Motion Parallax Based Restitution of 3D Images on Legacy Consumer Mobile Devices

Martin Řeřábek [#*1], Lutz Goldmann [#2], Jong-Seok Lee [#3], Touradj Ebrahimi [#4]

[#] *Multimedia Signal Processing Group (MMSPG), Ecole Polytechnique Fédérale de Lausanne (EPFL)*
Station 11, Lausanne, Switzerland
[1] martin.rerabek@epfl.ch
[2] lutz.goldmann@epfl.ch
[3] jong-seok.lee@epfl.ch
[4] touradj.ebrahimi@epfl.ch

[*] *Multimedia Technology Group (MMTG), Czech Technical University in Prague (CTU),*
Faculty of Electrical Engineering (FEE), Technicka 2, 166 27 Prague, Czech Republic
[1] rerabem@fel.cvut.cz

***Abstract*—While 3D display technologies are already widely available for cinema and home or corporate use, only a few portable devices currently feature 3D display capabilities. Moreover, the large majority of 3D display solutions rely on binocular perception. In this paper, we study the alternative methods for restitution of 3D images on conventional 2D displays and analyze their respective performance. This particularly includes the extension of wiggle stereoscopy for portable devices which relies on motion parallax as an additional depth cue. The goal of this paper is to compare two different 3D display techniques, the anaglyph method which provides binocular depth cues and a method based on motion parallax, and to show that the motion parallax based approach to present 3D images on consumer 2D portable screen is an equivalent way in comparison to the above mentioned and well-known anaglyph method. The subsequently conducted subjective quality tests show that viewers even prefer wiggle over anaglyph stereoscopy mainly due to a better color reproduction and a comparable depth perception.**

I. Introduction

Recently the interest in 3DTV, as one of the emerging multimedia trends, has remarkably increased due to availability of interesting 3D content and the rapid development of 3D technologies. Especially, 3D restitution technologies have improved considerably and gradually replace 2D technologies.

In order to perceive the world in 3D the human visual system relies on a variety of depth cues which can be grouped into 4 major categories [1]. Accommodation refers to the change of the refraction power of the lens in order to focus on objects at various distances. Monocular depth cues require only one eye and include interposition, perspective, gradients and shadows. Motion parallax occurs due to the relative motion of objects with respect to each other if the observer is moving. Binocular depth cues rely on two eyes and include stereopsis or binocular disparity and convergence. All these depth cues are fused by the human visual system and their importance varies with the scene distance. Motion parallax and binocular disparity are among the most powerful depth cues.

Even using conventional 2D displays humans can perceive depth due to monocular depth cues present within monoscopic images and videos. The large majority of 3D displays enhance the depth perception by adding binocular cues through stereoscopic images and videos. Furthermore, some displays offer motion parallax as an additional depth cue which may improve the perceived depth considerably [2]. The basic principle of a stereoscopic display is to provide different images to the left and the right eyes. Most stereoscopic display technologies require the viewer to wear passive (e.g. anaglyph, polarized) or active (e.g. shutter) glasses which filter the visual information appropriately. On the other hand, autostereoscopic displays use optical components (e.g. parallax barriers, lenticular lenses) to project the images directionally into the viewer's eyes without the need for glasses.

While 3D display technologies are already widely available for cinema and home or corporate use, only a few portable devices such as notebooks (e.g. Asus G74SX), tablets (e.g. LG G-Slate 3D) or mobile phones (e.g. LG Optimus 3D), cameras (e.g. FujiFilm W3 D3) and game consoles (e.g. Nintendo 3DS) currently feature 3D displays. In order to support 3D restitution on portable devices using conventional 2D displays only the following two technologies can be used. *Anaglyph stereoscopy* uses complementary color filters to deliver the different views of a stereoscopic image or video to each of the eyes. The resulting binocular disparity serves as an additional depth cue beside the monocular depth cues but requires additional glasses. *Wiggle stereoscopy* alternates rapidly between the views of a stereoscopic image. In that way it adds motion parallax as an additional depth cue without the need for glasses. In order to achieve a better depth impression this idea can be easily extended to multiple views and interactive view change.

The goal of this work is to study these alternative 3D display technologies for portable devices and compare them to conventional 2D restitution in terms of overall quality and perceived depth. The subjective quality evaluation is performed with a set of multi-view images on a mobile phone.

