Subjective Quality Assessment of Free Viewpoint Video Objects

Sara Kepplinger
Institute for Media Technology
Ilmenau University of Technology
98693 Ilmenau, Germany
0049 (0) 3677 69 2671
Sara.Kepplinger@tu-ilmenau.de

ABSTRACT
This paper presents an overview on the intended contribution to quality assessment of free viewpoint video representations in the video communication use case within the author's PhD proposal. This proposal will analyze opportunities and obstacles for free viewpoint video objects usage within video communication systems focussing on subjective quality of experience. Quality estimation of emerging free viewpoint video object technology in video communication has not been covered yet and adequate approaches are missing. The challenges are the definition of quality influencing factors, the formulation of a measure, and to link quality evaluation up with technical realization. The paper outlines a description on the theoretical background and intended work. A short description of the related project Skalalgo3d, which offers a useful application framework for the intended work, is included. Preliminary outlined results consist of a tentative research framework, and evaluations conducted so far.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Evaluation / methodology

General Terms
Algorithms, Measurement, Design, Experimentation, Human Factors

Keywords
Free viewpoint video, video communication, methodology, quality of experience

1. INTRODUCTION
Free viewpoint video applications enable the user to navigate interactively and freely within a visual real world scene representation. Applications, like free viewpoint choice on DVD, or similar approaches on TV or online, gain more and more attention in the field of interactive media. Free viewpoint video objects (or 3DVOB) usage within the context of video communication may offer sociability and communication support. This can be achieved by technical possibilities to overcome the obstacles of absent eye contact, or freedom of choice regarding the viewing angle and distance to the dialog partner, for example. These are activities which are possible and usual in real face-to-face conversations. There are different approaches to realize this way of representation using multiple views of the recorded scenes. This complex processing chain can be regenerated in different ways of acquisition, processing, scene representation, coding, transmission, and presentation. This paper is describing the planned efforts within the PhD proposal in order to pay more attention to the user's perception of these new visual representations allowing interactivity. One goal is to define an extended model or an absolute measure for overall quality including subjective quality assessment. Therefore, the correlation between the used algorithm(s) and achieved quality will be considered. The opportunity of this approach is to gain further insights which may be useful for system adaptivity and processing scalability. The challenges of this scheme are mainly (still) open questions about novel algorithms for image analysis and synthesis on one hand, and the development of evaluation and measurement methods of visual quality on the other hand. This emergent field of research is influenced by several different approaches in both: image processing (e.g. [13], [16]), and the inclusion of subjective quality assessment for overall quality estimation (e.g. [17], [12]). In the following, the most relevant work for the author's PhD proposal will be outlined, starting with a short introduction into the technical background. This proposal focuses on the quality assessment within the described technical context and use case. Two main questions are being addressed: How to include the subjective quality estimation by the user? How to identify the most relevant quality influencing factors in order to provide an extended quality model supporting technical optimization? This is outlined in the following way: The problem which will be worked on in this research project is stated in section 2 by explaining the theoretical starting point and intended goals. This is followed in section 3 by a general description of the project Skalalgo3d, approaches chosen within the project, and at related work. Section 4 describes the planned methodological approach. First evaluation steps and preliminary results of previous research will be outlined in short in section 5. This is concluded by section 6 Discussion leading to future work.

2. STATEMENT OF THE PROBLEM
The theoretical starting point is published research concerning the definition of a quality measure for free viewpoint video objects which was developed at the Ilmenau University of Technology [11]. This measure includes definitions of influencing quality parameters as well as measurable characteristics based on objectively quantifiable errors. It is clearly outlined in the description of the measure, called *3DVQM*, that it is open in terms

EuroITV'11, June 29–July 1, 2011, Lisbon, Portugal.

