COMPARISON OF 3D PORTABLE DISPLAY RESTITUTION TECHNIQUES BASED ON STEREO AND MOTION PARALLAX

Martin Řeřábek [12], *Touradj Ebrahimi* [1]

[1] Ecole Polytechnique Fédérale de Lausanne (EPFL),
Multimedia Signal Processing Group (MMSPG),
Station 11, Lausanne, Switzerland
[2] Czech Technical University in Prague (CTU),
Multimedia Technology Group (MMTG), Faculty of Electrical Engineering (FEE),
Technicka 2, 166 27 Prague, Czech Republic

ABSTRACT

Currently, the large majority of 3D display solutions rely on binocular perception phenomena. Furthermore, 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. In this paper, we study the alternative methods for restitution of 3D images on displays of commercially available portable devices and analyse their respective performance. This particularly includes the restitution method which uses multiscopic image data-set and which relies on motion parallax as an additional depth cue. The goal of this paper is to compare three different commercially available 3D display techniques, the auto-stereoscopic and anaglyph method which provide binocular depth cues and a method based on motion parallax. The subsequently conducted subjective quality tests show that the motion parallax based approach to present 3D images on consumer portable screen is an equivalent and quite competitive way in comparison to the above mentioned stereopsis based methods.

Index Terms— 3D restitution, portable device, binocular disparity, motion parallax, auto-stereoscopic display

1. INTRODUCTION

In order to perceive the world in 3D, the human visual system relies on a variety of depth cues which can be grouped into the following 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 and the observer with respect to each other. *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 video sequences. The large majority of 3D displays enhance the depth perception by adding binocular cues through stereoscopic images and video sequences. 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 glasses which filter the visual information appropriately. On the other hand, auto-stereoscopic displays use optical components to project the images directly into the viewer's eyes without the need for glasses.

The goal of this work is to enhance the previous study [3], where only three restitution techniques (2D, anaglyph and motion parallax) have been compared, by introducing new restitution method based on auto-stereoscopic display. Thus, this paper studies all above mentioned 3D display technologies for portable devices and compares them to conventional 2D restitution in terms of overall quality and quality of the perceived depth. Moreover, subjective quality evaluation is performed with a set of multi-view images on a mobile phone.

2. 3D TECHNOLOGIES FOR PORTABLE DISPLAYS

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, HTC EVO 3D), cameras (e.g. FujiFilm W3 D3) and game consoles (e.g. Nintendo 3DS) currently feature 3D displays. Currently, most portable devices with 3D features employ auto-stereoscopic 3D display based on occlusion (parallax

barrier) or refraction (lenticular lenses) approaches and allow to show a stereoscopic image without 3D glasses. On the other hand, assuming the usage of portable devices with conventional 2D displays, only the following two principles with corresponding technologies can be used for rendering and perception of 3D content. *Stereo parallax based anaglyph method* which 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. *Motion parallax based wiggle (multi-view) stereoscopy* which alternates rapidly between the views of a stereoscopic (multiscopic) image. In that way it adds motion parallax as an additional depth cue without the need for glasses.

2.1. Stereo parallax

Stereo parallax is related to the ability to see a scene from two different points which enables to gain two slightly different images of the scene. The resulting difference between the two images, commonly referred to as binocular disparity, serves as a strong depth cue. In order to provide binocular depth cues to the observer, different views have to be sent to each eye. This can be achieved with (stereoscopic) or without (auto-stereoscopic) glasses.

2.1.1. Auto-stereoscopic displays

There are many competing approaches [4] to resolve the issue of auto-stereoscopic display. We will address the parallax barrier as one of the technologies of direction multiplexed displays which are mainly used for the portable devices. Direction multiplexed auto-stereoscopic displays project the appropriate information directly into each eye by applying the various optical phenomena (diffraction, refraction, reflection and occlusion). In case of parallax barrier, a special layer and its occlusion effect is used in order to direct the light emitted by pixels of column interlaced stereo image exclusively to the appropriate eye, see the Fig. 1. Column interlaced stereo image is usually rendered automatically in device from left and right stereo image format. One of the main disadvantages of using auto-stereoscopic display for portable devices is the reduction of horizontal resolution.

