Mobile 3DTV System: Quality and User Perspective

Atanas Gotchev[1]   Satu Jumisko-Pyykkö[2]   Atanas Boev[1]   Dominik Strohmeier[3]

[1]Department of Signal Processing,
Tampere University of Technology
name.surmane@tut.fi

[2]Unit of Human-Centered Technology
Tampere University of Technology
satu.jumisko-pyykko@tut.fi

[3]Media Technology Department,
Technical University Ilmenau
dominik.strohmeier@tu-ilmenau.de

ABSTRACT
While mobile 3DTV system components such as stereo video compression techniques, transmission channels and auto-stereoscopic displays are available and with good level of maturity, their joint work crucially depend on the quality and user acceptance. We address these two key factors by rigorously investigating all possible artifacts appearing at all processing stages and influencing different layers of the human 3D vision, and by applying a novel user-centered design to the critical components of the system.

Categories and Subject Descriptors
I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *depth cues, motion, stereo, time-varying imagery.* I.4.1 [**Image Processing and Computer Vision**]: Digitization and Image Capture – *imaging geometry.* I.4.2 [**Image Processing and Computer Vision**]: Compression – *approximate methods.* I.4.3 [**Image Processing and Computer Vision**]: Enhancement – *filtering.* I.4.5 [**Image Processing and Computer Vision**]: Reconstruction – *transform methods.* I.4.10 [**Image Processing and Computer Vision**]: Image Representation – *multidimensional.* I.3.3 [**Computer Graphics**]: Picture/Image Generation – *display and viewing algorithms.* H.5.1 [**Information Interfaces**]: Multimedia Information Systems – *video.* H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, human information processing.*

General Terms
Algorithms, Measurement, Performance, Design, Reliability, Experimentation, Human Factors, Standardization, Verification.

Keywords
Mobile 3DTV, DVB-H, auto-stereoscopic displays, stereo video, user-centered design.

1.INTRODUCTION
3DTV and mobile TV are two emerging technologies in the area of audio-video entertainment and multimedia. While the former is

expected to bring 3D cinema to and replace the HDTV at users' homes, the latter is expected to appeal to mobile users. 3DTV has been supported by manufacturers and developers of large size displays, such as Philips, and Samsung. Research projects, such as ATTEST [1] and 3DTV [2] have addressed various aspects of 3DTV content creation, coding, delivery and system integration [3], [4], [5]. As of mobile TV, results from pilots on broadcast mobile TV services across Europe have revealed clear consumer demand for such technology. In addition, the European Commission has backed such services by identifying DVB-H as the single European standard for mobile TV. First commercial DVB-H TV broadcasts have started in several European countries with a number of compatible handheld terminals by Nokia, Samsung, Motorola and LG.

3D on large screens versus mobile multimedia… Would these two rather diverging technologies converge in a development to be called *Mobile 3DTV*? Experts have predicted that 3DTV would be adopted at homes at least 10 years far from now after the HDTV wave gets back [6]. If to appeal to the users sooner, 3D has to be introduced through the 'back doors' of some *more dynamic* and *novel technology receptive* market niches, such as mobile devices. Thus, the challenge is to adapt, modify and advance the 3D video technology, originally targeted for large screen experience, for the small displays of handhelds. This challenge has been addressed by the recently started FP7 project, entitled 'Mobile 3DTV Delivery over DVB-H system' (MOBILE3DTV). The goal of this paper is to briefly present our mobile 3DTV concept, with a focus on the user issue, as little is known about the user experience of 3D video content visualized on a portable screen. We present our approach to analyzing 3D-specific visual artifacts appearing in all processing stages. We then specify a user-centric approach to investigate the user needs and expectations of mobile 3DTV content. By putting these two approaches together we would like to bring a quality and user perspective of a technology being under research.

