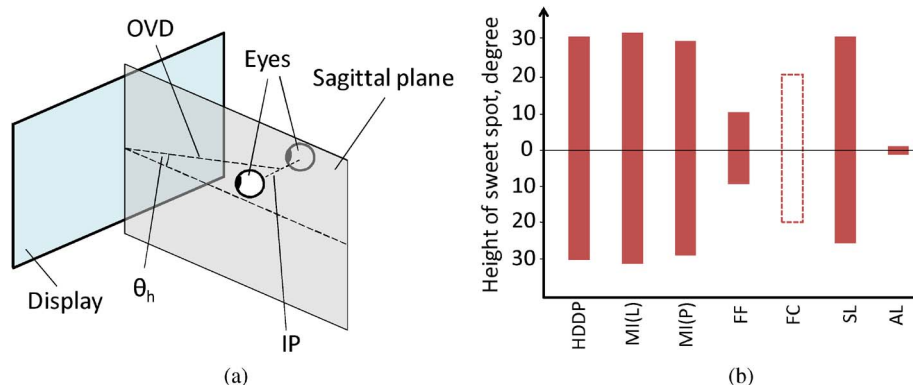
Fig. 16. *(a) Measurement of sweet spot height. (b) Sweet spot heights for various mobile 3-D displays.*

AL display is very sensitive to the vertical angle, and has a sweet spot height of −2° to 2°. In fact, this is the limiting factor defining the minimum observation distance for that display.

In contrast to 2-D displays, where the user is free to choose the observation distance, autostereoscopic 3-D displays deliver best results when observed at their OVDs. Since OVD varies from display to display, it is more suitable to compare angle-of-view (AOV) and angular resolution, rather than the absolute size and resolution of such displays. The area, which each display occupies in the visual field, when observed from its optimal observation distance, is given in Fig. 17(a). Next to each display is given its OVD. The angular size of all displays, observed at their OVD is given in Fig. 17(b). For MI, FF, and SL displays, both results for 2-D and 3-D modes are given as the resolutions are different. For comparison, the angular resolutions for the displays of two popular handhelds, Nokia N900 and Apple iPhone4, at 40-cm observation

distance are given. The theoretical angular resolution of the human retina (50CPD) is calculated for perfect 20/20 eyesight. Fig. 17 is instructive about the fact that 2-D and 3-D displays have comparable AOV but different angular resolution. Especially the horizontal angular resolution of mobile 3-D displays is much lower than the one of a typical mobile 2-D display.

IV. USER EXPERIENCE OF 3-D MEDIA FOR MOBILES

User experience seems to be the key factor for the adoption of the mobile 3-D media technology, as having a perceptually acceptable and high-quality 3-D scene on a small display is a challenging task. According to the holistic user-centered research framework, as formulated in Section II-B, research efforts have focused on optimizing the technology components, such as content creation and coding techniques, delivery channels, portable 3-D
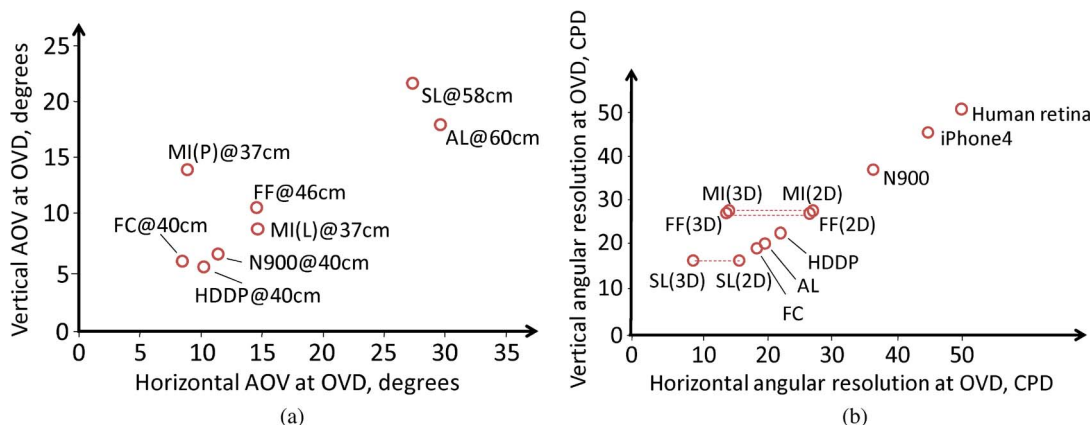

Fig. 17. *Angular size and angular resolution of various mobile 3-D displays: (a) angular size observed from OVD, in degrees; (b) angular resolution observed from OVD, in cycles per degree. Note: N900 and iPhone4 are 2-D displays given for comparison, as they appear at 40-cm observation distance.*

displays, and media-rich embedded platforms to deliver the best possible visual output. In this section, the 3-D media user experience is addressed methodologically by an interdisciplinary approach having threefold goals. First, the artifacts, which arise in various usage scenarios involving stereoscopic content, are analyzed and categorized so as to put them against the peculiarities of the human visual system and the way users perceive depth. Then, critical parts of the system, such as coding and transmission approaches, are studied for their performance both through objective comparisons and subjective tests so as to reveal the levels of acceptance and satisfaction of the new content and services. Eventually, 3-D graphical user interfaces complement the experience of 3-D media content.

A. Three-Dimensional-Specific Artifacts

Stereoscopic artifacts can be described with respect to the stage in the 3-D media delivery chain, as exemplified in Fig. 5 and how they affect different "layers" of human 3-D vision. In this way, artifacts can be clustered in a multidimensional space according to their source and *structure, color, motion, and binocular* "layers" of HVS, interpreting them. These layers roughly represent the visual pathways as they appeared during the successive stages of evolution. The *structure* layer denotes the spatial and colorless vision.

It is assumed that during the evolution human vision adapted for assessing the "structure" (contours and texture) of images [35], and some artifacts manifest themselves as affecting image structure. *Color* and *motion* layers represent the color and motion vision, correspondingly. The *binocular* layer denotes artifacts meaningful only when perceived in a stereo pair, and not by a single eye (e.g., vertical disparity). The result of multidimensional clustering is well illustrated by a circular diagram in polar coordinates given in Fig. 18 [39]. Such a wide nomenclature of clustered artifacts helps in identifying the stages at which they should be properly tackled. While some of the artifacts are less important in mobile context, some are quite typical and influential for the acceptance of the technology.

1) Artifacts Caused at Creation/Capture Stage: The most common and annoying artifact introduced in the process of capture or rendering a stereoscopic image is *unnatural disparity* between the images in the stereo pair. Special care should be taken when positioning cameras or when selecting rendering parameters and rectification is a standard preprocessing stage. However, often a perfectly rectified stereoscopic image needs to be visualized at different size than the originally captured one. Changing the size or resolution of stereoscopic pair can also introduce unnatural
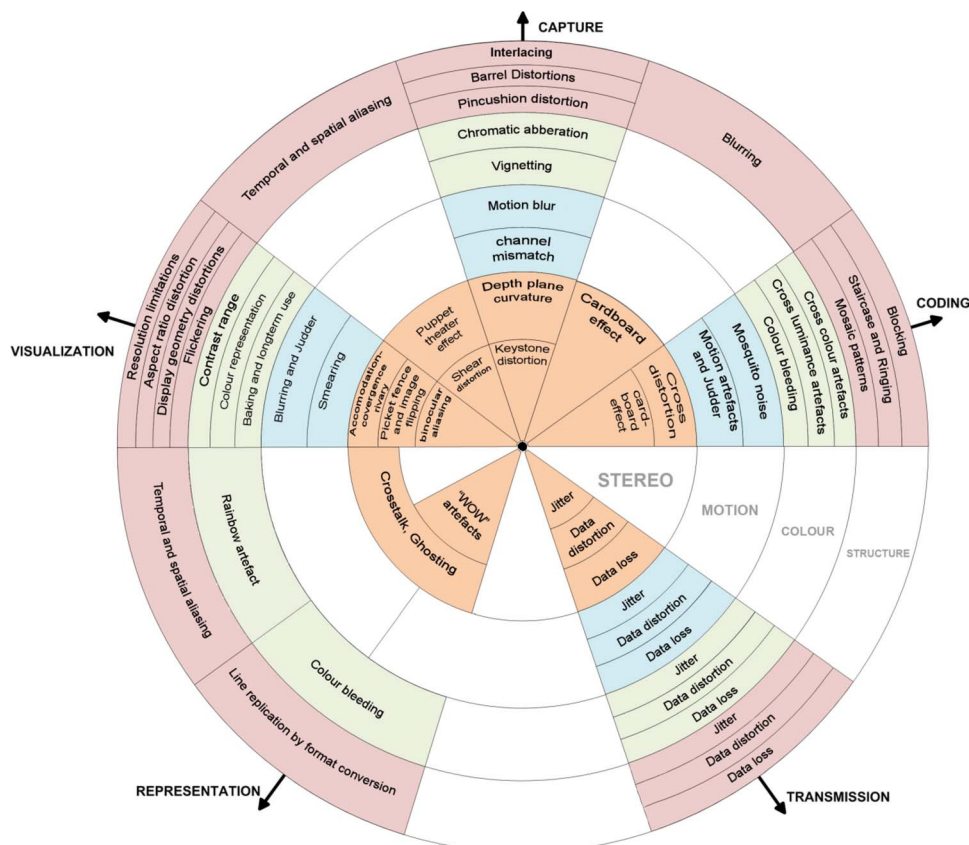

Fig. 18. *Artifacts caused by various stages of content delivery and affecting various "layers" of human depth perception [39].*
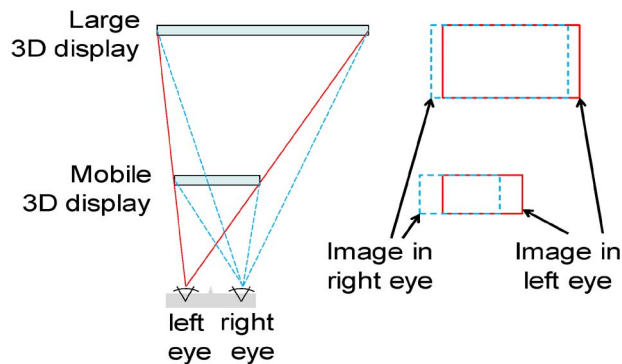
Fig. 19. *Change of relative disparity while rescaling stereoscopic image pair.*

disparity. When resizing a stereoscopic pair, the relative disparity is scaled proportionally to the image size. However, as the interocular distance remains the same, observing a closely positioned mobile 3-D display would require different relative disparity range compared to when observing large 3-D display placed further away. The effect is illustrated in Fig. 19. Even if the mobile and large 3-D displays have the same visual size, stereoscopic images on them have different disparity.

Two-channel stereo video, and video plus dense depth are the likely contenders for 3-D video representation for mobiles [1]. If the representation format is different from the one in which the scene has been originally captured, converting between the formats is a source of artifacts. A typical example is the occlusions areas in depth-from-stereo type of conversion.

2) Coding Artifacts: Various coding schemes utilize temporal, spatial, or interchannel similarities of a 3-D video [2]. Algorithms originally designed for single-channel video, might be improperly applied for stereo video, and important binocular depth cues might be lost in the process. The block discrete cosine transform (DCT), which is

in core of most compression video compression algorithms is a source of *blocking artifacts*. They are thoroughly studied for 2-D video, but their effect on stereoscopic quality is yet to be determined. Some authors propose that blocking might be considered as several, visually separate artifacts—*block-edge discontinuities*, *color bleeding*, *blur*, and *staircase artifacts* [35], [36]. Each of these artifacts introduces different amount of impairments to object edges and texture. The human brain has the ability to perceive single image by combining the images from left and right eyes (a so-called cyclopean image) [33]. As a result, the same level of DCT quantization might result in different perceptual quality, based on the depth cues present in a stereo image. In Fig. 20, both channels of a stereo pair are compressed with the same quality factor. When an object appears on the same place in both frames, it is equally affected by blocking in each frame, and the perceived cyclopean image is similar to the one shown in Fig. 20(a). When the object has different horizontal positions in each frame, the blocking artifacts will affect differently the object in each frame, which results in a cyclopean image similar to the one in Fig. 20(b).

3) Transmission Artifacts: In the case of digital wireless transmission a common problem is packet losses. Related artifacts are sparse and highly variant in terms of occurrence, duration, and intensity. At very low bit rates they may be masked by compression impairments. The presence of artifacts depends very much on the coding algorithms used and how the decoder copes with the channel errors. In the DVB-H transmission, the most common are burst errors, which result in *packet losses* distributed in tight groups [55]. In MPEG-4-based encoders, packet losses might result in *propagating* or *nonpropagating* errors, depending on where the error occurs with respect to key frames, and the ratio between key and predicted frames. Error patterns of wireless channels can be obtained with field measurements, and then used for simulation of channel losses [55], [56]. In multiview video encoding, where one channel is predicted from the other, usually error burst is long enough to affect both channels [57]. In that
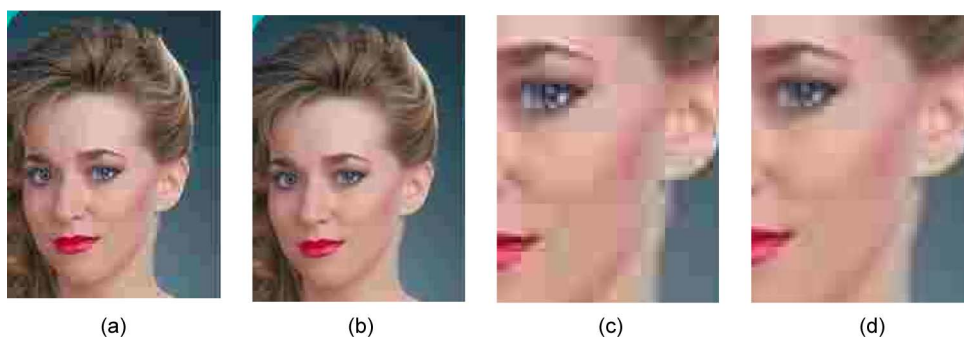

Fig. 20. *The impact of blocking on stereo pairs with different disparity: (a) q = 15, disparity = 0; (b) q = 15, disparity = 4; (c) zoomed detail of (a); (d) zoomed detail of (b).*

Fig. 21. *Packet loss artifacts affecting multiview encoded stereoscopic video [57].*

case, packet loss artifacts appear on the same absolute position in both images even though the appearance in one channel is mitigated due to the prediction. Fig. 21 illustrates the effect for the case of TU6 channel with channel signal-to-noise ratio (SNR) = 18 dB [57]. In the format V + D using a separate depth channel, usually depth is encoded in much lower bitrate than the video. In that case, burst errors affect mainly the video channel, and the relative perceptual contribution of depth map degradation alone is very small.