2.1.2. Anaglyph

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 [6]. One of the biggest advantages of anaglyph method is that they can be used with standard 2D displays and printers. Unfortunately, it suffers from a relatively poor 3D image quality due to the limited or inaccurate color reproduction,
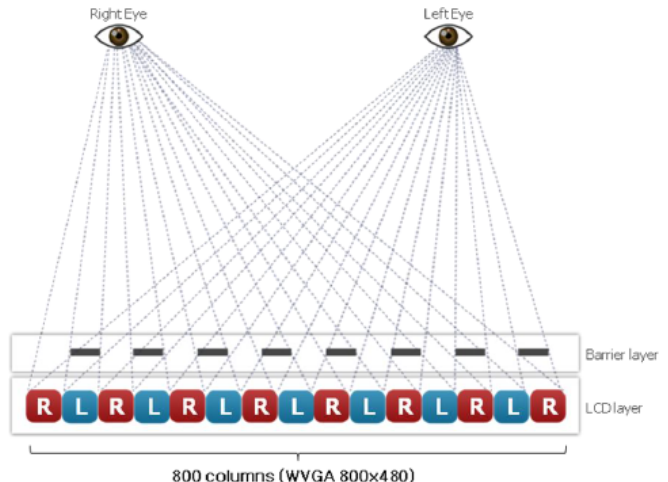

Fig. 1. The principle of parallax barrier implementation in LG Optimus 3D mobile phone [5].

increased level of cross-talk and retinal rivalry in comparison to the other restitution methods. A variety of algorithms have been proposed to convert a stereoscopic image pair into an anaglyph image. The most common algorithms [7] 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:

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

The generated anaglyph image can be directly displayed on any conventional 2D display with any standard image viewer.

2.2. Motion parallax

Motion parallax, which serves as a strong additional depth cue [8], derives benefit from the relative motion between different parts of the image and the observer. One of the method using the motion parallax is *wiggle stereoscopy* where the rapid alternation between the two views of a stereoscopic image occurs. However, wiggle stereoscopy does not provide true binocular depth perception and alternating between only two views leads to a resulting 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.
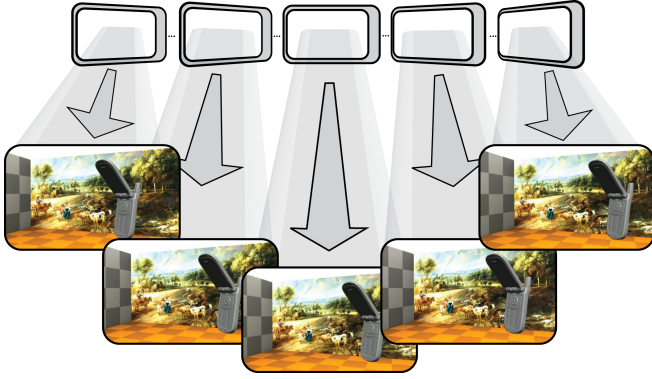
Fig. 2. Motion parallax based 3D restitution.

2.2.1. *Multi-view motion parallax*

The idea of portable device restitution method based on motion parallax employing the multi-view image data-set is illustrated in Fig. 2. By rotating the portable device, the viewer controls the view which is rendered on the display and so he/she can observe the scene from different positions. This way the viewer imitates a process similar to what he/she does when observing 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 [9]. The depth estimation and view synthesis tools of the 3D video coding (3DV) framework [10] developed by MPEG are used 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 the depth map of the center view. From the three operation modes (automatic, segmentation, semi-automatic) 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.

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 Fig. 3.

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 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
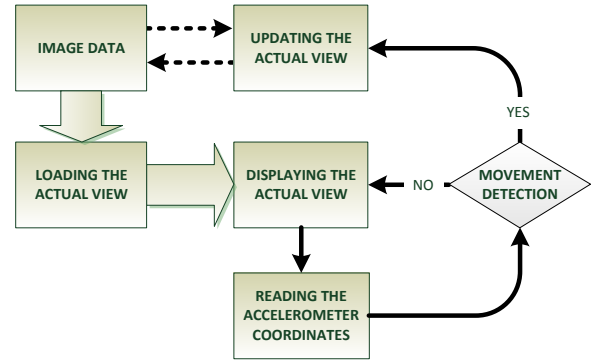

Fig. 3. Content restitution for the motion parallax based display.

corresponds to another view. Therefore, the display duration of a view depends solely on the speed with which the viewer moves the display.

3. QUALITY EVALUATION

3.1. Image data preparation

A subset of the multi-view sequences from the MPEG[1] 3D Video (3DV) data-set [11] has been used for the experiments. More specifically 4 sequences (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) [10] and the view synthesis reference software (VSRS 3.5) [10], 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 multi-view image sets were further converted to high quality JPEG images. Finally the spatial resolution of the multiscopic image set was adapted to match that of the portable device. For our experiments an Android based smart phone LG Optimus 3D with an autostereoscopic display and with screen resolution of 800x480 pixels was used.