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

At the stage of 3D content creation and coding currently there is no single and generally adopted representation format for stereo video, taking specific mobile channel conditions into account. Most natural is to have two-channel stereo video. Capture of such video by synchronized cameras is relatively easy and the coding can be done efficiently within the framework of the emerging

multi-view coding (MVC) amendment of the H.264 AVC standard [4]. However, compared to simulcast, the gain of MVC compression is not high. An alternative is offered by the representation known as single-view video plus depth format, already standardized under the notion of MPEG-C, Part 3 [4]. Such representation leads to good compressibility, as the depth adds less than 20% to the bit-budget of a single video channel [6] but requires additional techniques for depth estimation at the content creation side and depth image based rendering at the receiving side. A new concept of mixed spatial resolution is expected to cope with the problems of fast rendering and efficient compression [4]. All three above-mentioned data representations are being analyzed, compared and optimized within the scope of the MOBILE3DTV project.
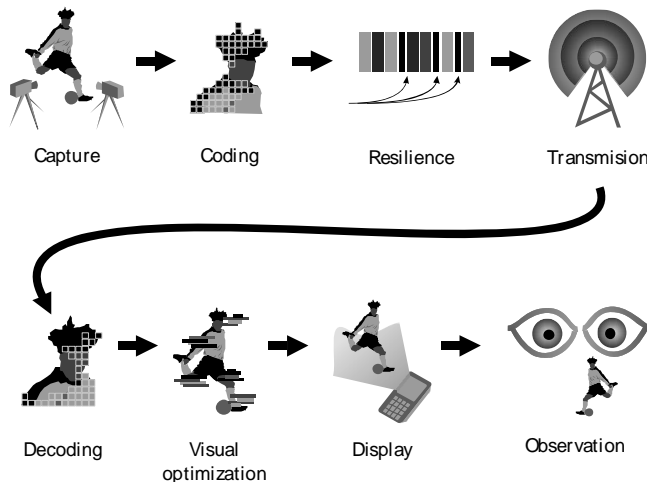

Figure 1. Mobile 3DTV concept

DVB-H is considered to be the broadcast media of future mobile 3DTV. The standard has been a very successful development from the initial idea through comprehensive research and development to commercial services. Key issue is its flexibility. It is not just a tiny TV channel but rather general and powerful data broadcast technology. Would it be appropriate for carrying stereo video is a tempting research problem being studied by the MOBILE3DTV project. So far, DVB-H has been extensively studied for its capability to provide error protection related with the importance of the content to be transmitted (i.e. unequal error protection - UEP) [8]. However, nothing has been done so far concerning the error protection of stereo-video content over such a channel. It might be well protected by the current tools but it might also turn that it needs novel and more comprehensive UEP schemes.

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

As seen from the above brief overview, the components of the future technology are there. However, they should be put and optimized to work together. There are two key issues in this assembly. *Quality* and *user experience*. The next two sections describe our approach in studying these.

3. STEREOSCOPIC ARTIFACTS

We address the issue of quality as affected by stereoscopic artifacts arising from different stages of processing and delivery of stereo video, as presented in Figure 1.

• Capture and content creation – special care should be taken when positioning cameras or selection of rendering parameters. Un-natural correspondences between the images in a stereo-pair (i.e. vertical disparity) are source of several types of artifacts.

• Representation –format conversion stage is a source of artifacts [3]. Some classes of artifacts are typical for one format but not appearing in another, e.g. in dense depth video, disocclusion artifacts are common, while vertical parallax does not occur.

• Coding – typical coding schemes utilize temporal, spatial or inter-channel similarities of a 3D video and algorithms designed for single-channel video are often adapted for stereo [7]. This might result in altering of image details, important for proper depth perception.

• Transmission – in digital transmission, the most common problem is packet loss. Resilience and error concealment algorithms attempt to mitigate the impact on the received video, however, if not designed for stereo-video, such algorithms might introduce additional artifacts on their own.

• Visualization – there are various approaches for 3D scene visualization, which offer different degree of scene approximation [9]. Each family of 3D displays has its own characteristic artifacts, and the artifacts are often scene dependant.