One common artifact introduced during receiving and decoding of 3-D video is *temporal mismatch*, where one channel gets delayed with respect to the other. It might be caused by insufficient memory or CPU, or error concealment in one channel. The outcome is that the image from one channel does not appear with a simultaneously taken image from the other channel, but with an image that is taken a few frames later. Even temporal mismatch of as low as two frames can result in a stereoscopically inadequate image pair. For comparison, two images are shown in Fig. 22—the left image is done by superimposing frame 112 from left and right channels of a movie; the right image is done by superimposing frame 112 from the left channel and frame 115 from the right channel of the same movie.

4) Visualization and Display Artifacts: Even a perfectly captured, transmitted, and received stereoscopic pair can exhibit artifacts due to various technical limitations of the

autostereoscopic display in use [58]–[60]. The most pronounced artifact in autostereoscopic displays is crosstalk, caused by imperfect separation of the "left" and "right" images and is perceived as *ghosting artifacts* [27]. Two factors affect the amount of crosstalk introduced by the display—position of the observer and quality of the optical filter in front of the LCD, as discussed in Section III-B. Due to the size of the subpixels, there is a range of observation positions, from where some subpixels appear partially covered by the parallax barrier, or are partially in the focal field of the corresponding lenticular lens. This creates certain *optimal observation spots* in the centers of the sweet spots, where the two views are optimally separated [the areas marked with I and III in Fig. 12(b)], and transitional zone (marked with II) where a mixture of the two is seen. However, even in the optimal observation spot one of the views is not fully suppressed—for example, part of the light might "leak" through the parallax barrier as shown in Fig. 23(a) and create the minimal crosstalk effect discussed in Section III-B.

Fig. 22. *Temporal mismatch in stereo video. Left: superimposed images of temporally synchronized stereo pair. Right: superimposed images of stereo pair with three frames temporal mismatch.*
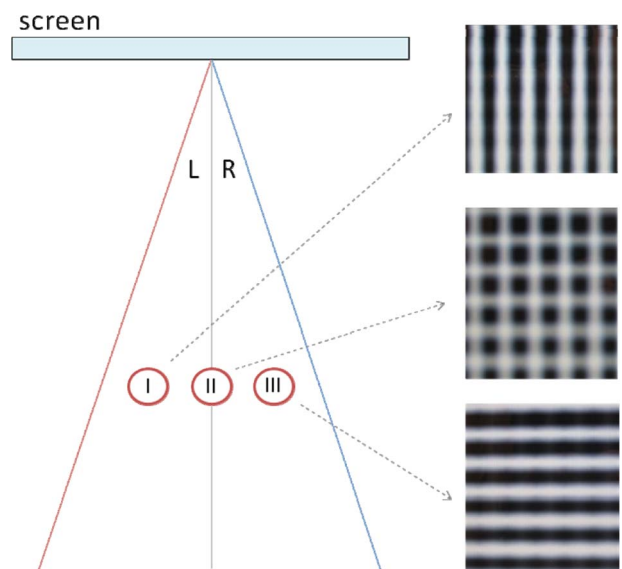

Fig. 23. *Effect of crosstalk in portable 3-D displays; from left to right: photographs taken of a 3-D display from positions I, II, and III.*

The effect is well illustrated by a special test stereoscopic pair, where the "left" image contains vertical bars, and the "right" image contains horizontal bars. This stereo pair has been visualized on a parallax-barrier-based 3-D display, and photographed from observation angles as marked with I, II, and III in Fig. 12(a). The resulting photos are shown in Fig. 23(c)–(e). Both position-dependent and minimal crosstalk effects can be seen. By knowing the observation position and the amount of crosstalk introduced by the display, the effect of crosstalk can be mitigated by precompensation [133].

There are darker gaps between subpixels of an autostereoscopic display. They are more visible from certain angles than from others. When an observer moves laterally in front of the screen, he perceives this as luminance changes creating brighter and darker vertical stripes over the image. Such effect is known as *banding artifacts* or *picket fence effect* and is illustrated in Fig. 24. The effect can be reduced by introducing a slant of the optical filter with respect to the pixels on the screen [15]. Tracking of the user position with respect to the screen can also help in reducing these artifacts.

Parallax-barrier and lenticular-based 3-D displays with vertical lenses arrangement have horizontal resolution twice lower than vertical one as only half of the subpixels of a row form one view. This arrangement requires spatial subsampling of each view, before both views are multiplexed, thus risking introducing aliasing artifacts. In 3-D displays, aliasing might cause false color or Moiré artifacts (illustrated in Fig. 25) depending on the properties of optical filter used. Properly designed prefilters should be used, in order to avoid aliasing artifacts.

Autostereoscopic displays that use parallax barrier usually have a number of interleaved "left channel" and "right channel" visibility zones, as shown in Fig. 26. Such display can be used by multiple observers looking at the screen at different angles, for example, positions marked with "1" and "2" in the figure. However, an observer in position "3" will perceive *pseudoscopic* (also known as *reversed stereo*) image. For one observer, this can be avoided by using face tracking and algorithm that swaps the "left"
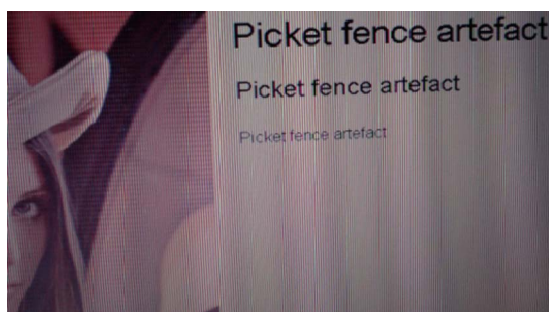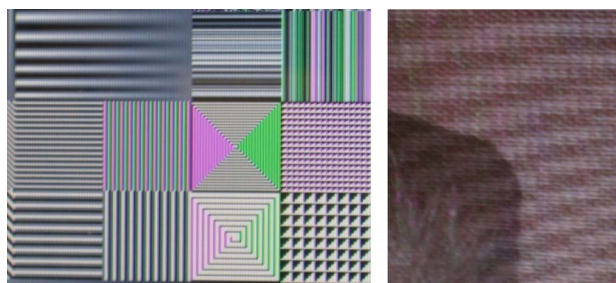

Fig. 25. *Aliasing in autostereoscopic displays. Left: false color. Right: Moiré artifacts.*

and "right" images on the display appropriately to accommodate to the observers viewing angle.

B. Optimized Delivery Channel

1) Evaluation of Coding Methods: The methods for 3-D video coding described in Section II-C contain a multitude of parameters that vary their performance in different scenarios. As all methods are based on H.264 AVC, the profiles of the latter (i.e., baseline, main, extended, and high profiles), its picture type (I, P, and B), and entropy coding methods (CABAC or CAVLC) determine the varying settings to be tested for mobile use [72].

In [73], candidate stereoscopic encoding schemes for mobile devices have been investigated for both encoding and decoding performance. Rate-distortion curves have been used to assess the coding efficiency and decoding speed tests have been performed to quantify the decoder complexity. It has been concluded that, depending on the processing power and memory of the mobile device, the following two schemes can be favored: H.264/AVC MVC extension with simplified referencing structure and H.264/AVC monoscopic codec with IPP + CABAC settings over interleaved stereoscopic content.
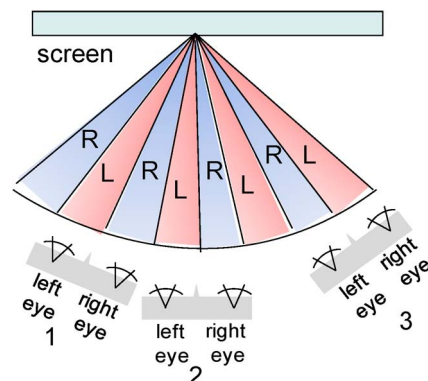

Fig. 24. *Banding/picked fence artifacts.*

Fig. 26. *True stereoscopic (1 and 2) and pseudoscopic (3) observation positions.*

Table 2 Target Bitrates for Different Quality Levels and Test Contents

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

In [74], H.264/AVC simulcast, H.264 stereo SEI message, H.264/MVC, MPEG-C Part 3 using H.264 for both video and depth and H.264 auxiliary picture syntax for video plus depth have been compared for their performance in mobile setting. A set of test videos with varying types of content and complexity have been used. The material has been coded at different bitrates using optimum settings for each of the aforementioned encoders. The quality has been evaluated by means of peak signal-to-noise ratio (PSNR) over bitrate. The results show that the overall rate-distortion (RD) performance of MVC is better than simulcast coding. It has also been shown that the overall RD performance of video plus depth is better than stereo video with simulcast coding.

The selection of an optimum coding method has recently been addressed in two publications by Strohmeier and Tech [108], [111] based on the results from subjective tests. Four different coding methods that had been adapted for 3-D mobile television and video were evaluated. H.264/AVC simulcast [120], H.264/MVC [121], and MRSC [114], [115] using H.264/AVC were chosen as coding methods for a video + video approach. Video plus depth coding using MPEG-C Part 3 [122] and H.264/AVC as a video + depth approach completed the coding methods under assessment. The depth maps of the test sequences were obtained using the hybrid-recursive-matching algorithm, described in [134]. The virtual views were rendered following the approach described in [135]. To further decrease the coding complexity with regard to limited calculation power of current mobile devices, the baseline profile was used for encoding. This includes a simplified coding structure of IPPP and the use of CAVLC. Six different contents were encoded at two different quality levels. To determine the different quality levels, the quantization parameters (QPs) of the encoder for simulcast coding were set to 30 for the high quality and 37 for the low quality. From these sequences, target bit rates for the other methods were derived and used in the test set creation, respectively. Table 2 presents the target bitrates for different quality levels and contents.

The test items were evaluated by 47 test participants. The evaluation followed the absolute category rating (ACR) [102] and test participants evaluated acceptance of (yes/no) and satisfaction with (11-point-scale) perceived overall quality [99]. The test items were presented using a NEC HDDP 3.5″ mobile display [123] with a resolution of 428 × 240 pixels.

All coding methods under test provided a highly acceptable quality at the high-quality level of 80% and higher. At the low-quality level, MVC and V + D still got an acceptance score of 60% and higher. Strohmeier and Tech [108] showed in their study that MVC and the video + depth provide the best overall quality satisfaction for both quality levels (see Fig. 27). These coding methods significantly outperform MRSC and simulcast. With respect to the different test contents the results show that coding methods show content-dependent performance. Video + depth gets the highest overall satisfaction scores for Car, Mountain, and Soccer2. MVC outperforms all other coding methods for content Butterfly.

The results of this study were extended in a follow-up study by Strohmeier and Tech [111]. While the first study was limited to the use of low coding complexity, the second study used the complex high profile, which enables hierarchical B-frames and CABAC. The other parameters, quality levels, test contents, and device were the same so that the follow-up study [111] allowed a direct comparison of the results of baseline and high profile. Forty participants evaluated the test set of high profile.

The results of the overall quality evaluation for the high-profile sequences confirmed the findings of the baseline sequences (see Fig. 28). The test items at the high-quality level got an overall quality acceptance score of at least 75%. For the low-quality level, MVC and video + depth reach an acceptance level of 55% and more. As in the baseline case, MVC and video + depth also outperform the other coding methods in terms of satisfaction with overall quality. The content-dependent results for the provided overall quality for all coding methods were shown in the results as well.