Given the resulting multi-view data-set, 7 subsets are created that simulate the different restitution methods (2D, autostereoscopic, anaglyph, motion parallax) for narrow $10cm$ and wide $20cm$ camera baseline leading to the following test conditions. *2D* - reference 2D image, chosen to be the center view of the multi-view data-set. *Sn* - stereoscopic image set[2] with narrow camera baseline resulting in a smaller depth

---

[1]http://mpeg.chiariglione.org
[2]Stereoscopic image set consists of left and right image. The resulting interlaced image is rendered automatically using the LG Real3D technology [**?**]
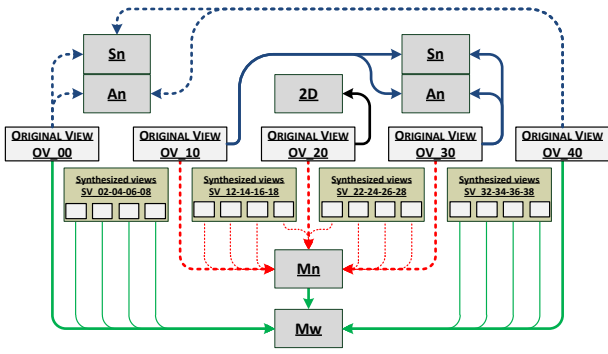
Fig. 4. Creation of the individual test conditions from multiscopic data-set.

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

range. *Sw* - stereoscopic image set with a wide camera baseline resulting in a larger depth range. *Mn* - multiscopic image set of 11 images with a narrow camera baseline resulting in smaller motion parallax. *Mw* - multiscopic image of 21 images with a wide camera baseline resulting in larger motion parallax. *An* - anaglyph image with narrow camera baseline resulting in a smaller depth range. *Aw* - anaglyph image with wide camera baseline resulting in a larger depth range.

The idea of image data-set creation process is illustrated in Fig. 4 for a multi-view image set which consists of 5 original views (OV) and 4 intermediate synthesized views (SV) 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 (Sn, Mn and An) 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 (Sw, Mw and Aw). 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.

3.2. Test methodology

Since judging the quality of different 2D and 3D restitution techniques individually may be quite difficult, the stimulus comparison (SC) quality evaluation method seems to be the most suitable. The subjective tests were performed as follows. An special Android application has been developed in order to display sequentially a pair test stimuli on a testing device. 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
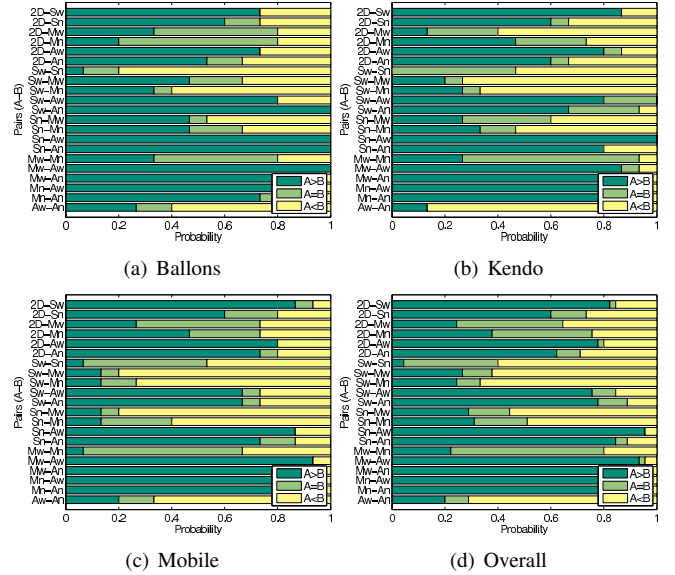
the 3 test video sequences (Balloons, Kendo, Mobile) all the possible combinations of the 7 test conditions (2D, Sw, Sn, Mw, Mn, Aw, An) were considered. This led to a test-set with $3 \times \binom{7}{2} = 63$ paired comparisons. Since the IPD (Inter-Pupillary Distance) of all participating subjects has not been measured, the two camera baselines, used to generate narrow and wide subset, are compared against each other.

Fifteen subjects (11 male and 4 female) with an average age of 29.3 participated in the subjective test experiments. They reported normal or corrected to normal vision according to [12]. All of them were nonexpert viewers with a marginal experience of 3D image and video viewing.