Human visual system is a set of separate subsystems, which operate together in a single process. It is known that spatial, colour and motion information is transmitted to the brain using largely independent neural paths [10]. Vision in 3D, in turn, also consists of different "layers" which provide separate information about depth of the observer scene. This is true both for perception and cognition – on perceptual level there are separate visual mechanisms and neural paths, and on cognitive level there are separate families of depth cues, with varying importance from observer to observer. The depth cues used for assessing the depth by different layers in human vision can be summarized as follows:

• Accommodation – this is the ability of the eye to change its refraction power in order to focus on objects at various distances. Accommodation is the primary depth cue for very short distances, where an object is hardly visible with two eyes. With the distance, the importance of this depth cue quickly decreases. However, the information from other depth-assessing systems is unconsciously used to correct the refraction power, to ensure clear image of the object being tracked. As a result, a discrepancy between accommodation and binocular depth cues leads to so-called *accommodation-convergence rivalry*, which is a major limiting factor for auto-stereoscopic displays.

• Binocular depth cues – these are a consequence of both eyes observing the scene at slightly different angles. The mechanism of binocular depth estimation has two parts – vergence and stereopsis. *Vergence* is the process, in which both eyes take a position which minimizes the difference of the visual information projected in both retinae. The angle between the eyes is used as a depth cue. With the eyes converged on a point, *stereopsis* is the process which uses the residual disparity of the surrounding area for depth estimation relative to point of convergence. Binocular depth cues are the ones most often associated with "3D cinema".

However, binocular vision is quite vulnerable to artefacts – lots of factors can lead to an "unnatural" stereo-pair being presented to the eyes. As HVS is not prepared to handle such information, binocular artefacts can lead to nausea and "cybersickness" [11]. It is worth saying, that around 5% of all people are "stereoscopically latent" and have difficulties assessing binocular depth cues [10]. Such people have a perfect depth perception, only they rely mostly on depth cues coming from other visual "layers".

• Pictorial cues – for longer distances, binocular depth cues become less important, and HVS relies on pictorial cues for depth assessment. These are depth cues that can be perceived even with a single eye – shadows, perspective lines, texture scaling. However, even for medium distances, stereoscopically good image can be "ruined" by the missing of subtle pictorial details, and the scene can be perceived as suffering "puppet theatre" or "cardboard effect" artifacts.

• Motion parallax – this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. The same mechanism is used by insects, and is commonly known as "insect navigation" [12]. Artifacts in the temporal domain (e.g. motion blur, display persistence) will affect the motion parallax depth cues.

Experiments with so-called "random dot stereograms" have shown that binocular and monocular depth cues are independently perceived [13]. Furthermore, the first binocular cells (cells that react to a stimulus presented to either of the eyes) appear at a late stage of the visual pathways – 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 [10]. This led to our assumption that "2D" (monoscopic) and "3D" (stereoscopic) artifacts would be independently perceived. The planar "2D" artifacts, such as noise, ringing, etc, are thoroughly studied in the literature [14]. In our approach, we concentrate on artifacts which affect stereoscopic perception. However, due to the "layered" structure of HVS, binocular artifacts might be inherited from other visual "layers", e.g. blockiness is a "purely" monoscopic artifact, which still can destroy or modify an important binocular depth cue.

In summary, stereoscopic artifacts might be created during various stages in the mobile 3DTV value chain, and might affect different "layers" of human 3D vision. A full nomenclature of 3D artifacts is being summarized within the scope of the MOBILE3DTV project, so to find out which of them are most pronounced for the considered value chain and the ways they can be effectively suppressed or mitigated.

4.USER PERSPECTIVE

This section presents the idea to apply user-centered design for mobile 3DTV. We outline basic principles of human-centered design and present the idea of applying it for the critical component of the system. We also describe existing user studies of mobile television and 3DTV systems to facilitate the discussion of careful user requirement elicitation for novel mobile 3D television.