Finally, the results of both studies allowed to directly comparing the performance of baseline and high profiles (see Fig. 29). Although the results show small differences for baseline and high codec profiles for some settings, the overall view on the results shows no differences among the two profiles. However, significantly lower bit rates can be realized for the high profile due to more efficient,
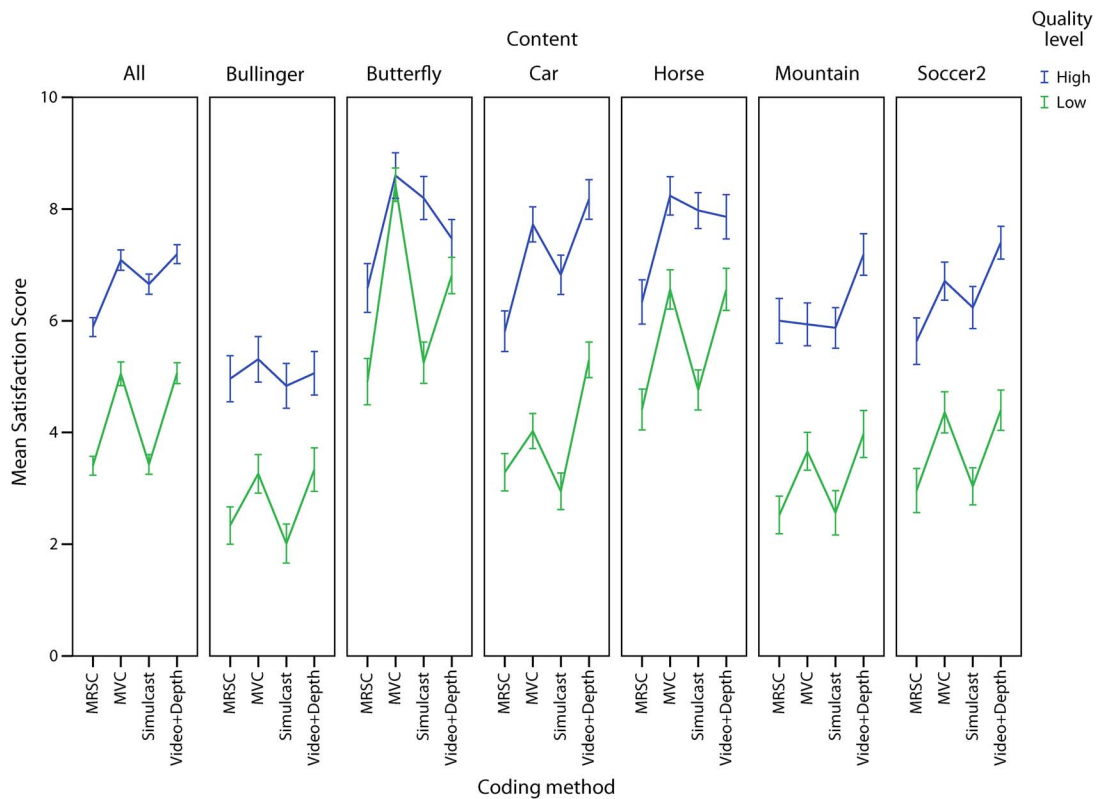
Fig. 27. *Mean satisfaction scores for different coding methods at baseline profile averaged over contents (all) and content-by-content given at high- and low-quality levels. Error bars show 95% confidence interval (CI) of mean.*

though more complex, coding structures. Altogether, Strohmeier and Tech [111] showed that the use of high coding profile, i.e., hierarchical B-frames and CABAC, can provide the same experienced quality as baseline profile using lower bit rates. This can result in advantages for the transmission of these sequences in terms of better error resilience [124].

2) Evaluation of Transmission Approaches: In order to illustrate the effects of channel characteristics on the received video quality, a typical 3-D broadcasting system is simulated as shown in Fig. 30 [85]. In this study, DVB-H is used as the underlying transmission channel. DVB-H is the extension of DVB project for the mobile reception of digital terrestrial TV. It is based on the existing DVB-T physical layer with introduction of two new elements for mobility: MPE-FEC and time slicing. Time slicing enables the transmission of data in bursts rather than a continuous transmission; explicitly signaling the arrival time of the next burst in it so that the receiver can turn on between and wake up before the next burst arrives. By this way the power consumption of the receiver is reduced. Multiprotocol encapsulation is used for the carriage of IP datagrams in MPEG2-TS. IP packets are encapsulated to MPE sections each consisting of a header, the IP datagram as a payload, and a 32-b cyclic redundancy check

(CRC) for the verification of payload integrity. On the level of the MPE, an additional stage of forward error correction (FEC) can also be added. This technique is called MPE-FEC and improves the C/N and Doppler performance in mobile channels. To compute MPE-FEC, IP packets are filled into an $N \times 191$ matrix where each square of the matrix has one byte of information and N denotes the number of rows in the matrix. The standard defines the value of N to be one of 256, 512, 768, or 1024. The datagrams are filled into the matrix columnwise. Error correction codes (RS codes) are computed for each row and concatenated such that the final size of the matrix is of size $N \times 255$. To adjust the effective MPE-FEC code rate, padding or puncturing can be used. Padding refers to filling the application data table partially with the data and the rest with zero whereas puncturing refers to discarding some of the rightmost columns of the RS-data table.

In the simulated system, 3-D video content is first compressed with a 3-D video encoder, operating in one of the modes: MVC, V + D, or simulcast. Resulting network abstraction layer (NAL) units (NALU) are fed to the stereo video streamer. The packetizer encapsulates the NAL units into real-time transport protocol (RTP) [84] mono-compatible only, user datagram protocol (UDP), and finally, internet protocol (IP) datagram for each view separately. The resulting IP datagrams are encapsulated in
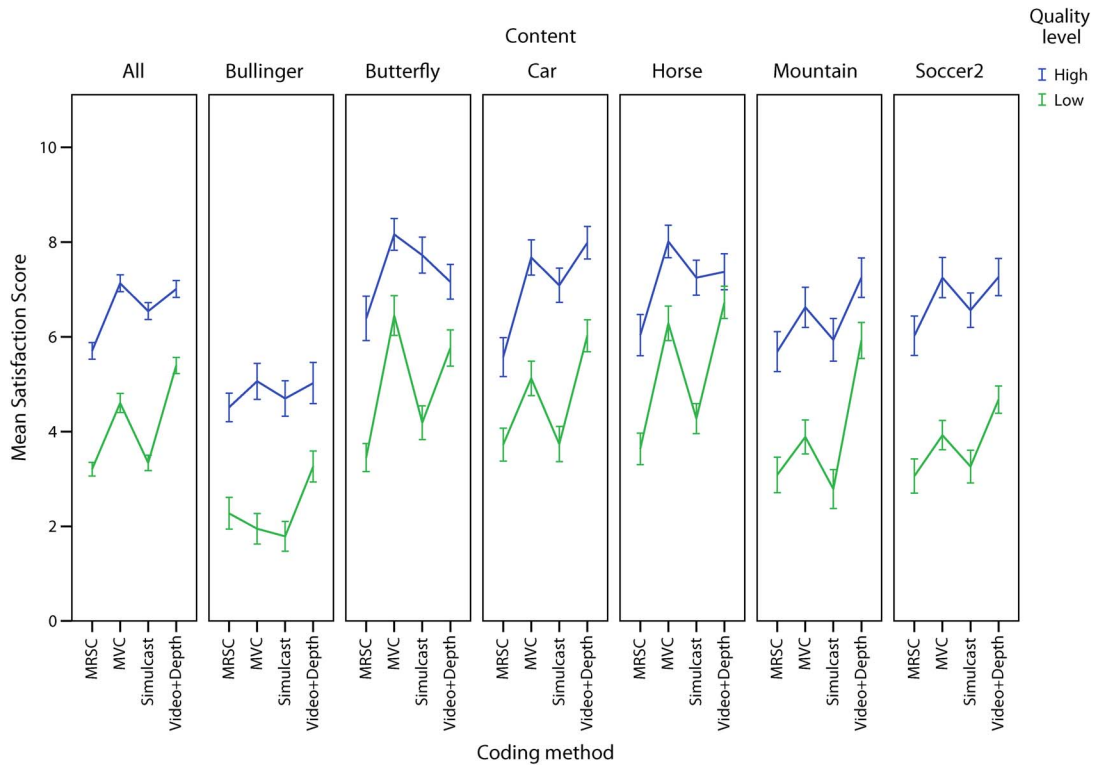
Fig. 28. *Mean satisfaction scores for different coding methods at high profile averaged over contents (all) and content-by-content given at high- and low-quality levels. Error bars show 95% CI of mean.*

the DVB-H link layer where the multiprotocol encapsulation forward error correction (MPE-FEC) and time slicing occurs [86]. Through the MPE-FEC mechanism, IP datagrams are protected by adding additional bytes for a variable-length Reed–Salomon (RS) coding. MPE-FER rate refers to the ration between application and total data. Time slicing allows sending the packets into time slices (bursts) for better power consumption at the receiver site. Different views are assigned different packet identifiers and encapsulated as different elementary streams. Therefore, they are transmitted in different time slices or bursts. The link layer output MPEG-2 transport stream (TS) packets are passed to the 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 modulator. The received stream is decoded using the section erasure method, i.e., the MPE-FEC frame is filled with contents of the error-free MPE and MPE-FEC sections and the empty bytes in the frame are marked as erasures, RS decoding is performed to reconstruct the lost data, and finally, the received and correctly reconstructed IP datagram are passed to the video client. IP datagram are handled in the depacketizer and resulting NAL units are decoded with the stereo video decoder to generate right

and left views. Finally, these views are combined with a special interleaving pattern to be displayed in stereo 3-D on an autostereoscopic display.

Within the Mobile3DTV project, extensive sets of tests have been performed in order to find an effective compromise between compression efficiency, FEC-code rates, and robustness with respect to typical channel conditions [88]. Simulations have been carried out involving 3-D video content with different characteristics as described in Table 3 and coded as simulcast, V + D, and MVC simplified structure. For all the tests, JMVC 5.05 (in monoscopic mode for simulcast) is used with a GOP size of 8. The QPs of the encoder are adjusted such that the total bitrate does not exceed 300 kilobits per second (kbs). For each coding structure, equal error protection (EEP) and unequal error protection (UEP) are applied at the link layer. For EEP, the left and right or video and depth bursts are protected with the same FEC rate. On the other hand, UEP requires the video bit streams to be partitioned into different segments with different priorities. Segments are then protected with unequal amount of FEC data. For partitioning the video bit streams, there are several methods such as data partitioning and spatial–temporal quality layering [89]. In the referred study, a partitioning based on the views only is performed, i.e., left/right views in different segments or left/depth data in different segments. More complex partitioning can also be
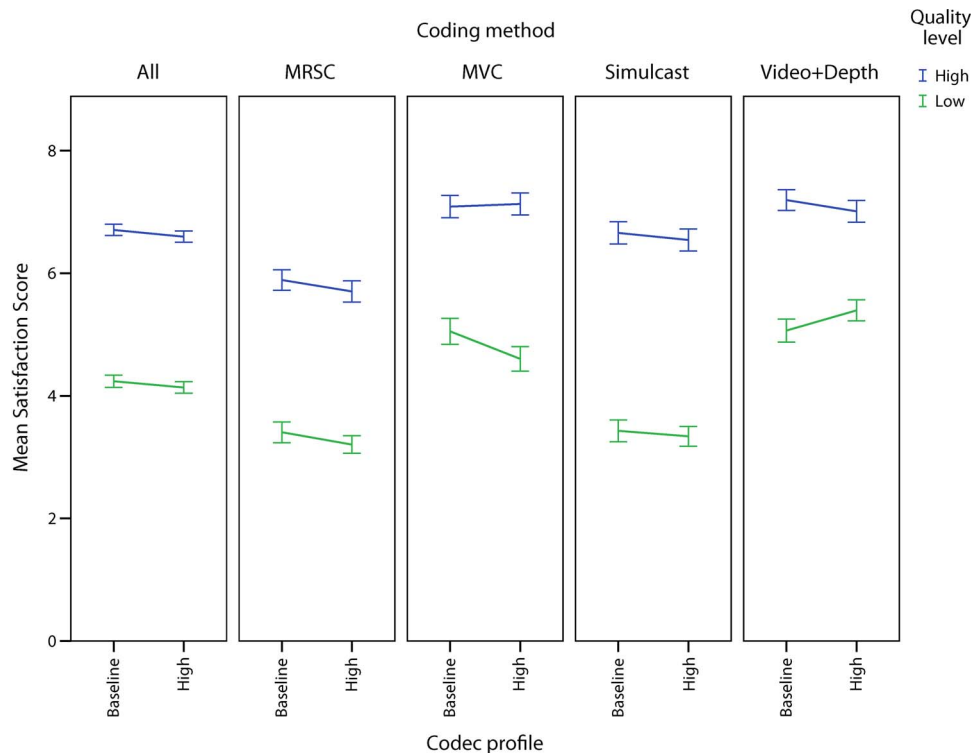
Fig. 29. *Comparison of the mean satisfaction scores for coding methods used in two studies [108], [111] for baseline and high profile. Error bars show 95% CI of mean.*

applied to the stereo data. Once segmented, several UEPs are derived where the channel coding ratio among the streams is determined according to the priority level of the streams.

In the transmission experiments conducted, a constant typical FEC rate (3/4) is chosen to protect the left and right bursts in the EEP mode since applying an MPE-FEC code rate below $R = 3/4$ at a medium frame size is not recommended without further measures [91]. Then several unequal protection schemes are derived using this EEP structure. Using the FEC rate chosen, 1/4, 2/4, 3/4, and 1% of the RS columns of right burst (right view or depth) are transferred to the left burst (left view), respectively, corresponding to the UEP1, UEP2, UEP3, and UEP4.