II. Restitution methods

A. Anaglyph stereoscopy (binocular disparity)

The anaglyph method for displaying stereoscopic images relies on the multiplexing of the individual views into complementary color channels at the display side and a pair of glasses with the corresponding color filters at the viewer side [3]. The resulting difference between the two retinal images, commonly referred to as binocular disparity, serves as a strong depth cue beside the monocular depth cues. Although this method has been already proposed in 1853 [4], it remains a common 3D display technique for conventional 2D displays. Unfortunately, it suffers from a relatively poor 3D image quality due to the inaccurate color reproduction and the high level of crosstalk.

A variety of algorithms have been proposed to convert the stereoscopic image pair into an anaglyph image. The most common algorithms [5] derive the color of pixel p_a in the anaglyph image through a linear combination $p_a = p_l \cdot M_l + p_r \cdot M_r$ of the corresponding pixel p_l and p_r in the left and the right image, respectively. The color anaglyph algorithm used in this work is based on following conversion matrices:

$$M_l = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_r = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

The generated anaglyph image can be directly displayed on the portable device with any standard image viewer.

B. Wiggle stereoscopy (motion parallax)

Wiggle stereoscopy alternates rapidly between the two views of a stereoscopic image. The resulting relative motion between different parts of the image, commonly referred to as motion parallax, serves as a strong additional depth cue beside the monocular cues [6]. However, wiggle stereoscopy does not provide true binocular depth perception and alternating between only two views leads to an annoyingly jerky image. Furthermore, it is only applicable to still images.

In order to improve the 3D quality, wiggle stereoscopy can be extended in two directions. First, more than two views can be used to achieve a smoother restitution. Furthermore, instead of alternating automatically between the images, they can be switched interactively according to the relative position between the display and the user. This idea is illustrated in Figure 1. By rotating the portable device, the viewer controls the view which is rendered on the display. In this way he can view the scene from different positions similar to what he could also do in the real world.

When the number of images in a stereoscopic or multiscopic image set is too small to achieve a smooth restitution, intermediate images are generated using depth image based rendering [7]. The depth estimation and view synthesis tools of the 3D video coding (3DV) framework [8] developed by MPEG are utilized for the content generation process within our experiments. The *depth estimation reference software (DERS)* uses three camera views (left, center, right) together with the intrinsic and extrinsic camera parameters to estimate

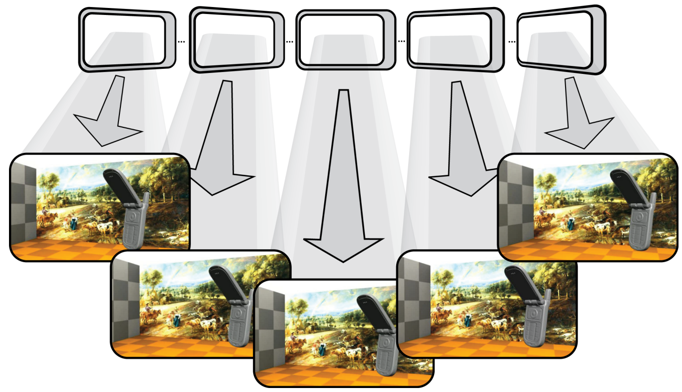


Fig. 1. Motion parallax based 3D restitution.

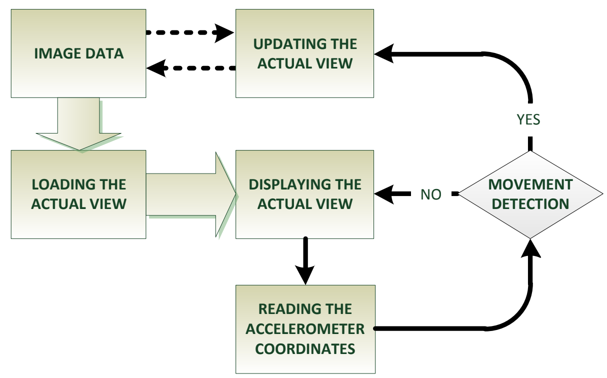


Fig. 2. Content restitution for the motion parallax based display.

the depth map of the center view. It employs graph cuts to find the most likely disparity for every pixel by minimizing a cost energy function that consists of a similarity and a smoothing term. From the three operation modes (automatic, segmentation, semiautomatic) of DERS, the automatic mode has been used. Once the depth maps were obtained the *view synthesis reference software (VSRS)* was used to synthesize intermediate views. A virtual view is generated based on two reference views with the corresponding depth maps as well as the intrinsic and extrinsic camera parameters. This is achieved by depth and texture mapping and hole filling for each of the reference views followed by image blending and inpainting. Finally the spatial resolution of the multiscopic image set is adapted to match that of the portable device. For our experiments a Samsung Galaxy S with a screen resolution of 800x480 pixels was used.