4.1Human-centered design

The nature of human-centered designs process is cyclic. It includes active user involvement to understand user's requirements, iterative design and evaluation and moreover, a multidisciplinary approach [15]. Figure 2 describes the overall

user-centered design process. The benefits of applying user-centered design methods are announced for example in the terms of better end-user satisfaction and lower costs of system development [16].
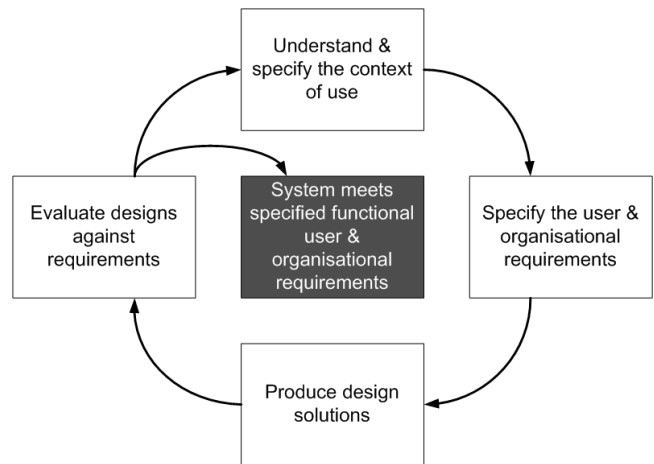

Figure 2 Human-centered design process for interactive systems based on ISO 13407 Standard [15]

User requirements are the earliest steps in defining the system. To become successful, the system needs to satisfy its end-users. MOBILE3DTV project is characterized as a system development project and therefore high-level emphasis is given for the user requirement elicitation *at the beginning*. The requirement elicitation starts in interacting with users or customers about their needs. User needs and expectations reflect the aspects of desires and concerns about the system. They are problems that hinder users in achieving their goals, or as opportunities to improve the likelihood of users achieving their goals in certain context [17]. At the end, user requirements represent any externally visible function, non-functional property or constrain that is required in order to satisfy the user needs [17].

We apply human-centered design to overall system development, but also as a part of development of the critical system components. To provide the well-functioning mobile 3DTV in future, the effort of different players in the field need to be combined from content production and packaging, its delivery and transmission and reception including device and its display. Throughout the whole value chain, perceived quality is one of the main critical components impacting on the final user-experience of 3D. For example, the visible impairments (artifacts) resulted from different factors in the chain can decrease, destroy the 3D viewing experience, or cause discomfort. When improving and evaluating these critical non-functional system components we apply gathered user requirements to simulate the characteristics of end-product. Our aim us to use early stage prototyping parallel to the critical component improvement and evaluation.

4.2Experienced quality of critical components

The approach to examine experienced quality of critical system component is recently introduced [18]. Subjective quality evaluation (~sensorial, perceptual, affective, experienced evaluation) is based on human judgments of various aspects of experienced material based on human perceptual processes [cf. [18][19]. These experiments can be conducted for example to compromise display artifacts, coding or transmission parameters.

Majority of the subjective quality evaluation experiments follow Telecommunication Union (ITU) research methodological recommendations which offer guidelines to design and run the experiments in the controlled environment [20]. However, many of these studies disregard the user characteristics, viewed contents, context of use and actual goals of viewing even the experiments are targeted consumer products or services.

To go beyond these shortcomings, user-centered approach has started to take its first steps. Experienced quality of critical system components aim to ensure that qualities of components, developed in isolation from the end-product, are not barriers for adoption of end-products, and therefore their acceptability should be studied in their optimization process [18]. We use recent work of mobile TV to illustrate this approach. For example, selection of potential viewers and impact of their background factors, like interests in the content and technology attitude has been as a part of quality evaluation studies [21]. Potential television content has been also used and the first steps to introduce contextual quality evaluation procedure parallel to controlled laboratory measures have been taken [22]. Finally, there is also attempt to understand the interpretation of constructed quality [19]. Our aim is to develop further user-centered approach to experienced quality of critical system component and this work needs careful user need and expectation elicitation as a base.