Fig. 30. *Block diagram of 3-D broadcasting system over DVB-H.*

Table 3 Spatial and Temporal Characteristics of Test Contents Used in Transmission Tests

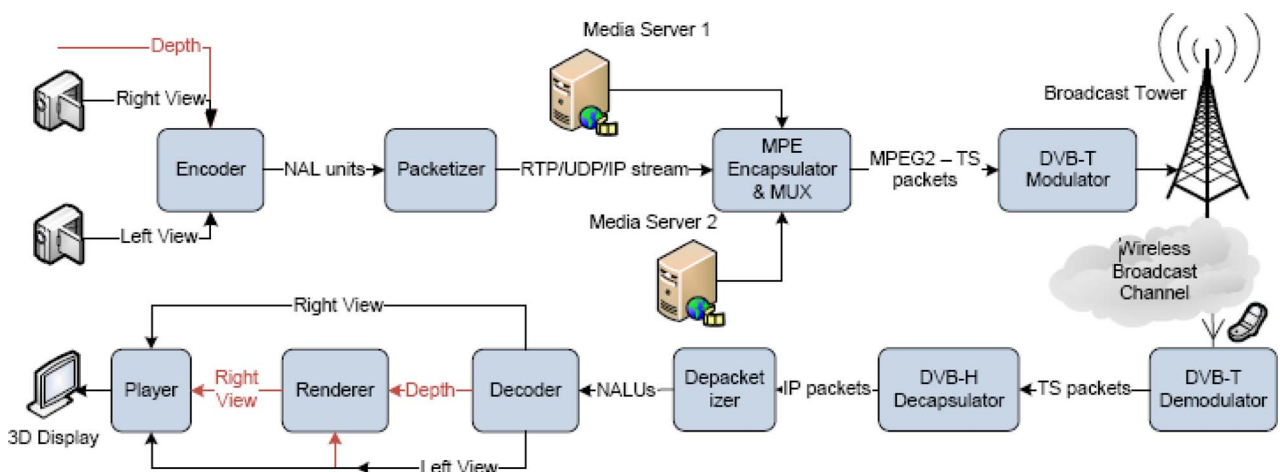| Content	Characteristics	Width	Height	Fps
HeidelbergAlleys	Low camera motion, Low Motion, High Detail	432	240	12.5
KnightsQuest	Computer – Generated,	432	240	12.5
RhineValleyMoving	High Camera and Object Motion, Low Detail	432	240	12.5
RollerBlade	Stationary camera, High Object Motion	320	240	15

For simulating the physical transmission channel, a MATLAB/Simulink tool that models the DVB-T/H modulation and demodulation processes and the physical transmission channel has been used [90]. The channel is modeled as multipath Rayleigh fading channel with additive white Gaussian noise. A mobile use case with Cost 207 radio channel model TU6, having maximum Doppler frequency of 24 Hz, is used to obtain the channel specific error patterns. These patterns are then used for modeling the TS packet loss due to channel conditions.

In all the simulations, PSNR values have been used as the distortion metric. First, mean squared errors (MSEs) are calculated individually for the left and right channels. They are used to calculate the PSNR for the left and right channels and the average of the two MSEs is used to calculate the average PSNR. At this point, we would like to mention that perceptually driven objective quality metric for stereo images would be more appropriate for comparison than PSNR. There has been an active research toward developing such metrics, however, they are still deficient in delivering simple, interpretable, and reliable results for the mobile case of interest [136].

In case of V + D sequences, since even for the lossless case there is an existing distortion (for PSNR metric) due to imperfections during depth estimation and rendering, original right view is not taken as the reference sequence. Instead, the distortions of the V + D transmissions are given as the PSNR of the received left sequence using original left view as reference; and the PSNR of the right sequence rendered from the received left and depth views using the right view rendered from original left and original depth.

Figs. 31 and 32 show the PSNR results for different coding and protection methods and for the *RollerBlade* and *KnightsQuest* videos. The results show that MVC performs better than simulcast because of the compression efficiency (bitrate of MVC coded video is chosen to be equal to that of simulcast coded video). UEP in general results in rather marginal improvement over EEP especially under low channel SNR. Also it has been shown that the results depend heavily on the content. If the depth map is accurate as seen in the *RollerBlade* video, V + D representation outperforms other methods. If the depth map is not accurate, MVC outperforms V + D representation for high
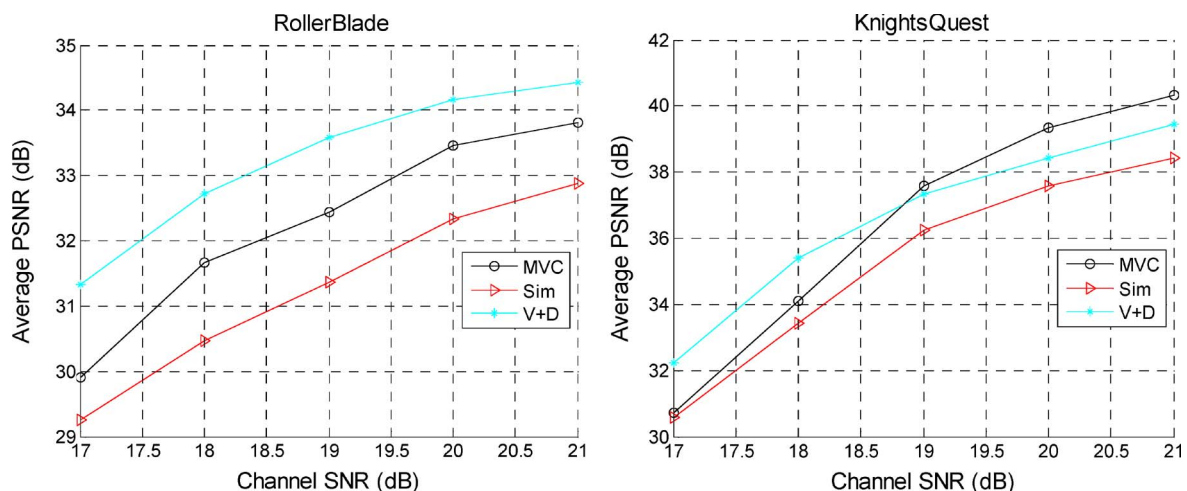
Fig. 31. *Average PSNR results for coding method comparison in EEP mode.*
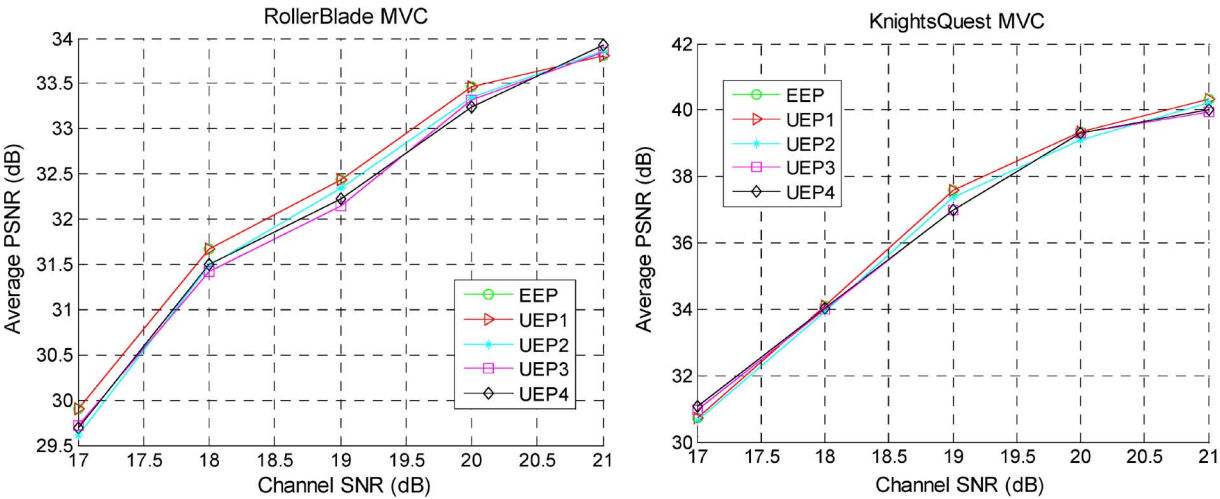
Fig. 32. *Average PSNR results for protection method comparison in MVC mode.*

SNR cases, however due to the compression efficiency of V + D representation it yields better results for low SNR. On the other hand, at the receiver side, view synthesis has to be performed after decoding to generate the second view of the stereo pair, which is rather challenging for mobile devices to achieve in real time.

Subjectively, transmission parameters for mobile 3-D media have been evaluated by Strohmeier *et al.* [109]

under the constraint of the studies on coding methods for mobile 3-D video by Strohmeier and Tech [108], [111]. This large-scale study has targeted channel transmission parameters taking into account different error resilience methods at the link layer (equal and unequal MPE-FEC) of the DVB-H channel. Regarding the transmission channel, equal (EEP) and unequal (UEP) error protections have been assessed at two different error rates of 10% and 20%

Fig. 33. *Results of transmission study [109] given as overall results (all) as well as content-per-content. Error bars show 95% CI of mean.*

Fig. 34. *Correlation plot of descriptive study on different coding methods for mobile 3-D television and video [108], [110]. Inner circle and outer circle mark 50% and 100% explained variance, respectively.*

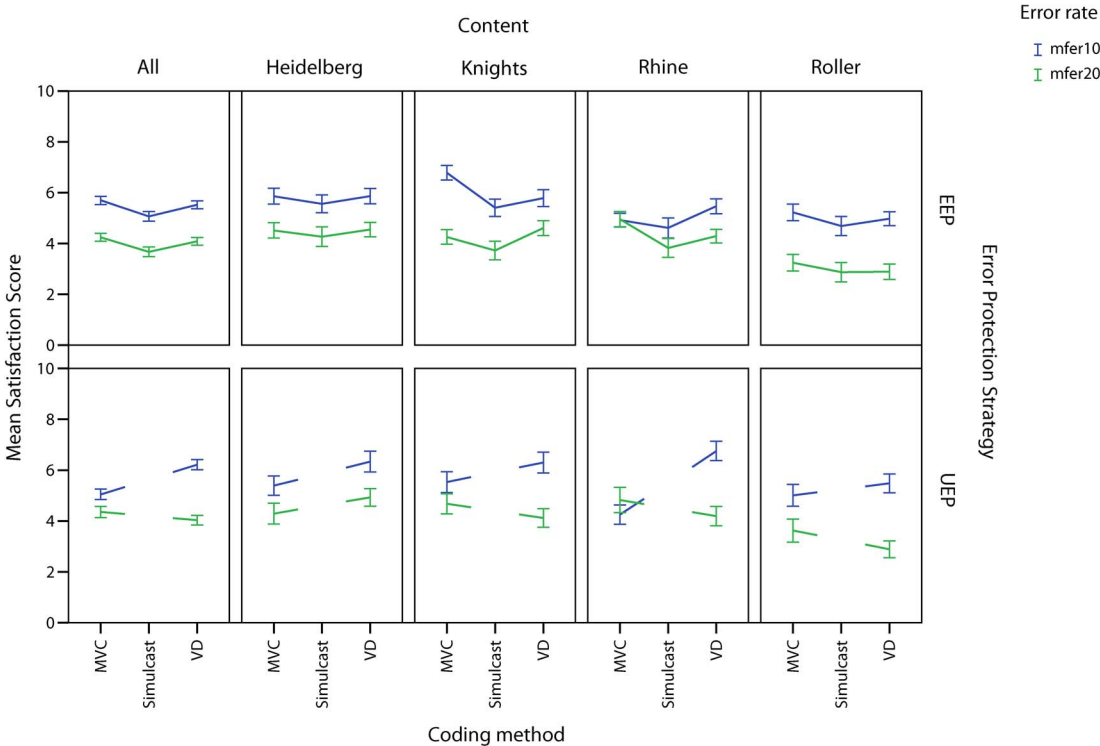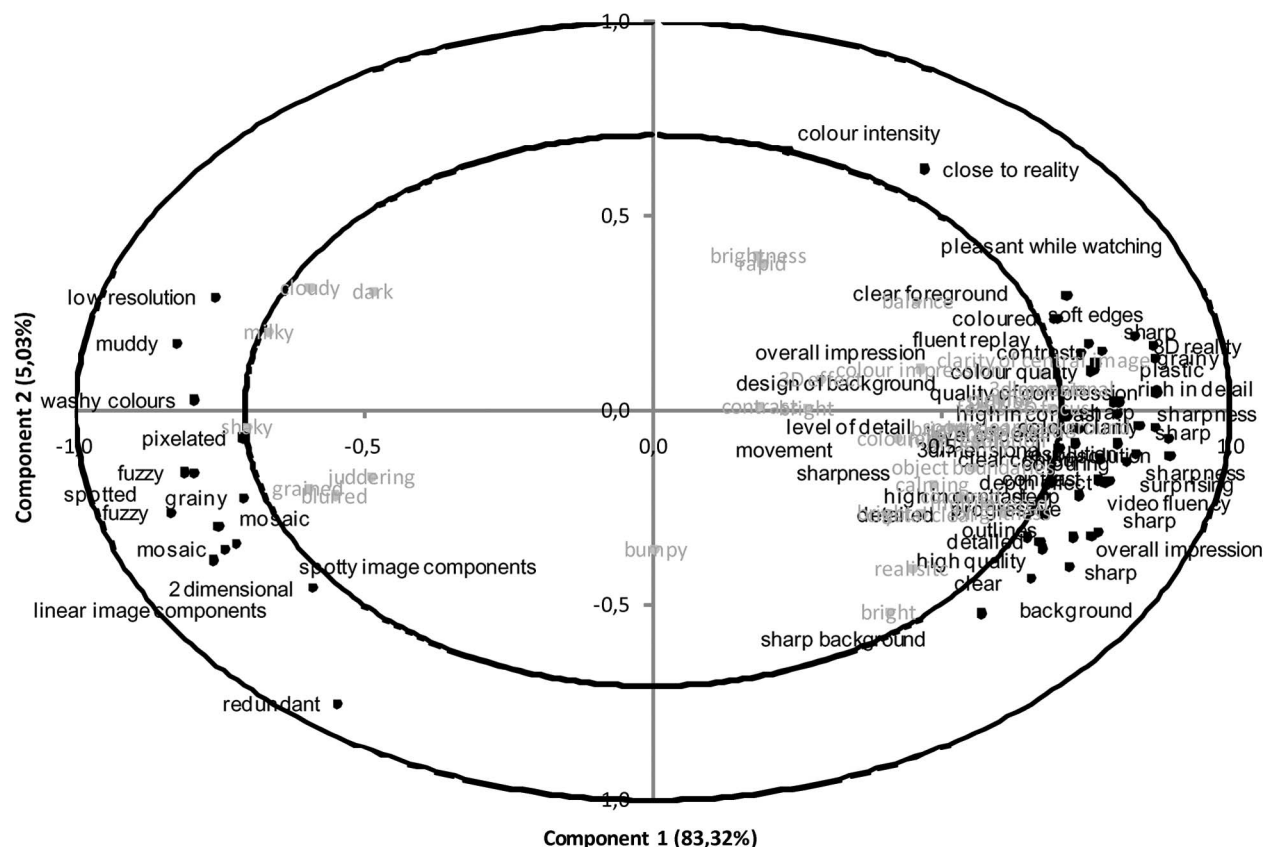corresponding to low and high channel SNRs. According to the results of the coding methods evaluation study, only MVC and the video + depth approach were evaluated for all parameters. Four different contents chosen to match the user requirements of mobile 3DTV [103], [104] have been used. Seventy seven test participants took part in the subjective quality evaluation. ACR was chosen as test method and test participants again evaluated acceptance of (yes/no) and satisfaction with (11-point-scale) perceived overall quality.