of extension by subjective quality estimation and the definition of to-be identified quality influencing factors and their evaluation [5]. Initial efforts towards an extension of measurements by subjective quality of experience were already made under the usage of synthetic free viewpoint video objects [3]. However, up to now, subjective assessment of natural free viewpoint video objects and the resulting user experience received only little attention and demand more efforts in incorporating early user inclusion [11] (see also section 3. Related Work). A video communication use case is used as a framework regarding eye-contact and other communication based factors. Based on this, and on the technical further development in terms of processing steps, the aim is the identification of further quality influencing factors based on subjective quality assessment. This will lead to the definition of terms and quantifiers and proposed patterns of application of the results to prospective technical developments. Throughout the author's experience, literature analysis, and work within the project, a number of questions arose for the PhD proposal:

- What kind of methodology accounts for reliable subjective quality assessment of free viewpoint video objects in the particular use case of video communication?
- Which further factors influencing free viewpoint video object quality can be identified?
- To which extent do factors, identified by means of subjective quality assessment, influence the overall quality of experience?
- How can the identified factors benefit prospective technical development and processing algorithms of free viewpoint video objects?

These questions within this interdisciplinary approach address mainly methodological questions and practical realization within the area of human computer interaction as well as intended impact on processing development. The novelties the PhD intends to bring about are definitions of (further) quality influencing factors, the way of linking quality evaluation up with technical realization of free viewpoint videos, and therefore the formulation of an adequate measure.

3. RELATED WORK

The work related to the PhD proposal consists of three main topics. Once, there are the evaluation approaches with similar goals. Then, there is the technical realization of free viewpoint video objects in general and their usage for eye contact support in video communication. Preliminary, the project Skalalgo3d is described in this section as the PhD proposal arose within the framework of this project.

3.1 Skalalgo3d

The project Skalalgo3d (Scale able algorithms for 3D video objects under consideration of subjective quality factors) intends to improve free viewpoint video objects and eye contact support used within the context of video communication. The project work of Skalalgo3d is divided into two general working areas. These are the technical realization of free viewpoint video objects and the identification of subjective quality factors. This concerns the optimum processing as well as qualitative displaying under different conditions. It is funded by the German Research Foundation (DFG).

3.2 Technical Realization of 3DVOB

In general, the procedure of 3DVOB generation starts with the acquisition of a time variable and a three dimensional object. The methods of the reconstruction processes differ in principle [5]. They can be either model based or based on disparity analysis or a combination of both. The differences are due to different usage of interpolation, warping, morphing and the recording of several different camera views. The technical development within the project Skalalgo3d is based on the following actual processing chain as shown in Figure 1. This includes the software usage of MATLAB and the project's internal ReVOGS (Realistic Video Object Generation System). A person is recorded by at least two ordinary webcams. This is followed by the processing of first representations out of the recorded scene. This may include adequate pre-processing like colour correction, keying, and calibration. Thereof, second representations are generated by rectification and analysis for accurate disparity determination.

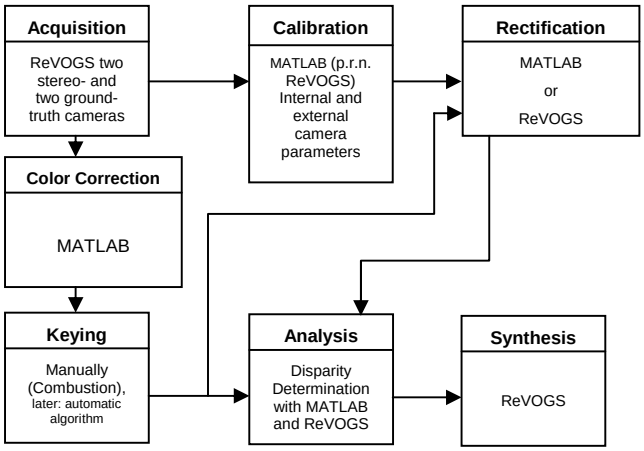

Figure 1. Current status of 3DVOB generation in *Skalalgo3d*

After this, the view synthesis leads to the intended 3DVOB, provided by different and new views.