Given the multiscopic image set the goal of the multiview restitution is to display the appropriate view according to the relative orientation between the device and the viewer as shown in Figure 2. Measuring the orientation of the device with respect to the viewer can be achieved through a built-in accelerometer. From the three possible rotations (pan, tilt, roll) of the mobile device only the rotation around the vertical axis (pan) is used. Given the initial pan angle and the predefined

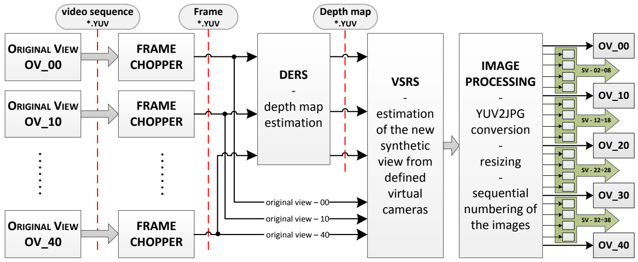


Fig. 3. Content generation for the motion parallax based display.

pan range, the current view can be computed based on the current pan angle. It is displayed until the pan angle reaches an angle that corresponds to another view. Therefore, the display duration of a view depends solely on the speed with which the viewer moves the display.

III. SUBJECTIVE EVALUATION

In order to study the 3D restitution methods for conventional 2D displays on portable devices a subjective quality test has been conducted. It compares the anaglyph and the motion parallax based restitution methods against each other and a normal 2D restitution in terms of overall quality and perceived depth. In other words, it tries to answer to the question of whether the motion parallax based restitution of 3D content on conventional displays could be used as an alternative to the commonly used color anaglyphs.

A. Image dataset

A subset of the multiscopic videos from the MPEG[1] 3D Video (3DV) dataset [9] has been used for the experiments. More specifically 4 videos (Lovebird1, Ballons, Kendo, Mobile) from the class C set have been considered. The first one was used for training and the latter three for testing.

Suitable frames have been selected and extracted from each video. Using the latest version of the depth estimation reference software (DERS 5.1) [8] and the view synthesis reference software (VSRS 3.5) [8], the required number of views were synthesized. Since the display application on the mobile phone does not support the YUV format used by the MPEG tools, the resulting multiscopic image sets were further converted to high quality JPEG images. Finally, frames were converted and cropped to the display dimensions (800x480 pixel resolution, 5:3 aspect ratio) of the used mobile phone (Samsung Galaxy S). The complete image processing chain used for image data preparation is illustrated in Figure 3 and described in more details in section II-B.

Given the resulting multi-view dataset, 5 subsets are created that simulate the different restitution methods (2D, anaglyph, motion parallax) for different camera baselines (narrow, wide) leading to the following test conditions:

[1]http://mpeg.chiariglione.org

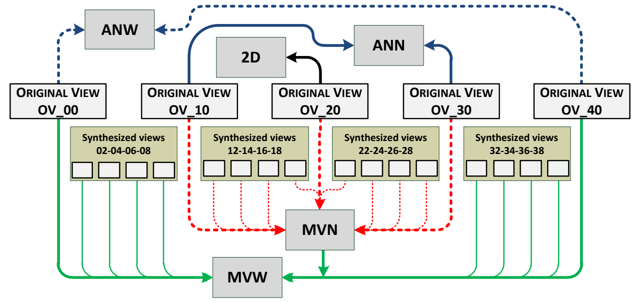


Fig. 4. Creation of the individual test conditions from multiscopic dataset.

2D image (2D): 2D reference image chosen to be the center view of the multi-view dataset.

Narrow Anaglyph(ANN): Anaglyph image set with $10cm$ camera baseline resulting in a smaller depth range.

Wide anaglyph (ANW): Anaglyph image set with a wider camera baseline ($20cm$) resulting in a larger depth range.