4.3Related work: mobile TV and 3DTV
To understand user needs and expectations about mobile 3DTV, we first examine the relation to existing systems of 3DTV and mobile TV. Knowledge about users, content and context of use of both services is an important basis to develop and introduce user-oriented mobile 3DTV services into the consumer market.

Related studies about users of mobile TV and 3DTV services show that users are described rather differently. Research into mobile TV user description clearly describes a general user. According to Carlsson and Walden [24], the typical user of mobile TV is aged between 23 and 35, he is male, has got a university degree and a yearly income of 20,001 to 30,000 euro. His motivation to use the system is to kill time while waiting or to stay up-to-date about daily news while being on the move [31]. In contrast, users' motivation to use 3DTV is to get entertained. For them, presence [32], [30], the feeling of being there seems to be the killer experience.

The different motivations of using mobile TV and 3DTV also describe the preferred content of users in both systems. While mobile TV users prefer short clips of news and weather, sport program or music videos [25], users of 3DTV expect to watch action movies, films with special 3D effects or documentaries [26] with possibilities to experience the new feeling of presence and to explore the content sufficiently [33].

But besides understanding users and contents, it is important to know the usage context. Currently, contextual studies about 3DTV are rare due to few running systems. Users expect entertainment of 3DTV in environments like 3D cinemas. In contrast to 3DTV, heterogeneous mobile TV contexts are studied better. The main situations to use mobile TV are at work during breaks, while commuting, at home to create privacy, and in waiting or waste-time situations [27], [29], [31]. Related to social context, mobile TV seems to be private viewing which counts for most situations. However, users also use shared viewing like to co-view during lunch or just to share funny stories or clips [27], [31].

These requirements of different contexts using mobile TV also reflect the main principles of designing services. Users want to access content on-demand [28] and intuitively without complicated searches. Services need to provide possibilities to pause and replay content or to switch to audio only cases to facilitate the transitions between the mobile contexts. [29], [31].

Summarizing, user requirement studies into mobile TV and 3DTV provide detailed knowledge about users, content and usage context but also reveal that mobile 3DTV is not just combining mobile TV and 3DTV. To include both requirements into one existing system will be an important step to successfully develop and introduce mobile 3DTV.

4.4User requirement elicitation
User-requirement elicitation process for mobile 3DTV applied methodological triangulation to tackle explicit and implicit aspects of user needs and expectations. The survey and focus group methods represent explicit, conventional, and popular methods in the requirement gathering whereas probe study represents innovative data-collection targeting more on implicit needs [23]. In these three studies we examined user, system and usage context characteristics: Firstly, we examined on user characteristics and motivations for using mobile 3D television. Secondly, we studied preferred system including television and other video contents, device and service features. Thirdly, we explored characteristics of mobile usage context including the aspects of temporal, physical, social and task environment.

The data-collection phase has finished resulting in broad view to requirements. Altogether 350 respondents filled the questionnaire. Eight focus groups were conducted with four different target groups. Data from a month lasting probe study with ten participants was collected. The next steps will be detailed analysis of the results.

5.CONCLUSIONS
We presented our approach to quality and user studies within the MOBILE3DTV project. As the project started few months ago the first results on studying stereoscopic artifacts and user needs and expectations are just being analyzed. A rather broad nomenclature of stereoscopic artifacts resulting from processing modules put versus human visual system layers is being summarized so to help in revealing which of these artifacts are more pronounced and annoying in the mobile 3DTV value chain. As for the user studies, a user-centered approach has been adopted in order to help in optimizing critical components of the system.

6.ACKNOWLEDGMENTS

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

[2] www.3dtv-research.org

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

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

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

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

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

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

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

[10] Wandell, B.A. 1995. Foundations of vision. Sinauer Associates.

[11] McCauley, M., and Sharkey, T. 1992. Cybersickness: Perception of Self-Motion in Virtual Environments. in Presence: Teleoperators and Virtual Environments, 1(3), 311-318.