There are different approaches available for the view synthesis, the disparity analysis and refinement, as well as for the usage of 3DVOB for eye-contact support. They are outlined in following sub-sections.

3.2.1 Different disparity and synthesis methods

Within the development process of the most adequate algorithms to create a qualitative 3DVOB representation, different approaches, concerning disparity and synthesis methods, are considered. These approaches differ in their cost-benefit ratio.

Table 1. Summary of used disparity and synthesis methods

Processing step	**Different methods used**	
View Synthesis	Linear interpolation	Linear interpolation plus median filtering
Disparity analysis	Windowed NCC cost measure	
Disparity refinement	with / without hole filling after cross check	with / without temporal cleaning

In Table 1 a summary of up to now used methods is given. The view synthesis is either only done by accounting the neighbour pixel, or a classical equalization filter is used additionally, in

order to reduce the so called “salted pepper noise”, a visual disorder. The disparity analysis undertaken for the test items is realized by the classical usage of a cost based measure. Within the refinement differences are made by the usage of a hole filling filter or temporal cleaning. These different approaches may result in differently perceived quality concerning representation. A more detailed description on the analysis of disparity in general is outlined in [13].

3.2.2 Support of eye contact in video communication

The problem of usual video communication systems is the impossibility of eye-contact.

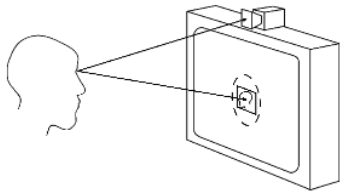

Figure 2. Problem of eye-contact in video communication

The user either has to look on the display to get information or to look into the camera in order to simulate eye-gaze (see Figure 2). Eye-contact is seen as a critical factor in the fields of communication, psychology, and sociology. In 1976 [1] analyzed the role of gaze and mutual gaze in conversations and communication. A possibility offered by the usage of free viewpoint video objects is to support eye-contact via video communication on computer, television, or mobile devices. This can be realized by ways of eye-adjustment or the use of the so called Wollaston illusion by adjusting the displayed person’s position (without manipulating the eyes). Skalalgo3d allows this support by a virtual camera positioning. This approach is described to some extent in [8]. There are already some approaches published concerning eye-contact support. In [10] an approach of virtual view image synthesis for eye-contact in TV conversation system is described. In [15] the effects of gaze direction and depth on perceived eye contact and perceived gaze direction compared between 2D and 3D display conditions are described. In [2] the role of eye gaze in avatar mediated conversational interfaces is analysed. One possible approach to technically realize eye contact via a camera/display system for videophone applications is described in [7].

3.3 Existing Evaluation Methods

Objective quality measures compute metrics representing compare able reference values, mostly focused on technical feasibility. Several defined measures, parameters, and assessment methods are available in order to rate general video quality objectively [17]. Subjective quality measurements intend to include the users and their opinion. This is expressed for example via judgment (e.g. yes/no) or adjustment (e.g. user changes influencing factors and chooses the preferred outcome) resulting in a measure (e.g. mean opinion score) representing this judgments. However, developments, especially in the field of 3D technology, ask for the inclusion of more sophisticated subjective measures in order to reach an adequate consideration of human perception which may differ from the objective rating [6]. In the context of 3DVOB quality assessment, preliminary defined subjective quality factors derive for example from occlusion, distortion, and shape, as outlined in [11]. However, the identification and the tighter definition of the extent of influences ask for further exploration. This is intended by the author’s PhD proposal. Researchers of related research areas (mainly associated with user interface design) already worked on measurement of user experience and user acceptance also in a pre-prototype stage of product development (e.g. [9]). Activities are being undertaken in order to clarify the different usage of efforts to understand the quality of experience of new technologies as summarized in [4]. Research on free viewpoint video object technology up to now mainly regards the technical feasibility. Hence, in this particular emergent field of research user inclusion did not attract much attention up to now. Specific approaches are available concerning the evaluation of video quality in different usage contexts by means of subjective (e.g. [5]) as well as objective measurement (e.g. [6]). However, the quality estimation of emerging free viewpoint video object technology in video communication has not been covered yet [14].