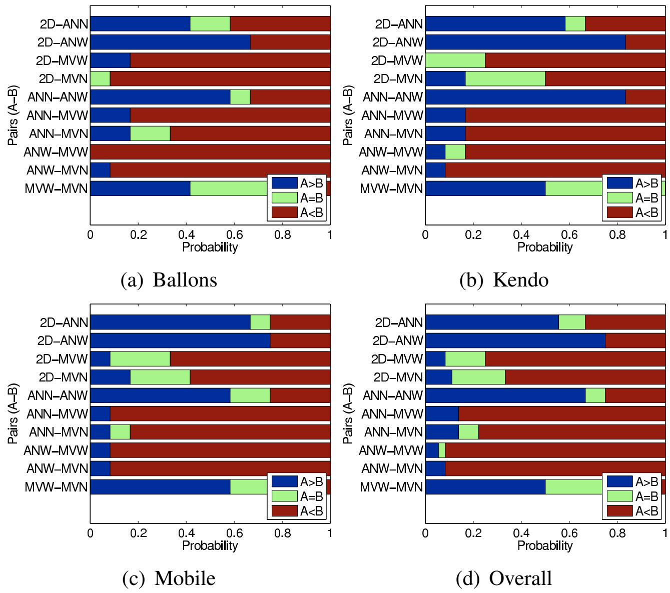
Narrow multi-view (MVN): Multi-view set of 11 images with a narrower camera baseline ($10cm$) resulting in smaller motion parallax.

Wide multi-view (MVW): Multi-view set of 21 images with a wider camera baseline ($20cm$) resulting in larger motion parallax.

This idea of this process is illustrated in Figure 4 for a multi-view image set which consists of 5 original views (OV_00, OV_10, OV_20, OV_30, OV_40) and 4 intermediate synthesized views (SV_02 - SV_08, SV_12 - SV_18,SV_22 - SV_28, SV_32 - SV_38) between each of them. As usual, the camera baseline corresponds the distance between the left and right view. Since camera distance affects directly the perceived depth but also the visual comfort, two different camera distances were considered to identify the optimal one for each of the restitution methods. Subsets with narrower camera baseline (ANN and MVN) utilize the original views OV_10 and OV_30 as the left and right views, respectively. On the other hand, the original views OV_00 and OV_40 are used as left and right views within the wider camera baseline subsets (ANW and MVW). Since the viewing angle range of the mobile device is fixed, the motion smoothness and the depth range are larger for the wider camera baseline.

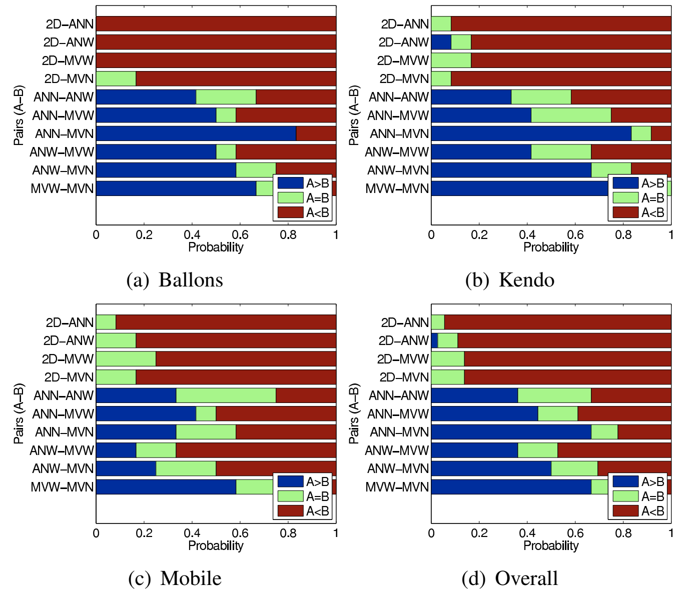
B. Test methodology

Several methods have been proposed for the quality evaluation of 2D [10] and 3D [11] images and videos including single stimulus (SS), double stimulus (DS) and stimulus comparison (SC). Since judging the quality of different 2D and 3D restitution techniques individually may be quite difficult, the



(a) Ballons (b) Kendo (c) Mobile (d) Overall

Fig. 5. Preference and tie probabilities of the individual pairs considering the overall quality.



(a) Ballons (b) Kendo (c) Mobile (d) Overall

Fig. 6. Preference and tie probabilities of the individual pairs considering the perceived depth.