The results of the study (see Fig. 33) [109] confirm the findings of Strohmeier and Tech [108], [111]. At low error rates, the acceptance rate for all test items was at least 60%. The results of the acceptance rate are promising that the current state of the art in mobile 3-D media encoding and transmission can already reach a good quality acceptance at the end user. Genuinely 3-D coding methods, MVC and video + depth, have outperformed simulcast at all settings. While for low error rates, MVC and simulcast provided the same quality, MVC has been rated better at higher error rates. Regarding the transmission parameters for low error rates, the results show that MVC performs best at EEP, while video + depth is significantly better at UEP. The error protection did not show any impact on the perceived quality

at high error rates. An explanation for these results can be found in the fact that UEP allows for better protecting the video view in the video + depth approach. The better performance for MVC at EEP can be explained with the additional interview dependencies of left and right views. Taking together the results of the study, Strohmeier *et al.* conclude that MVC is the strongest coding method contender for mobile 3-D media due to its higher error robustness in time-varying wireless channels [109].

C. User-Centered Evaluation Studies on Mobile 3-D Media

Beyond the quantitative analysis of satisfaction with the overall quality of mobile 3-D media systems, the UC-QoE approach [95] targeted a deeper evaluation of the different components that contribute to QoE for mobile 3-D media. The application of the OPQ approach [110] resulted in deeper knowledge about the interaction of video quality and depth perception on forming 3-D QoE. In sensory profiling, test participants, in addition to a quantitative profiling, develop their individual quality attributes. These attributes are then used individually to describe the perceived quality. The data are then analyzed using generalized procrustes analysis [100], which results in a

low-dimensional and a perceptual model of the experienced quality factors. Two studies by Strohmeier *et al.* [107], [108] have shown that the video quality and artifact-free video perception is still the key factor for high subjective quality of mobile 3-D video. The results, as illustrated in Fig. 34, demonstrate that quality mainly depends on one component that has been identified as "*video quality*" as its polarities are described with attributes like mosaic, fuzzy, or grainy on the negative polarity and with sharp, high in contrast, and clear on the positive polarity. Surprisingly enough, a depth-related component has not been identified. Attributes describing depth like *3-D reality* or *3dimensional* are included on the positive polarity. These results are in line with previous studies [97]. The artifact free perception of 3-D video is also important in the comparison of mobile 2-D versus 3-D video. The added value of 3-D is only positively attributed to quality perception if the artifact levels of 2-D and 3-D stimuli are comparable [107]. According to the findings, subjective quality of current mobile 3-D devices is still limited by the quality of the applied displays.

The collection of individual quality factors from several descriptive studies on mobile 3-D media was used by Jumisko-Pyykkö *et al.* [125] to derive a general set of components of QoE for mobile 3-D video and television systems. The results generalize the individual attributes into four main components of QoE, namely, visual quality, viewing experience, content, and other modalities interactions. The component of visual quality is further divided into subcomponents of depth, spatial, and motion. Jumisko-Pyykkö *et al.* [125] provide definitions for each of the components and the underlying factors that contribute to each of these components. Jumisko-Pyykkö *et al.*'s study transformed the content-dependent findings of the descriptive quality studies [107], [108], [118] into a generalized set of components that can be used in further system developments to guide the design for user-centered quality optimization.

Another main focus in the UC-QoE evaluation has been set to the QoE evaluation in the context of use [94], [118]. It aimed at extending the external validity of the results gained in controlled environments. A recent work on the evaluation of mobile 3-D media in the context of use has compared perceived quality in laboratory and different mobile contexts [96], [98], [118]. The work combined quantitative and qualitative evaluation tasks as well as in-depth analysis of contexts and task effort [118]. The results confirm the findings of the user requirements in terms of heterogeneity of the different contexts [103]. Further, the studies have revealed that the results of the quality evaluation differ between controlled environments and the contextual settings. Test participants were less critical in the contextual environments [98]. The studies also showed that quality in the context depends on the contextual circumstances. Body movements to adjust the viewing distance as well as gaze shifts due to shared attention were significantly higher in the context in comparison to the

controlled environment. The strong conclusion is that mobile 3-D media systems, besides the 3-D experience, need to guarantee ease of viewing as well as a high viewing comfort to provide a high viewing experience in heterogeneous usage contexts [94], [103], [118].

D. Three-Dimensional Graphical User Interfaces

It is desirable that the user engage with 3-D content actively instead of just being a passive consumer. In addition, the users should also be able to search, browse, and annotate 3-D media content, using 3-D input modalities. Three-dimensional media will benefit from interactivity on mobile devices more than on desktops, because of the limitations of the mobile context, including small physical screen size and limited input modalities. With users demanding ever larger screens and attractive interfaces from mobile devices, the graphical user interface is becoming the most prominent feature of a mobile device.

Several works have studied the creation of 3-D interaction techniques that approach the richness of reality, particularly for desktop and large-scale interaction. Shneiderman *et al.* [47] have examined the features for increasing the usability of 3-D user interfaces (UIs) primarily for desktop and near-to-eye displays, and have proposed general guidelines for 3-D UI developers. These guidelines include: better use of occlusion, shadows, and perspective; minimizing the number of steps in navigation in the UI; and improving text readability with better rendering, limited angle to the view position, contrast with the background, and so on. Bowman *et al.* have analyzed the interaction techniques that are common to applications in 3-D user interfaces, and have developed a taxonomy of universal tasks for interacting with 3-D environments [48]. These tasks include: selection and manipulation of virtual objects, travel and wayfinding within a 3-D environment, issuing commands via 3-D menus, and symbolic input such as text, labels, and legends. Defining appropriate 3-D interaction techniques is still an active field in itself [48].

1) Three-Dimensional Widgets: For 3-D graphics, however, there is a lack of standardized 3-D UI widgets. This is partially due to the lack of commonly accepted list of UI metaphors, and partially due to the lack of an effort to structure a comprehensive and flexible set of existing widgets into a common 3-D UI library. Also, when designing 3-D user interfaces, new challenges emerge compared to traditional 2-D UI design. A major difference between 2-D and 3-D UIs is how the possibility to position objects in depth (along the *z*-axis) affects information density. Recent efforts have attempted to standardize a list of 3-D widgets [49]–[51]. The most popular 3-D widgets are based on metaphors that can be listed as tree, card, elevator, gallery, mirror, book, and hinged metaphors [51]. For example, Apple's Coverflow interface that is used in

Fig. 35. *Three alternatives for 3-D media browser layout [54].*

iPhone and Mac OSX Finder applications makes use of the card metaphor.

According to the application and the targeted task, different layout techniques can be selected. Undeniably, depth positioning adds complexity to the design of UIs since more layout options emerge. A stereoscopic 3-D UI looks quite different than a 3-D UI rendered on a 2-D screen. To designers without a lot of prior experience of the characteristics of stereoscopic design, guessing the visual effects of positioning UI elements in depth can be difficult. In Fig. 35, 3-D graphics is used to display a number of media content for a media browser in a circle seen in different layouts. In Fig. 36, another UI example by TAT AB, Malmö, Sweden, called SocialRiver, is shown, where photos, videos, and applications are dropping down in at the far end, and move towards the front. The user can "catch" a photo, video, or application and make it active. This includes showing the video or photo in higher resolution, or activating the application, as shown in the figure. Programmable vertex and pixel shaders are used to render depth-of-field effect and motion blur to direct the focus to the front-most icons, as well as to animate "wobbly" windows using vertex skinning.

2) Three-Dimensional UI Performance: In 3-D UIs, it is essential to optimize the graphics rendering for power consumption. In stereoscopic rendering, the images for the left and right eyes are very similar, and there is an opportunity to exploit this inherent coherency. With a brute-force implementation, the scene is first rendered to the left eye, and then to the right eye. In general, however, it makes sense to render a single triangle to both views before proceeding with the next triangle [52]. Kalaiah and Capin [53] use this rendering order to reduce the number of vertex shader computations. By splitting the vertex shader into parts that are view independent (and hence only computed once) and view dependent, vertex shader computations can be greatly reduced. In the per-pixel processing stage that follows, a simple sorting procedure in a generalized texture space greatly improves the texture cache hit ratio [52], keeping the texture bandwidth very

close to that of monoscopic rendering. In addition, Hasselgren and Akenine-Möller [52] introduce approximate rendering in the multiview pipeline, so that fragment colors in all neighboring views can be approximated from a central view when possible. Otherwise, the pipeline reverts to full pixel shader evaluation. When approximate rendering is acceptable, this technique can save a lot of per-pixel shader instructions executions.

To achieve good graphical performance and low power consumption, it is necessary to reduce the internal traffic between the processing elements and the memory. Therefore, mobile graphic solutions make use of data and texture compression to decrease that traffic. This is made even more important with the trend that computation power increases at a faster rate than memory bandwidth. For example, in a recent work, based on the International Technology Roadmap for Semiconductors (ITRS), Owens [12] reports that the processing capability growth is about 71%, while DRAM bandwidth only grows by 25%.

Fig. 36. *TAT's SocialRiver application.*

One of the most viable approaches for reducing memory traffic to graphics processing unit is compression of textures and buffers [113]. *Textures* can be considered as read-only images that are attached to graphical data. The main requirements of a texture compression/decompression algorithm include fast random access to the texture data, very fast decompression, and inexpensive hardware implementation. The requirement of random access usually implies that a block of pixels is compressed to a fixed size. These requirements have given rise to codecs, such as the Ericsson Texture Compression (ETC) and the PowerVR Texture Compression (PVRTC), which allow developers to compress textures down to 4 b/pixel or more without any perceived loss of quality. *Buffers* are different from textures in that they are symmetric: both processes must be performed on hardware in real time. For example, the color buffer can be compressed, and so when a triangle is being rendered to a block of pixels (say, 4 × 4) in the color buffer, the hardware attempts to compress this block.

Another approach for reducing memory traffic is based on tiling architectures. Tiling architectures are built on the goal to reduce the memory traffic related to frame buffer accesses, which may be one of the costly parts of an application. Instead of storing the full frame buffer in memory, thus transmitting it to the CPU repeatedly during rendering for different objects, only a small tile of the frame buffer is stored on the graphics chip, and the rendering is performed one tile at a time. This approach allows many possibilities for optimization and culling techniques, avoiding processing of data that will not contribute to the final image. Commercially, both Imagination Technologies Ltd., Hertfordshire, U.K. and ARM, Cambridge, U.K. provide a mobile 3-D accelerator using a tiling architecture.

3) Three-Dimensional User Input: Utilizing 3-D input techniques with autostereoscopic displays provides additional challenges related to the finger occluding the stereo information and virtual buttons being at different depth levels compared to the physical display, as well as problems related to the limited viewing area of the autostereoscopic display. A number of alternatives currently exist on mobile devices for 3-D interaction, including the use of touchscreen-based input, inertial trackers, camera-based tracking, GPS tracking—each with its own advantages and disadvantages.

- With *touchscreen-based input*, efficient use of screen space is essential. For single-touch or multi-touch screen UIs, the main limitation is that interactive elements should be presented in at least 1 × 1 cm^2 on the touch surface to be picked by an average finger [30]. In return, this limits how many UI elements can be rendered on display. A possible solution is to layer the 3-D UI elements, such that the elements in the top layer are large enough to support finger touch input, while rendering can be denser in the underlying layers.

- With *inertial tracker (accelerometer or gyroscope)-based input*, there is an advantage that there is no limit on the size of the UI elements. On the other hand, a major problem with inertial trackers is that they suffer from error accumulation due to sensor biases, noise, and drift [43]. In addition, because mobile devices are assumed to be used while on the move, mechanisms are necessary to filter out the jitter created by user's movement (e.g., acceleration due to walking, user in a car) from the user's actual input to the application. Thus, recent research studies have attempted to detect the context from accelerometer input [44].

- *Camera-based input* solutions have also been proposed. Face tracking allows enhancing the depth effect in the applications by supplying motion parallax for enhancing human–computer interaction [44]. In addition, eye-position tracking allows adapting the stereo views to compensate for the zones with high crosstalk and to prevent pseudo-stereo [127]. Camera input can also be used for tracking the self-movement of the device, which can be used for controlling scroll and view direction in an application [45]. Researchers have also proposed a finger tracking system, which allows the user to interact with 3-D scenes floating in 3-D [46].

V. USE SCENARIOS AND RESEARCH CHALLENGES FOR THE NEXT GENERATION 3-D MOBILE DEVICES

Three-dimensional media-enabled mobile devices are part of a bigger revolution bringing the next generation networked media devices, services, and applications where Internet is expected to play the central role. In the incoming years, Internet is expected to become larger, faster, and more reliable. Its use will grow beyond simple tasks such as searching for movies or buying food online. The web will evolve from a place for *sharing* content (Web 2.0) to a common environment where content is cocreated (Web 3.0) [61]. The media, occupying most of the today's internet—images and video, will evolve to the more realistic, 3-D images and 3-D video. Consumer electronics will transform from digital (CE 2.0) to connected (CE 3.0) [62] and will support 3-D media as well [63]. Today, most of the 2-D media exists in the flat world of Web 1.0 and Web 2.0 pages. Naturally, the 3-D media of tomorrow will "live" on a 3-D canvas—the 3-D media internet. Instead of a proprietary 3-D virtual world created by a single organization (such as Second Life [64] or Google Lively [65]), the 3-D internet of the future will be created by its users.