4. METHODOLOGY

First and foremost, the goal of the PhD effort is to identify factors of subjective quality experience (e.g. disturbing fringes, recognized holes, missing eye contact...) and their respective extent of impact as formulated into a measure. To achieve this, an adequate method needs to be defined. There are several methodological approaches paying little attention to subjective factors, besides evaluation efforts on objective measurements (e.g. concerning system processes) in early system development phases, as described in section 3.3. As a consequence, the first step within the proposed evaluation framework is an explorative approach in order to deduct non-critical factors and to define a range of applicable methods for evaluation. In a second step, the application of an - at that time specified - applicable method to collect data about quality influencing subjective factors is verified. The final step is the formulation of a most adequate methodological approach providing a quality measure. This measure provides the possibility of a mathematically formulation of quality influencing factors derived from the users’ perception and therefore may be able to be integrated into the technical processing chain (e.g. in a form of a perceptual coder or something similar).

5. PRELIMINARY RESULTS

As a first approximation to the described topic several methods were conducted in 2009 in order to gain more information. Expert interviews, focus groups, and online questionnaires were held to collect information on possible pre-experiences and users’ ideas about possible free viewpoint video object usage. In 2010 a methodological framework, in cooperation with the Institute of Psychology at the University of Salzburg, Austria, was established by the systematic evaluation of pre-produced test items. The goal was to detect to what extent the resulting quality of different processing steps was acceptable and to examine subjective factors influencing the experienced quality. This included the experience of eye-contact and the measurement of possible influence by several characteristics (e.g. appeal, trustworthiness…) of the shown conversational partner or different conversation contexts (private talk vs. professional conference). The test items were different free viewpoint video objects (produced by the usage of the described processing chain) showing four different people (two men, two women) resembling a possible video communication partner. With the conduction of

this study a total of 322 data sets were collected. The data collection was carried out within three weeks in November 2010 in a laboratory providing a standardized environment (i.e. lightning conditions) in five separate rooms with personal computers and 19" LCD displays. Table 2 summarizes the design of the evaluation study and shows the different pre-defined independent variables.

Table 2. Combination of different evaluation variables

Technical items	Test items (i.e. videos, 10 sec.), with/without eye-contact, different view synthesis and disparity analysis (Table 1)			
Usage context	Private talk with friend		Professional talk to adviser	
Content shown	Man	Woman	Man	Woman

With the possibility of collecting this amount of data sets every possible test setting contains at least 15 data sets. The settings vary in the combination of the different evaluation variables. For first evaluation activities, a set of different test items were created with the usage of the described different methods of view synthesis and disparity analysis (see also Table 1 in section 3.2.1).

6. DISCUSSION AND FUTURE WORK

There are several open questions concerning the identification of subjective quality factors and their measurement. In a first step, within the framework of the author's PhD proposal, data is collected providing a basis for explorative analysis. The analysis of the preliminary data will be organized in three different phases. First of all, quality influencing factors will be identified via categorization and correlation analysis paying attention to the different evaluation variables (Table 2). A weighing of the identified factors leading to a list of influences on quality will be carried out in a second step. This is followed by the subtraction of non-critical factors and the conception of a further evaluation in May 2011. Therefore, it is planned to allow the user to define him or herself the best combination of provided processing steps in order to assess a free viewpoint video object representation with the best experienced quality. The identification or development of an ideal methodology in order to reach above mentioned goals is seen as accompanying needed effort and therefore main part of the overall result. Results will be published gradually within conference publications, the intended PhD work, and within the project report of Skalalgo3d until the end of February 2012.