[12] Wexler, M., and Boxtel, J. 2005. Depth perception by the active observer. Trends Cogn Sci 9 (Sept 2005), 431-438.

[13] Julesz, B. 1971. Foundations of Cyclopean Perception, The University of Chicago Press, Chicago.

[14] Yuen, M. 2005. Coding Artifacts and Visual Distortions. In Digital Video Image Quality and Perceptual Coding. H. Wu and K. Rao, eds, CRC Press.

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

[16] Maguire, M., Bevan, N. 2002. User requirement analysis – A review of supporting methods. In Procs IFIP 17th World Computer Congress, (Motreal, Canada 2002), 133-148.

[17] Kujala, S. 2002. User Studies: A Practical Approach to User Involvement for Gathering User Needs and Requirements. Doctoral dissertation. Espoo: the Finnish Academies of Technology. ISBN 951-666-599-3.

[18] Jumisko-Pyykkö, S. Kumar Malamal Vadakital, V., Hannuksela, M.M. 2008. Acceptance Threshold: Bidimensional Research Method for User-Oriented Quality Evaluation Studies. To appear in Int J Dig MM Broadcast.

[19] Jumisko-Pyykkö, S. Häkkinen, J., Nyman, G. 2007. Experienced Quality Factors - Qualitative Evaluation Approach to Audiovisual Quality. In Proc. SPIE 6507.

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

[21] Jumisko-Pyykkö, S. & Häkkinen, J. 2008. Profiles of the Evaluators - Impact of Psychographic Variables on the Consumer-oriented Quality Assessment of Mobile Television. in Proc. SPIE 6821.

[22] Jumisko-Pyykkö, S. & Hannuksela, M. M. 2008. Does Context Matter in Quality Evaluation of Mobile Television? In Proc. 10th Int. Conf. on Human Computer Interaction with Mobile Devices and Services (MobileHCI).

[23] Hanington, B. M. 2003. Methods in the Making: A Perspective on the State of Human Research in Design.Design Issues 19 (4), (MIT Press, Cambridge) 9 –18.

[24] Carlsson, P., Walden, P. 2007. Mobile TV--To Live or Die by Content. In Proc. 40th Hawaii International Conference on System Sciences. (Jan. 2007), 51.

[25] Knoche, H. O., McCarthy, J. D. 2004. Mobile Users Needs and Expectations of Future Multimedia Services. In Proc. WWRF12. (Nov. 2004, Toronto, Canada)

[26] Freeman, J., Avons, S. E. 2000. Focus Group Exploration of Presence through Advanced Broadcast Services. In Proc. SPIE 3959 , 3959-3976.

[27] O'Hara, K., Mitchell, A. S., and Vorbau, A. 2007. Consuming video on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03, 2007), 857-866. DOI= http://doi.acm.org/10.1145/1240624.1240754

[28] Stockbridge, L. 2006. Mobile TV: Experience of the UK Vodafone and Sky service. In Proceedings of EuroITV'06.

[29] Oksman, V., Noppari, E., Tammela, A. , Makinen, M., and Ollikainen, V. 2007. Mobile TV in Everyday Life Contexts--Individual Entertainment or Shared Experiences? Lect Notes Comput Sc 4471, 215- 225.

[30] Slater, M. 2003. A Note on Presence Terminology. In Presence connect. Bd. 3.

[31] Chipchase, J., Yanqing, C., Jung, Y., and Design, N. 2006. Personal Television: A Qualitative Study of Mobile TV Users, Lect Notes Comput Sc 4471,195-204.

[32] IJsselsteijn, W.A. 2002. Understanding Presence. Proceedings of the AIIA (Siena, Italy, Sep. 2002).

[33] Häkkinen, J., Kawai, T., Takatalo, J., Leisti, T., Radun, J., Hirsaho, A. and Nyman, G.. 2008. Measuring stereoscopic image quality experience with interpretation based quality methodology. Proceedings of the SPIE 6808, 68081B-68081B-12 .