The vision for 3-D internet is not brand new. However, earlier attempts—like VRML, X3D, and numerous other standards—were not widely accepted by the public. One of the reasons is that creating a VRML model requires too

much time and skills compared to shooting a photo and sharing it on Picasa [66] or Flickr [67]. The future 3-D internet should allow people to cocreate contents and knowledge, and key factor for its success is that the users have tools to create 3-D media as simply as it is for making a photo or video today.

We foresee a universal, scalable, user-centric service that will allow 3-D media to be cocreated and positioned on the 3-D continuum of the Future Internet. Such service will combine 3-D audio–video data and 3-D models both anchored to position in 3-D space. In the *3-D media cloud*, the data will be continuously updated in a recursive manner, as illustrated in Fig. 37. Incoming 3-D video streams will be used to update the models; the models will be used to register positions in space, which will refine the coordinates of the video streams as well as their 3-D quality.

A key element of such service is what we call next-generation 3-D-enabled mobile device or Mobile Multi-media Device 3.0 (MMD3.0). It is a portable wireless network terminal, capable of capturing images and video in 3-D, recording 3-D audio, and being aware of its position and orientation in 3-D world. It can capture 3-D data, tag it with 3-D location, and send it to the cloud. It can browse through 3-D audio–video streams and 3-D models, and visualize them on a 3-D display. Many MMD3.0 devices will record data from the 3-D world, and sharing this data will create the canvas of the 3-D media internet.

The construction of the virtual world will gradually evolve through three stages. At the beginning, user-created content will be roughly positioned on a 2-D map. Such services already exist—one example is Google Maps, which relies on volunteers to create 3-D models and position them manually on the map space [68]. Another service, soon to appear, is Nokia Image Space [69] where 2-D photos are automatically geo-positioned based on

GPS and compass data. Our vision combines both concepts—media will be in 3-D and will be automatically geo-located on a 3-D map.

In the second stage, the collected 3-D audio–video data will be used to create *3-D models*—in the form of point clouds—of the real world. One example for such paradigm is Microsoft Photosynth [70], where 2-D photos are used for building rough point cloud of a scene. The downside of Photosynth is that it requires many 2-D images to reconstruct a 3-D model, and expects the images to be *manually* tagged as belonging to a certain place. On the contrary, the 3-D audio–visual data gathered by an MMD3.0 type of device will allow reconstruction using much fewer sources. As a result, more precise 3-D models will appear at a faster rate in the 3-D media internet.

In the third stage, most of the geographic locations in the world will be presented as 3-D models. Naturally, the important landmarks will be reconstructed first. As new audio–video streams are available, the 3-D models will be continuously updated. The 3-D media will appear on the map, and will be available for browsing by location or following hyperlinks. The users might volunteer to improve the quality of the virtual map, since adding new data will be an easy "point-and-click" operation, or, they might contribute by simply sharing their holiday 3-D videos. In the 3-D media internet, a MMD3.0 compatible device will serve both as a distributed sensor network and a terminal. By aiming the device towards a landmark in the real world, it will "know" 1) what is in front of the camera and 2) the direction of the camera. This will enable services such as 3-D location search, 3-D position and time-shifting, and 3-D content browsing and creation.

1) Challenges: The current architecture of the Internet is progressively reaching a saturation point in meeting increasing user's expectations [61]. Future Internet should be able to grow both in size and throughput to accommodate tomorrow's communication requirements. In order to identify the key research challenges of 3-D media for mobiles we will follow the information flow between users of 3-D media and services providing it. Fig. 38 illustrates the path of 3-D media as it is being captured by and reconstructed on an MMD3.0 device, transmitted over the network, stored in the "cloud," forwarded on request, enhanced, and played back on an MMD3.0 network terminal.

A main research challenge is to make MMD3.0 truly personal. This includes understanding the features and limitations of user-created 3-D audio–video content and addressing the QoE as perceived by subscribers of 3-D media services.

Another research challenge is to seamlessly integrate different sensors for capturing 3-D audiovisual information—stereo microphones, stereo camera, range sensor, GPS and acceleration sensor along with autostereoscopic 3-D display and 3-D audio output. Sensor fusion and 3-D data reconstruction will be performed in the MMD3.0
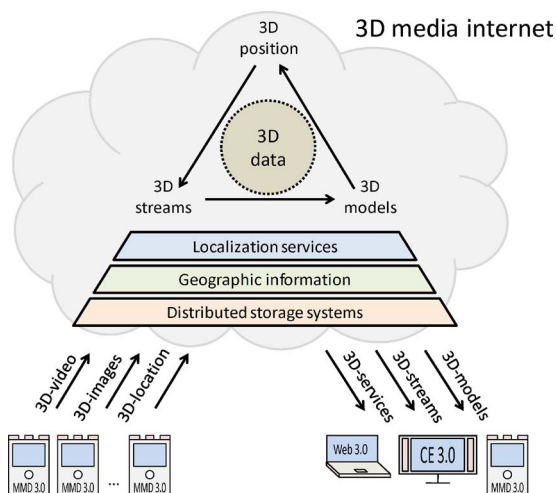

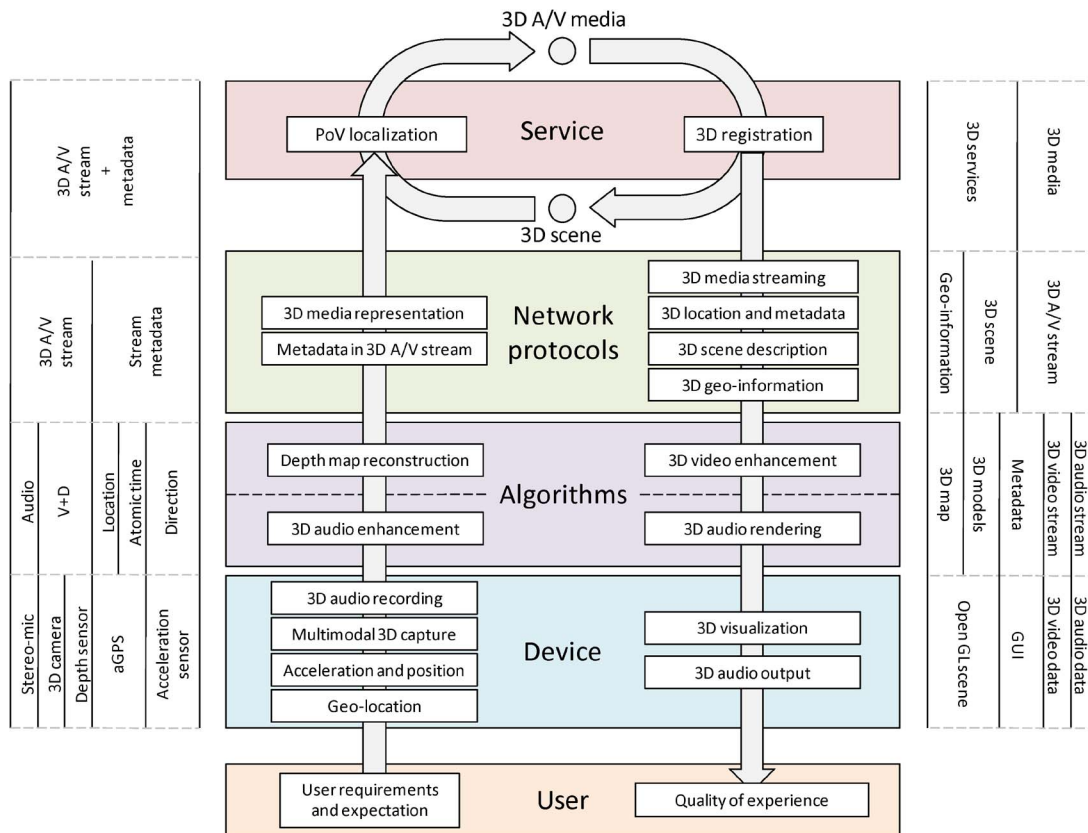**Fig. 37.** *Three-dimensional media internet concept.*

Fig. 38. *Three-dimensional media path from/to device.*

terminal. This will require powerful algorithms for converting multisensory data into *3-D media content*, i.e., video, augmented with dense depth and location information.

Next challenge is to enable network standards and protocols for representing 3-D media as interconnected network objects, or in other words, "*3-D things*," Format in which 3-D audio–video data, enriched with geographic coordinates, 3-D orientation and timestamps should be defined. On the way back to the MMD3.0 device, it should provide descriptions of 3-D scene augmented with 3-D objects, as well as 3-D media streams. Location, geographic information, and other services provided by the 3-D media internet should be requested by and delivered to an MMD3.0 in a scalable manner. It is also expected that geo-information will become core service of the Future Internet, as search is core service today.

The last challenge in our concept is to deliver 3-D media service that contains 3-D maps of the real world, libraries of 3-D models, and 3-D audio–video streams located on the maps. This is precisely the stage, when *user-created* content will become *cocreated*. The following functionalities should be supported by such service: 3-D models ("point clouds") reconstructed from available 3-D streams; position (point-of-view) of MMD3.0 device registered with respect to the models. The "point-of-view localization"

and "point cloud reconstruction" tasks will be re-executed as new 3-D media is available, yielding better 3-D models and better localization of the 3-D media available in the library. According to the "network tussle" principle [71] several 3-D media services in various stages of precision can coexist in the Future Internet being compatible and standardized. Further research challenges are related with distributed network storage and "cloud computing."

VI. CONCLUSION

In this paper, specifics of delivery of 3-D media to mobile devices have been addressed. Recent studies have confirmed that the users of such media expect higher realism of and closer emotional relation with the new content. Achieving such realism and emotional effect on a portable device is a challenge both for the optical quality of the display and the methods for creation and delivery of 3-D content. To address this challenge in a proper way, the studies of user experience have to be scheduled already at the beginning of the design of the overall system. Furthermore, new methodologies for user studies have to be developed to tackle the complexity of the problems with content, formats, delivery, and consumption. Two such methodologies, namely the OPQ and user-centered QoE

evaluation in the context of use, have been developed and successfully applied with the aim of optimizing new technology components and gathering new knowledge about how users tend to consume 3-D content on mobiles. The studies have especially emphasized the importance of visual quality of 3-D content for the acceptance of the new technology. The results of these studies have strong implications to the choice of displays, 3-D video formats, and coding and transmission methods as well as the receiver-side processing and 3-D UIs.

In the successive stages of development and deployment of 3-D services and applications for mobiles, new high quality 3-D displays will be available at first. Portable autostereoscopic displays have been the main contender for delivery of 3-D visual experience on mobiles. The user studies have elicited the principal characteristics of such displays. They need to be switchable in order to provide the freedom of choosing between 2-D and 3-D contents and their combination. They need to provide the same quality in 2-D and 3-D as 3-D with decreasing quality is immediately discarded by the user. For such displays the spatial resolution does matter and it should not be compromised for the price of delivering the 3-D effect. Portable 3-D displays should guarantee the ease of viewing and ensure high viewing comfort in heterogeneous usage contexts.

After displays, it is the content to be delivered. It is highly determined by the dynamism of mobile users. It should be content for "fast" consumption: sport events, short documentary, and news. No long watching is expected but 15–30 min of use. In addition to television-like content, mobile applications to be used in heterogeneous environments such as interactive navigation and 3-D games are highly expected.

Three-dimensional video seems the most mature content for mobile delivery. Again, the quality issue is of primary importance, as the user studies revealed that 3-D video is accepted as superior to 2-D video *only if* artefact-free. This determines the research challenges for the format and coding and transmission methods. Among coding methods, MVC has demonstrated the best rate-distortion performance and robustness in varying channel conditions. These results are consistent with the choice of MVC as the coding format for Blu-Ray discs. However, this consistency specifies also the next research challenge: how to effectively repurpose high-definition (HD) 3-D content

for its mobile visualization as it is expected that 3-D video will be mainly created in HD. Simple resizing of stereo video effects in changing the 3-D geometry of the scene and diminishing the 3-D effect. It seems that there is a need of a genuine master format for 3-D video where the depth map of the scene is explicitly presented so to allow a realistic rendering in different perspectives and spatial resolutions.

Precisely because of the demand for high quality, error protection for robust transmission of 3-D video over wireless channels is a must. Optimal combinations of effective coding and effective error protection have been studied especially for the case of DVB-H broadcast and the results have favored the combination of MVC with application-layer slice-mode error protection and MPE-FEC EEP. Still, UEP approaches bear the potential to achieve higher performance especially if combined with cross-layer optimization.

Along with the quality of 3-D content, it is the attractive graphical user interface that must appeal to the mobile users. In contract to content delivery where scalable solutions are likely (i.e., repurposing of HD content, rendering of mobile stereo from multiview plus multi-depth representations), the graphical user interfaces should be unique and scalable solutions are not possible. Instead, GUIs have to be especially designed for the portable 3-D platforms addressing the issue of realism, emotion exploiting the main difference between 2-D and 3-D—the availability of depth to be used for increasing and enriching the information density.

The first stage of deployment of 3-D media to mobiles will be mainly related with media consumption, i.e., delivery of video, GUIs, games. The next stage is to turn the mobile user from a consumer to a creator of 3-D content. This would require substantial research efforts, as to make capture in 3-D a trivial task. Current state of the art dictates that 3-D capture is highly professional work related with the 3-D-specific visual artifacts, which requires a professional planning and shooting combined with postprocessing. For mobile 3-D capture, these things should be made automatic. In the beginning, mobile 3-D camera devices will be with limited quality yet being able to contribute to cocreation of high-quality 3-D models and 3-D environments, where 3-D audio and video, augmented with positioning information, will be a basis of novel services and applications. ∎

REFERENCES

[1] A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Müller, C. Erdem, and C. Weigel, "Scene representation technologies for 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1587–1605, Nov. 2007.