7. REFERENCES

[1] Argyle, M., Cook, M. 1976. Gaze and mutual gaze. Cambridge University Press, New York, USA.

[2] Colburn, A., R. 2000. The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. In: Avatar Mediated Conversational Interfaces, Technical Report MSRTR-2000-81, Microsoft Research, Microsoft Corporation, One Microsoft Way.

[3] Fan, F., 2008. Analyse von Qualitätsparametern von 3D-Videoobjekten. Diplomarbeit, Technical University Ilmenau, Ilmenau, GER.

[4] Geerts, D., De Moor, K., Ketykó, I., Jacobs, A., Van den Bergh, J., Joseph,W., Martens, L., De Marez, L., 2010. Linking an Integrated Framework with Appropriate Methods for Measuring QoE. In: *Proceedings of the QoMEX 2010, Second International Workshop on Quality of Multimedia Experience* (June 21–23, 2010. Trondheim, Norway), Trondheim, Norway, Paper No. 158.

[5] ITU-R BT.500-11. ITU Recommendation. 2002. Methodology for the subjective assessment of the quality of television pictures. ITU.

[6] Jumisko-Pyykkö, S., Reiter, U., Weigel, C., 2007. Produced quality is not perceived quality – A qualitative approach to overall audiovisual quality. IEEE Xplore.

[7] Kollarits, R. V., Woodworth, C., Ribera, J., F., and Gitlin, R., D. 1995. An eye contact camera/display system for videophone applications using a conventional direct-view lcd. SID 1995, Digest.

[8] Korn, T. 2009. Kalibrierung und Blickrichtungsanalyse für ein 3D Videokonferenzsystem. Diplomarbeit, Technical University Ilmenau, Ilmenau, GER.

[9] Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J. (2009) Understanding, Scoping and Defining User Experience: A Survey Approach. In Proc. Human Factors in Computing Systems, CHI'09. April 4-9, 2009, Boston, MA, USA.

[10] Murayama, D., Kimura, K., Hosaka, T., Hamamoto, T., Shibuhisa, N., Tanaka, S., Sato, S., Saito, S. 2010. Virtual view image synthesis for eye-contact in TV conversation system. In: *Conference Proceedings of Electronic Imaging*, (17-21 January 2010. San Jose Convention Center, San Jose, California, United States) San Jose, USA, Paper No.7526-11.

[11] Rittermann, M. 2007. Zur Qualitätsbeurteilung von 3D-Videoobjekten. Dissertation, Technical University Ilmenau, Ilmenau, GER.

[12] Satu Jumisko-Pyykkö, Dominik Strohmeier, Timo Utriainen, and Kristina Kunze. 2010. Descriptive quality of experience for mobile 3D video. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (NordiCHI '10). ACM, New York, NY, USA, 266-275. DOI=10.1145/1868914.1868947.

[13] Scharstein, D., Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. In: *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

[14] Schreer, O., Kauff, P., Sikora, Th., (eds.) 2005. 3D Video communication: Algorithms, concepts and real-time systems in human centered communication. John Wiley & Sons, Ltd., UK.

[15] Van Eijk, R., Kuijsters, A., Dijkstra, K., IJsselsteijn, W., A. 2010. Human Sensitivity to eye contact in 2D and 3D videoconferencing. In: *Proceedings of the QoMEX 2010, Second International Workshop on Quality of Multimedia Experience* (June 21–23, 2010. Trondheim, Norway), Trondheim, Norway, Paper No. 76.

[16] Weigel, C., Schwarz, S., Korn, T. 2009. Wallebohr, M., Interactive free viewpoint video from multiple stereo. In: *Proceedings of the 2009 3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video* (3DTV-CON 2009, Potsdam, Germany, May 4-6, 2009). IEEE Xplore, DOI = 10.1109/3DTV.2009.5069663.

[17] Winkler, S. 2005. Digital Video Quality – Vision Models and Metrics. John Wiley & Sons, UK.