3.3. Results and discussion

Statistical tools have been applied in order to analyze the individual users preferences ratings 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 video sequences. Fig. 5 and Fig. 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 quite similar. Multi-view (Mn, Mw) is rated similarly as 2D and together in general perceived much better (preference probability between 60% and 90%) when compared to anaglyph (An, Aw). Multi-view and 2D comparison with auto-stereoscopic (Sn, Sw) brings quite competitive results with slight advantage on side of multi-view
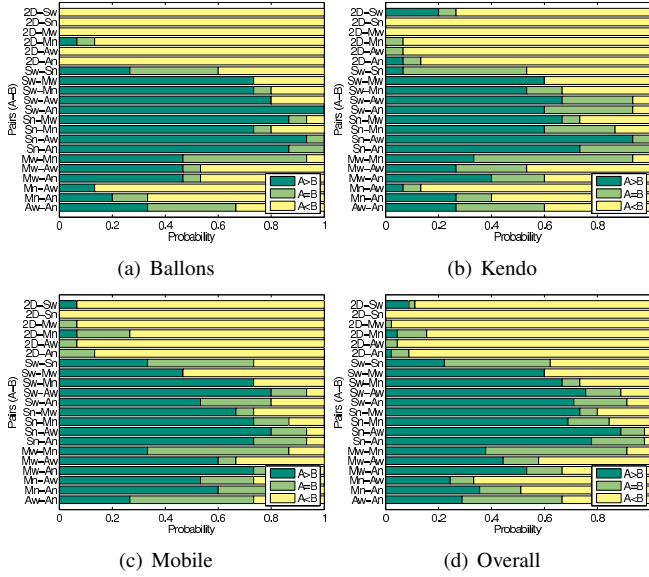
(a) Ballons (b) Kendo

(c) Mobile (d) Overall

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

(a) Overall quality (b) Perceived depth

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

(a) Overall quality (b) Perceived depth

Fig. 8. Preference probability of choosing the given restitution technique among the others.

(preference probability between 60% and 70%). While camera baseline doesn't play role in the peer comparison of both multiscopic sets (Mn, Mw), it does have significant impact on preference probability for auto-stereoscopic image sets (Sn, Sw) and anaglyph images (An, Aw).

The results for the perceived depth are again quite similar across the different scenes. As expected the perceived depth with the 3D restitution methods (Sn, Sw, Mn, Mw, An, Aw) is much better (between 90% and 100 %) when compared to the 2D restitution. The perceived depths of anaglyph and wide multi-view (Mw) seem to be quite comparable with slight shifts (0.3-0.7) of preference probability depending on the video sequence. In general the narrow anaglyph (An) and the wide multi-view (Mw) achieve similar depth perception and narrow auto-stereoscopic is mostly preferred among the other methods.

A preference matrix from the individual paired comparisons by discarding the ties was constructed in order to better analyse the individual test conditions and their performance with respect to each other. It provides the preference probabilities of a test condition A versus another test condition B along the rows. Fig. 7 shows the preference probability matrices averaged over all the video sequences for the overall quality and the perceived depth, respectively.

Analysis of the overall quality matrix shows that multi-view and 2D is clearly preferred (preference probabilities between 70 % and 90 %) over both auto-stereoscopic and anaglyph with slight dependency on the camera baseline for auto-stereoscopic method. With respect to the perceived depth, auto-stereoscopic achieve the best depth quality (preference probabilities between 60 % and 95 %) followed by the
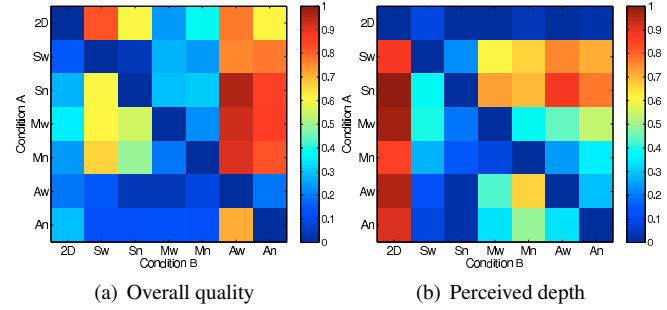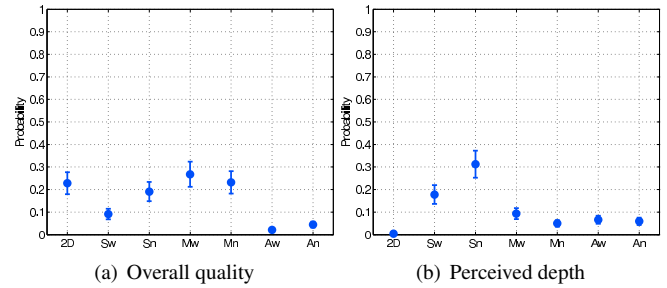
wide multi-view (Mw) and anaglyph method. Obviously, 2D clearly looses against all 3D restitution methods.