[2] A. Smolic, K. Müller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*,

vol. 17, no. 11, pp. 1606–1621, Nov. 2007.

[3] G. B. Akar, A. M. Tekalp, C. Fehn, and M. R. Civanlar, "Transport methods in 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1622–1630, Nov. 2007.

[4] Y.-K. Chen, C. Chakrabarti, S. Bhattacharyya, and B. Bougard, "Signal processing on platforms with multiple cores—Part 1: Overview and methodologies," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 24–25, Nov. 2009.

[5] G. Blake, R. Dreslinski, and T. Mudge, "A survey of multicore processors," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 26–37, Nov. 2009.

[6] *OMAP 4 Platform: OMAP 4430/OMAP4440.* [Online]. Available: http://focus.ti.com/general/docs/wtbu/wtbuproductcontent.tsp?templateId=6123&navigationId=12843&contentId=53243

[7] *LH7A400 32-Bit System-on-Chip by NXP Semiconductors.* [Online]. Available: http://www.nxp.com/#/pip/pip=[pip=LH7A400_N_1]|pp=[t=pip,i=LH7A400_N_1]|

[8] *Marvel PXA320 Processor Series.* [Online]. Available: http://www.marvell.com/products/cellular/application/pxa320.jsp

[9] *NVidia Tegra Product Page.* [Online]. Available: http://www.nvidia.com/page/handheld.html

[10] *Next Generation NVIDIA Tegra Page.* [Online]. Available: http://www.nvidia.com/object/tegra_250.html

[11] *The Snapdragon Platform.* [Online]. Available: http://www.qualcomm.com/products_services/chipsets/snapdragon.html

[12] J. D. Owens, "Streaming architectures and technology trends," in *GPU Gems 2.* Reading, MA: Addison-Wesley, 2005, pp. 457–470.

[13] S. Pastoor, "3D displays," in *3D Video Communication*, Schreer, Kauff, and Sikora, Eds. New York: Wiley, 2005, pp. 235–260.

[14] P. Surman, K. Hopf, I. Sexton, W. K. Lee, and R. Bates, "Solving the 3D problem—The history and development of viable domestic 3-dimensional video displays," in *Three-Dimensional Television: Capture, Transmission, and Display*, H. M. Ozaktas and L. Onural, Eds. New York: Springer-Verlag, 2007.

[15] C. Van Berkel and J. Clarke, "Characterization and optimization of 3D-LCD module design," in *Proc. Stereoscopic Displays Virtual Reality Syst. IV*, 1997, vol. 2653, pp. 179–186.

[16] W. L. IJzerman, S. T. de Zwart, and T. Dekker, "Design of 2D/3D switchable displays," *J. Soc. Inf. Display*, vol. 36, pp. 98–101, May 2005.

[17] Sharp Laboratories of Europe. [Online]. Available: http://www.sle.sharp.co.uk/research/optical_imaging/3d_research.php

[18] S. Uehara, T. Hiroya, H. Kusanagi, K. Shigemura, and H. Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," *Proc. SPIE—IS&T Electron. Imaging Stereoscopic Displays Appl. XIX*, vol. 6803, pp. 68030O–68030O-8, 2008.

[19] G. J. Woodgate and J. Harrold, "Autostereoscopic display technology for mobile 3DTV applications," in *Proc. SPIE—IS&T Electron. Imaging Stereoscopic Displays Appl. XVIII*, 2007, vol. 6490, p. 64900K.

[20] *3D-LCD Product Brochure*, MasterImage. [Online]. Available: http://masterimage.co.kr/new_eng/data/masterimage.zip?pos=60

[21] H. Hong and M. Lim, "Determination of luminance distribution of autostereoscopic 3D displays through calculation of angular profile," *J. Soc. Inf. Display*, vol. 18, p. 327, 2010, DOI: 10.1889/JSID18.5.327.

[22] M. Salmimaa and T. Jarvenpaa, "3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays," *J. Soc. Inf. Display*, vol. 16, pp. 1033–1040, 2008, DOI: 10.1889/JSID16.10.1033.

[23] G. Woodgate and J. Harrold, "Key design issues for autostereoscopic 2-D/3D displays," *J. Soc. Inf. Display*, vol. 14, pp. 421–426, 2006, DOI: 10.1889/1.2206104.

[24] P. Boher, T. Leroux, V. Collomb Patton, T. Bignon, and D. Glinel, "A common approach to characterizing autostereoscopic and polarization-based 3-D displays," *J. Soc.*

[25] F. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, no. 2–3, pp. 99–108, Aug. 2004, DOI: 10.1016/j.displa.2004.07.004, ISSN: 0141-9382.

[26] *FinePix REAL 3D Series,* product brochure, Fujifilm, 2010. [Online]. Available: http://fujifilm.co.uk/media/dContent/mediaCentre/Brochures/0_FinePix-Real-3D-catalogue.pdf

[27] J. Konrad and P. Angiel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 128–140, Jan. 2006.

[28] B. A. Wandell, *Foundations of Vision.* Sunderland, MA: Sinauer Associates, 1995.

[29] D. Chandler, *Visual Perception (Introductory Notes for Media Theory Students)*, University of Wales, Aberystwyth, Wales, MSC portal site. [Online]. Available: http://www.aber.ac.uk/media/sections/image.html

[30] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis.* New York: Oxford Univ. Press, 1995.

[31] S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich-Hertz-Institut Berlin," in *Proc. 2nd Int. Display Workshop*, 1995, pp. 69–72.

[32] D. B. Diner, "A new definition of orthostereopsis for 3-D television," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 1991, pp. 1053–1058.

[33] B. Julesz, *Foundations of Cyclopean Perception.* Chicago, IL: Univ. Chicago Press, 1971.

[34] M. Halle, "Autostereoscopic displays and computer graphics," in *ACM SIGGRAPH*, vol. 31, no. 2, pp. 58–62, May 1997.

[35] M. Yuen, "Coding artefacts and visual distortions," in *Digital Video Image Quality and Perceptual Coding*, H. Wu and K. Rao, Eds. Boca Raton, FL: CRC Press, 2005.

[36] M. Yuen and H. R. Wu, "A survey of MC/DPCM/DCT video coding distortions," *Signal Process.*, vol. 70, pp. 247–278, Nov. 1998.

[37] M. McCauley and T. Sharkey, "Cybersickness: Perception of self-motion in virtual environments," *Presence: Teleoper. Virtual Environ.*, vol. 1, pp. 311–318, 1992.

[38] M. Wexler and J. Boxtel, "Depth perception by the active observer," *Trends Cogn. Sci.*, vol. 9, pp. 431–438, Sep. 2005.

[39] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content," in *Proc. SPIE, Stereoscopic Disp. Appl.*, vol. 7237, 2009, pp. 72371F-1–72371F-12.

[40] W. IJsselsteijn, P. Seuntiens, and L. Meesters, "Human factors of 3D displays," in *3D Video Communication*, Schreer, Kauff, and Sikora, Eds. New York: Wiley, 2005.

[41] M. Hassenzahl and N. Tractinsky, "User experience—A research agenda," *Behaviour Inf. Technol.*, vol. 25, no. 2, pp. 91–97, Mar./Apr. 2006.

[42] V. Roto, "Web browsing on mobile phones—Characteristics of user experience," Ph.D. dissertation, Dept. Comput. Sci. Eng., Helsinki Univ.

Inf. Display, vol. 18, pp. 293–300, 2010, DOI: 10.1889/JSID18.4.293.

Technology, Helsinki, Finland, 2006, TKK Dissertations 49.

[43] E. Foxlin, "Motion tracking requirements and technologies," in *Handbook of Virtual Environments: Design, Implementation, and Applications*, K. M. Stanney, Ed. New York: Erlbaum, p. 167.

[44] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *Proc. PERVASIVE 2004*, vol. 3001, Berlin, Germany: Springer-Verlag, 2004, pp. 1–17.

[45] T. Capin, A. Haro, V. Setlur, and S. Wilkinson, "Camera-based virtual environment interaction on mobile devices," *Lecture Notes in Computer Science*, vol. 4263. Berlin, Germany: Springer-Verlag, 2006, pp. 765–773.

[46] J. Hannuksela, S. Huttunen, P. Sangi, and J. Heikkilä, "Motion-based finger tracking for user interaction with mobile phones," in *Proc. 4th Eur. Conf. Vis. Media Production*, London, U.K., 2007, pp. 1–6.

[47] B. Shneiderman and C. Plaisant, *Designing the User Interface,* 4th ed. Reading, MA: Addison-Wesley, 2004.

[48] D. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, *3D User Interfaces—Theory and Practice.* Reading, MA: Addison-Wesley, 2004.

[49] R. Dachselt and A. Hubner, "Virtual environments: Three-dimensional menus: A survey and taxonomy," *Comput. Graph.*, vol. 31, pp. 53–65, 2007.

[50] *3D Components Web Site.* [Online]. Available: www.3dcomponents.org

[51] "Interim progress report for 3D UI/Direct manipulation solutions," EC FP7 3DPHONE Project, Tech. Rep. D6.1.2, 2009.

[52] J. Hasselgren and T. Akenine-Möller, "An efficient multi-view rasterization architecture," in *Proc. Eurograph. Symp. Rendering*, 2006, pp. 61–72.

[53] A. Kalaiah and T. Capin, "Unified rendering pipeline for autostereoscopic displays," in *Proc. 3DTV Conf.*, Kos, Greece, May 2007, DOI: 10.1109/3DTV.2007.4379427.

[54] TAT Inc., "Application built on 3D rendering & 3D display, joystick interaction," 3DPHONE Project Tech. Rep. D6.3.3, 2009.

[55] J. Poikonen and J. Paavola, "Error models for the transport stream packet channel in the DVB-H link layer," in *Proc. Int. Conf. Commun.*, Istanbul, Turkey, 2006, pp. 1861–1866.

[56] COST207, "Digital Land Mobile Radio Communications (Final Report)," Commission of the European Communities, Directorate General Telecommunications, Information Industries and Innovation, 1989, pp. 135–147.

[57] G. Akar, M. Oguz Bici, A. Aksay, A. Tikanmäki, and A. Gotchev, "Mobile stereo video broadcast," Mobile3DTV Project Rep. D3.2. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D3.2_Mobile3DTV_v1.0.pdf

[58] T. Marc, M. Lambooij, W. IJsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," *Proc. SPIE—IS&T Electron. Imaging*, vol. 6490, *Stereoscopic Displays and Virtual Reality Systems XIV*, Mar. 5, 2007, DOI: 10.1117/12.705527.

[59] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE*, vol. 1915, *Stereoscopic Displays and Applications*, pp. 36–48, Sep. 23, 1993.

[60] M. J. Meesters, W. A. IJsselsteijn, and P. J. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[61] *Future Internet: The Cross-ETP Vision Document,* European Future Internet Portal. [Online]. Available: http://www.future-internet.eu/fileadmin/documents/reports/Cross-ETPs_FI_Vision_Document_v1_0.pdf

[62] *The Digital Home Experience—Consumer Electronics 3.0*, Intel. [Online]. Available: http://www.intelconsumerelectronics.com/Consumer-Electronics-3.0/

[63] L. Onural, "Television in 3-D: What are the prospects?" *Proc. IEEE*, vol. 95, no. 6, pp. 1143–1145, Jun. 2007.

[64] *Second Life*. [Online]. Available: http://secondlife.com

[65] *Lively by Google*. [Online]. Available: http://www.lively.com

[66] *Picasa by Google*. [Online]. Available: http://picasa.google.com

[67] *Flickr by Yahoo*. [Online]. Available: http://flickr.com

[68] *Google SketchUp*. [Online]. Available: http://sketchup.google.com

[69] *Nokia Image Space*. [Online]. Available: http://research.nokia.com/imagespace

[70] *Photosynth by Microsoft Live Labs*. [Online]. Available: http://livelabs.com/photosynth/

[71] D. Clark, "Tussle in cyberspace: Defining tomorrow's internet," *IEEE/ACM Trans. Netw.*, vol. 13, no. 3, pp. 462–475, Jun. 2005.

[72] T. Wiegand, G. J. Sulivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[73] A. Aksay and G. Bozdagi Akar, "Evaluation of stereo video coding schemes for mobile devices," in *Proc. 3DTV Conf.*, Potsdam, Germany, May 2009, DOI: 10.1109/3DTV.2009.5069664.

[74] P. Merkle, H. Brust, K. Dix, Y. Wang, and A. Smolic, "Adaptation and optimization of coding algorithms for mobile 3DTV," Mobile3DTV Tech. Rep. D2.2, Nov. 2008.

[75] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand, "An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution," in *Proc. 27th Picture Coding Symp.*, Chicago, IL, May 6–8, 2009, pp. 389–392.

[76] A. Vetro, W. Matusik, H. Pfister, and J. Xin, "Coding approaches for end-to-end 3D TV systems," in *Proc. Picture Coding Symp.*, 2004.

[77] *Video Group: Text of ISO/IEC 14496-10:2009/FDAM1: Constrained Baseline Profile, Stereo High Profile and Frame Packing Arrangement SEI Message*, ISO/IEC JTC1/SC29/WG11 Doc. N10707, 2009, London, U.K.

[78] H. Lee, K. Yun, N. Hur, J. Kim, B. C. Min, and J. K. Kim, "A structure for 2D/3D mixed service based on terrestrial DMB system," in *Proc. 3DTV Conference 2007*, Kos Island, Greece, 2007.

[79] B. Furht and S. Ahson, *Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T, and MediaFLO*. New York: Auerbach Publications, 2008.