SC method seems to be the most suitable method for the subjective test.

The subjective test were performed as follows. A pair of test stimuli was shown sequentially on the mobile phone using the developed Android viewer application. After comparing the two test stimuli the subject was asked to choose his/her preference (“first” or “second”) in terms of *overall quality* and *depth quality*. The option “same” was also included to avoid random preference selections. For each of the 3 test videos (Ballons, Kendo, Mobile) all the possible combinations of the 5 test conditions (2D, ANN, ANW, MVN, MVW) were considered. This le d to a test set with $3 \times \binom{5}{2} = 30$ paired comparisons.

Since the IPD (Inter-Pupillary Distance) of all participating subjects has not been measured, the two camera baselines, used to generate narrow and wide subset, are compared against each other.

Fifteen subjects (8 male and 7 female) participated in the subjective test experiments. They reported normal or corrected to normal vision according to [11].

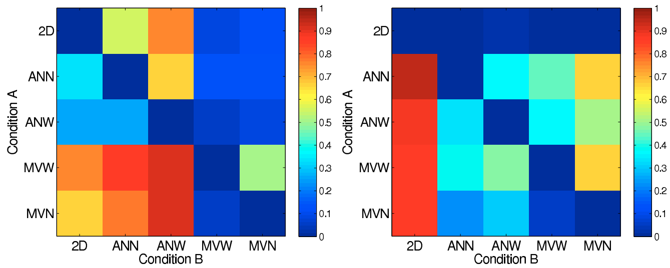
C. Results and discussion

After collecting the preference ratings from the individual users statistical tools were applied to analyze the preferences for the different scenes and test conditions.

The simplest way to analyze a set of paired comparisons is to compute the distribution of the votes over the different categorical levels (first, same, second) and normalizing them by the number of subjects. This can be done individually for each or jointly over all the videos. Figures 5 and 6 show the resulting probabilities for the overall quality and the perceived depth, respectively. With respect to the overall quality the results across the different scenes are very similar. 2D is slightly better (preference probability between 40% and 70%) perceived when compared to anaglyph (ANN, ANW), while it is worse (preference probability between 0% and 20%) when compared to the multi-view (MVN, MVW). Multi-view is generally preferred over anaglyph with a probability between 60% and 90%. The results for the perceived depth are again quite similar across the different scenes. As expected the perceived depth with the 3D restitution methods (ANN, ANW, MVW, MVN) is much better (between 80% and 100 %) when compared to the 2D restitution. The perceived depths of anaglyph and multi-view seem to be quite comparable with slight preference shifts (0.3-0.7) depending on the video. In general the narrow anaglyph (ANN) and the wide multi-view (MVW) are the preferred methods and achieve similar depth perception.

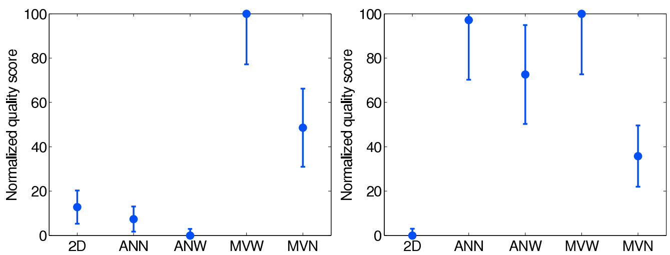
For a more detailed analysis of the individual test conditions and their performance with respect to each other one can construct a preference matrix from the individual paired comparisons by discarding the ties. It provides the preference probabilities of a test condition A versus another test condition B along the rows. Figure 7 shows the preference probability matrices averaged over all the video sequences for the overall quality and the perceived depth, respectively. Analysing the overall quality matrix shows that multi-view is clearly preferred (preference probabilities between 70 % and 90 %) over both 2D and anaglyph. Furthermore, 2D is usually preferred over anaglyph depending on the camera baseline. With respect to the perceived depth 2D clearly looses against all the 3D restitution methods. The narrow anaglyph (ANN) and the wide multi-view (MVW) achieve the best depth quality (preference probabilities between 40 % and 60 %) followed by the wide anaglyph (ANW) and the narrow multi-view (NWV).

Based on the preference probability matrix one can obtain continuous quality scores by applying the Bradley-Terry-Luce



(a) Overall quality (b) Perceived depth

Fig. 7. Preference probabilities of condition A vs. B.