Based on the preference probability matrix one can obtain continuous quality scores equivalent to MOS for further quality examination and comparison. This can be done by applying the Bradley-Terry-Luce (BTL) model [13]. In this model the preference probability P_{ij} of choosing i over j can be represented as

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

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} [13]. 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. Fig. 8 shows the preference probability of a given method among the others for the overall and the depth quality. The final, most common representation of quality scores equivalent to MOS is obtained by normalizing π_i into the range $[0, 100]$. Fig. 9 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 preference probability and quality scores for the different display techniques shows that wide multi-view slightly outperforms narrow multi-view, 2D and narrow auto-stereoscopic method. The rest of restitution
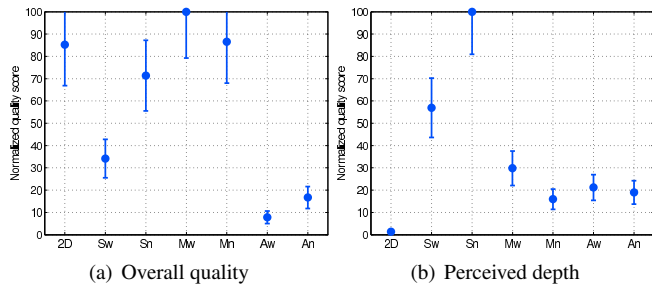
(a) Overall quality (b) Perceived depth

Fig. 9. MOS/CI of the individual test conditions.

methods are preferred with lower quality score in the following order Sw, An and Aw. In other words, the wide multi-view is rated as the best restitution methods in 30% of cases, Mn and 2D in 25%, Sn in 19%, etc. For the depth preference probability and quality scores, the situation is slightly different. The scores for all the 3D restitution techniques are much better when compared to the 2D restitution. Auto-stereoscopic (Sn, Sw) methods achieve the highest score when Sn with MOS 100 is rated approximately twice better than Sw with MOS 55 and three times better that Mw with MOS 30.

4. CONCLUSION

Alternative 3D restitution techniques for displays of the commercially available portable devices have been studied within this work. Conventional 2D together with different 3D restitution techniques including anaglyph, multi-view motion parallax and auto-stereoscopic method have been compared one to another in terms of overall quality and depth perception.

The subjective quality tests show that the additional depth cues provide a better depth perception when compared to simple 2D restitution. While motion parallax method is preferred in terms of overall quality, the depth quality of 3D is clearly on a side of auto-stereoscopic method. Well-known anaglyph method is loosing in both, overall and depth quality, most likely due to the inaccurate color rendering and the crosstalk of anaglyph stereoscopy.

As future work, the significant influence of camera baseline, content and size of multiscopic image set on the overall quality and perceived depth for all restitution methods will be studied in more details.

5. REFERENCES

[1] A. Gotchev, S. Jumisko-Pyykkö, A. Boev, and D. Strohmeier, "Mobile 3dtv system: Quality and user perspective," *Procs of EUMOB*, 2008.

[2] Cha Zhang, D. Florencio, and Zhengyou Zhang, "Improving immersive experiences in telecommunication with motion parallax [applications corner]," *Signal Pro-*

cessing Magazine, IEEE, vol. 28, no. 1, pp. 139 –144, jan. 2011.

[3] M. Rerabek, L. Goldmann, J.S. Lee, and T. Ebrahimi, "Motion parallax based restitution of 3d images on legacy consumer mobile devices," in *Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*. IEEE, 2011, pp. 1–5.

[4] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow, "A survey of 3dtv displays: Techniques and technologies," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1647 –1658, nov. 2007.

[5] LG Electronics Inc., "Introduction to real3d," Tech. Rep., LG Electronics Inc., April 2011.

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

[7] David F. McAllister, Ya Zhou, and Sophia Sullivan, "Methods for computing color anaglyphs," in *Stereoscopic Displays and Applications XXI*.

[8] G. Peters, "Theories of three-dimensional object perception - a survey," Recent Research Developments in Pattern Recognition, 2000.

[9] Christoph Fehn, "Depth image based rendering (dibr), compression and transmission for a new approach on 3d-tv," 2004.

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

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

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

[13] M.E. 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.