[80] S. Cho, N. Hur, J. Kim, K. Yun, and S. Lee, "Carriage of 3D audio-visual services by T-DMB," in *Proc Int. Conf. Multimedia Expo*, 2006, pp. 2165–2168.

[81] *3dphone Project*. [Online]. Available: http://www.3dphone.org/

[82] *Mobile3dtv: Mobile 3dtv Content Delivery Over DVB-H System*. [Online]. Available: www.mobile3dtv.eu

[83] [Online]. Available: cordis.europa.eu

[84] S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, *RFC 3984: RTP Payload Format for H.264 Video*. [Online]. Available: http://tools.ietf.org/html/rfc3984

[85] M. O. Bici, A. Aksay, A. Tikanmaki, A. Gotchev, and G. Bozdagi Akar, "Stereo video broadcasting simulation for DVB-H," in *Proc. NEM-Summit*, 2008, paper A2–11, 4 p.

[86] ETSI, "Digital Video Broadcasting (DVB): DVB-Himplementation Guidelines," TR 102 377 V1.3.1, 2009.

[87] [Online]. Available: www.dvb.org/groups_modules/commercial_module/cm3dtv/

[88] D. Bugdayci, M. Oguz Bici, A. Aksay, M. Demirtas, G. B. Akar, A. Tikanmäki, and A. Gotchev, "Stereo DVB-H broadcasting system with error resilient tools," Mobile3DTV Tech. Rep., D3.4. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D3.4_Mobile3DTV_v1.0.pdf

[89] S. Worrall, A. H. Sadka, P. Sweeney, and A. M. Kondoz, "Prioritisation of data partitioned MPEG-4 video over mobile networks," *ETT-Eur. Trans. Telecommun.*, vol. 12, no. 3, pp. 169–174, May/Jun. 2001.

[90] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[91] M. Kornfeld, "Optimizing the DVB-H time interleaving scheme on the link layer for high quality mobile broadcasting reception," in *Proc. IEEE Int. Symp. Consumer Electron.*, Dallas, TX, Jun. 2007, DOI: 10.1109/ISCE.2007.4382140.

[92] R. Calonati, "Techniche a diversita compatiili con il livello fisico dello standard DVB-T/H," Laurea in Ingegneria thesis, Dipartimento di Elettronica e Telecomunicazioni, Universita degli studi di Firenze, Firenze, Italy, 2007.

[93] S. Jumisko-Pyykkö and D. Strohmeier, "Report on research methodologies for the experiments," MOBILE3DTV Tech. Rep. D4.2. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.2_Mobile3dtv_v2.0.pdf

[94] S. Jumisko-Pyykkö and T. Vainio, "Framing the context of use for mobile HCI," *Int. J. Mobile-Human-Computer-Interaction*, to be published.

[95] A. Gotchev, A. Smolic, S. Jumisko-Pyykkö, D. Strohmeier, G. B. Akar, P. Merkle, and N. Daskalov, "Mobile 3D television: Development of core technological elements and user-centered evaluation methods toward an optimized system," in *Proc. IST/SPIE Conf. Electron. Imaging*, 2009, vol. 7256, DOI:10.1117/12.816728.

[96] S. Jumisko-Pyykkö and T. Utriainen, "User-centered quality of experience of mobile 3DTV: How to evaluate quality in the context of use," *Proc. SPIE*, vol. 7542, p. 75420W, DOI: 10.1117/12.849572.

[97] W. Tam, L. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," *Proc. SPIE*, vol. 3295, pp. 226–235, 1998.

[98] S. Jumisko-Pyykkö and T. Utriainen, "User centered quality of experience: Is mobile 3D video good enough in the actual context of use?" in *Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron. (VPQM)*, Scottsdale, AZ, 2010.

[99] S. Jumisko-Pyykkö, V. Kumar Malamal Vadakital, and M. Hannuksela, "Acceptance threshold: Bidimensional research method for user-oriented quality evaluation studies," *Int. J. Digit. Multimedia Broadcasting*, vol. 2008, Article ID 712380, 2008.

[100] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, pp. 33–51, 1975.

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

[102] *Subjective Video Quality Assessment Methods for Multimedia Applications,* Recommendation ITU-T P.910, 1999.

[103] S. Jumisko-Pyykkö, M. Weitzel, and D. Strohmeier, "Designing for user experience—What to expect from mobile 3D television and video," in *Proc. 1st Int. Conf. Des. Interactive User Experiences TV Video*, Mountain View, CA, Oct. 2008, pp. 183–192.

[104] D. Strohmeier, S. Jumisko-Pyykkö, M. Weitzel, and S. Schneider, "Report on user needs and expectations for mobile stereo-video," Mobile3DTV Rep. D4.1. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.1_Mobile3DTV_v1.0.pdf

[105] *Human-Centred Design Processes for Interactive Systems*, ISO 13407, The International Organization for Standardization, 1999.

[106] J. Freeman and S. E. Avons, "Focus group exploration of presence through advanced broadcast services," *Proc. SPIE—Human Vis. Electron. Imaging*, vol. 3959, pp. 3959–76, 2000.

[107] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "New, lively, and exciting or just artificial, straining, and distracting? A sensory profiling approach to understand mobile 3D audiovisual quality," in *Proc. 5th Int. Workshop Video Process. Quality Metrics Consumer Electron.*, Scottsdale, AZ, Jan. 2010, article no. 36.

[108] D. Strohmeier and G. Tech, "Sharp, bright, three-dimensional: Open profiling of quality for mobile 3DTV coding methods," in *Proc. Multimedia Mobile Devices*, San Jose, CA, Jan. 2010, article no. 75420T.

[109] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, G. Tech, D. Bugdayci, and M. O. Bici, "Results of quality attributes of coding, transmission and their combinations," Tech. Rep. Mobile 3DTV, Jan. 2010.

[110] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "Open profiling of quality: A mixed method approach to understand multimodal quality perception," *Adv. Multimedia*, 2010, DOI: 10.1155/2010/658980, Article ID 658980.

[111] D. Strohmeier and G. Tech, "On comparing different codec profiles of coding methods for mobile 3D television and video," in *Proc. 2nd Conf. 3D Syst. Appl.*, Tokyo, Japan, May 2010, article no. 28.

[112] H. Stone and J. L. Sidel, *Sensory Evaluation Practices,* 3rd ed. San Diego, CA: Academic, 2004.

[113] T. Capin, K. Pulli, and T. Akenine-Möller, "The state of the art in mobile graphics research," *IEEE Comput. Graph. Appl.,* vol. 28, pp. 74–84, Jul. 2008.

[114] H. Brust, A. Smolic, K. Müller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in *Proc. 3DTV Conf.,* Potsdam, May 2009, DOI: 10.1109/3DTV.2009.5069629.

[115] H. Brust, G. Tech, K. Müller, and T. Wiegand, "Mixed resolution coding with interview prediction for mobile3DTV," in *Proc. 3DTV Conf.,* Tampere, Finland, Jun. 2010, DOI: 10.1109/3DTV.2010.5506723.

[116] *Vocabulary for Performance and Quality of Service. New Appendix I Definition of Quality of Experience (QoE),* International Telecommunication Union, Geneva, Switzerland, ITU-T Recommendation P.10 Amendment 1, 2008.

[117] B. Escofier and J. Pagès, "Multiple factor analysis (AFMULT package)," *Comput. Stat. Data Anal.,* vol. 18, no. 1, pp. 121–140, Aug. 1994, DOI: 10.1016/0167-9473(94) 90135-X0167-9473.

[118] S. Jumisko-Pyykkö and T. Utriainen, "A hybrid method for quality evaluation in the context of use for mobile (3D) television," *Multimedia Tools and Applications,* 2010 in press.

[119] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proc. 17th ACM Int. Conf. Multimedia,* pp. 481–490, 2009. [Online]. Available: http://doi.acm.org/10.1145/1631272.1631338

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

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

[122] *ISO/IEC CD 23002-3: Representation of Auxiliary Video and Supplemental Information,* Doc. N8259, Klagenfurt, Austria, ISO/IEC JTC1/SC29/WG11, Jul. 2007.

[123] S. Uehara, T. Hiroya, H. Kusanagi, K. Shigemura, and H. Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," *Proc. SPIE—IS&T Electron. Imaging Stereoscopic Displays Appl. XIX,* vol. 6803, Jan. 2008, DOI: 10.1117/12.766936.

[124] D. Bugdayci, M. O. Bici, A. Aksay, M. Demirtas, G. B. Akar, A. Tikanmäki, and A. Gotchev, "Stereo DVB-H broadcasting system with error resilient tools," Tech. Rep. D3.4, Dec. 2009.

[125] S. Jumisko-Pyykkö, D. Strohmeier, T. Utriainen, and K. Kunze, "Descriptive quality of experience for mobile 3D video," presented at the 6th Nordic Conf. Human-Computer Interaction, Reykjavik, Iceland, Oct. 2010.

[126] B. J. Abbott, *An Integrated Approach to Software Development.* New York: Wiley, 1986.

[127] A. Boev, M. Georgiev, A. Gotchev, N. Daskalov, and K. Egiazarian, "Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display," in *Proc. 17th Eur. Signal Conf.,* Glasgow, Scotland, Aug. 2009, pp. 490–494.

[128] Mobile Broadcast/Multicast Service (MBMS), TeliaSonera White Paper. [Online]. Available: http://www.medialab.sonera.fi/workspace/MBMSWhitePaper.pdf

[129] A. Bourge, J. Gobert, and F. Bruls, "MPEG-C Part 3: Enabling the introduction of video plus depth contents," in *Proc. Workshop Content Generation Coding 3D-Television,* 2006. [Online]. Available: http://vca.ele.tue.nl/events/3Dworkshop2006/proceedings.html

[130] L. Pasquier and J. Gobert, "Multi-view renderer for auto-stereoscopic mobile devices," in *Proc. IEEE Int. Conf. Consumer Electron.,* 2009, DOI: 10.1109/ICCE.2009.5012210.

[131] *Vision on 3D Video,* Doc. N10357, Lausanne, Switzerland, ISO/IEC JTC1/SC29/WG11, Feb. 2009.

[132] *ST Ericsson: U8500—The Best Smartphone Platform.* [Online]. Available: http://www.stericsson.com/platforms/U8500.jsp

[133] J. Konrad, B. Lacotte, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Process.,* vol. 9, no. 5, pp. 897–908, May 2000.

[134] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 14, *Special Issue on Immersive Telecommunications,* no. 3, pp. 321–334, Mar. 2004.

[135] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Process., Image Commun.,* vol. 22 no. 2, pp. 217–234, Feb. 2007.

[136] A. Boev, M. Poikela, A. Gotchev, and A. Aksay, "Modeling of the stereoscopic HVS," MOBILE3DTV Tech. Rep. D5.3, 2010.

ABOUT THE AUTHORS

Atanas Gotchev (Member, IEEE) received the M.Sc. degrees in radio and TV engineering, and applied mathematics, and the Ph.D. degree in telecommunications from the Technical University of Sofia, Sofia, Bulgaria, in 1990, 1992, and 1996 respectively; and the Dr. Tech. degree in information technologies from Tampere University of Technology, Tampere, Finland, in 2003.

Currently, he is an Academy Research Fellow with the Department of Signal Processing, Tampere University of Technology. Before that he has held research and teaching positions with Bulgarian Academy of Sciences and Technical University of Sofia. His research interests are in sampling and interpolation theory, and spline and spectral methods with applications to multidimensional signal analysis.

Gozde Bozdagi Akar (Senior Member, IEEE) received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1988 and the M.S. and Ph.D. degrees from Bilkent University, Bilkent, Turkey, in 1990 and 1994, respectively, all in electrical and electronics engineering.

She was with the University of Rochester and Center of Electronic Imaging Systems as a Visiting Research Associate from 1994 to 1996. From 1996 to 1998, she worked as a member of research and technical staff at Xerox Corporation—Digital Imaging Technology Center, Rochester, NY. From 1998 to 1999, she was with the Department of Electrical and Electronics Engineering, Baskent University, Ankara, Turkey. During summer 1999, she was a Visiting Researcher at the Multimedia Labs, New Jersey Institute of Technology (NJIT), Newark. Currently, she is a Professor with the Department of Electrical and Electronics Engineering, Middle East Technical University. Her research interests are in 2-D and 3-D video compression, multimedia streaming, and human activity recognition.

Tolga Capin received the Ph.D. degree in computer science from the Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, in 1998.

He is an Assistant Professor at the Department of Computer Engineering, Bilkent University, Bilkent, Turkey. Before joining Bilkent, he worked at the Nokia Research Center as a Principal Scientist. He has contributed to various mobile graphics standards, including Mobile SVG, JCP, and 3GPP. He has published more than 20 journal papers and book chapters, 50 conference papers, and a book. He has four patents and ten pending patent applications. His current research interests include perceptually aware graphics, mobile graphics platforms, human–computer interaction, and computer animation.

Dominik Strohmeier received the M.S. degree in media technology from Ilmenau University of Technology, Ilmenau, Germany, in 2007, where he is currently working towards the Ph.D. degree in media technology on audiovisual quality evaluations.

He is currently a Researcher at the Ilmenau University of Technology. His research interests are focused on the application of mixed method research approaches in quality of experience evaluation. He has published journal and conference publications in the field of subjective audiovisual quality perception of mobile 3-D video and the related development of new evaluation methods.

Atanas Boev received the M.Sc. degree in micro-electronics from the Technical University of Varna, Varna, Bulgaria, in 2001.

From 2001 to 2002, he was a Researcher with the Institute of Communication Engineering, Tampere University of Technology, Tampere, Finland. In 2003 and 2004, he was Marie Curie Research Fellow with the Department of Signal Processing at the same university. Since 2005, he has been a Researcher with the same department. His research interests are in subjective quality of stereoscopic video and algorithms for optimized visualization on autostereoscopic displays.