(a) Overall quality (b) Perceived depth

Fig. 8. MOS/CI of the individual test conditions.

(BTL) model [12]. In this model the preference probability P_{ij} of choosing i over j can be represented as

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \quad (2)$$

Given that $\pi_i \geq 0$ and $\sum_i \pi_i = 1$ for all i, the individual π_i can be computed through maximum likelihood estimation based on the empirical probabilities P_{ij}. Ties between a pair ij are considered as half way between the two preference options and therefore equally distributed between P_{ij} and P_{ji} [12]. In addition, the CI (Confidence Interval) for the maximum likelihood estimates of the scores can be obtained from the Hessian matrix of the log-likelihood function. The final quality scores are obtained by normalizing π_i into the range $[0, 100]$. Figure 8 shows the obtained MOS (Mean Opinion Score) and CI of all the scenes for the overall and the depth quality, respectively. A comparison of the overall quality scores for the different display techniques shows that multi-view clearly outperforms anaglyph and 2D. Furthermore, MVW achieves an overall quality score of 100 which is approximately twice higher than MVN. For the depth quality scores the situation is slightly different. The scores for all the 3D restitution techniques are much better when compared to the 2D restitution. ANN and MVW achieve the highest score with a MOS of 100, followed by ANW with 70 and MVN with 35.

IV. Conclusion

Within this work we have studied alternative 3D restitution techniques for conventional 2D displays including well-known anaglyph stereoscopy and an extension of wiggle stereoscopy towards interactive multi-view. The subjective quality test shows that the additional depth cues provide a better depth perception when compared to simple 2D restitution. While the depth quality of 3D resitution techniques is comparable, wiggle stereoscopy is preferred in terms of overall quality due to the inaccurate color rendering and the crosstalk of anaglyph stereoscopy.

This initial study may be extended into several directions. Since the camera distance has a large influence on the perceived 3D quality for both restitution methods, it should be studied in more detail. Furthermore, the considered techniques will be also compared against other 3D restitution techniques which require specialized displays.

Acknowledgment

This work has been performed under Scientific Exchange Programme NMS-CH project 09.040 - CHVI, EC funded Network of Excellence Petamedia and COST Action IC1003 Qualinet.

References

[1] A. Gotchev, S. Jumisko-Pyykkö, A. Boev, and D. Strohmeier, "Mobile 3dtv system: Quality and user perspective," *Procs of EUMOB*, 2008. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/Gotchev08_EUMOB.pdf

[2] C. Zhang, D. Florencio, and Z. Zhang, "Improving immersive experiences in telecommunication with motion parallax [applications corner]," *Signal Processing Magazine, IEEE*, vol. 28, no. 1, pp. 139 –144, jan. 2011.

[3] A. Woods and C. Harris, "Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3d glasses (proceedings paper)," 2010.

[4] R. Zone, "Good old fashion anaglyph: High tech tools revive a classic format in spy kids 3-d," *Stereo World*, vol. 29, 2002.

[5] D. F. McAllister, Y. Zhou, and S. Sullivan, "Methods for computing color anaglyphs," in *Stereoscopic Displays and Applications XXI*.

[6] G. Peters, "Theories of three-dimensional object perception - a survey," Recent Research Developments in Pattern Recognition, 2000. [Online]. Available: http://www.inf.fh-dortmund.de/personen/professoren/peters/publication_sources/RecResDevInPatRecRes.pdf

[7] C. Fehn, "Depth image based rendering (dibr), compression and transmission for a new approach on 3d-tv," 2004. [Online]. Available: http://iphome.hhi.de/fehn/Publications/fehn_EI2004.pdf

[8] MPEG, "Draft report on experimental framework for 3d video coding," ISO/IEC JTC1/SC29/WG11, Tech. Rep. N11478, July 2010.

[9] M. video group, "Description of exploration experiments in 3d video coding," *ISO/IEC JTC1/SC29/WG11 N9466*, no. N11630, Oct. 2010.

[10] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep. BT.500-11, 2002.

[11] ——, "Subjective assessment of stereoscopic television pictures," ITU-R, Tech. Rep. BT.1438, 2000.

[12] M. Glickman, "Parameter estimation in large dynamic paired comparison experiments," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 48, no. 3, pp. 377–394, 1999. [Online]. Available: http://www.jstor.org/stable/pdfplus/2680831.pdf