ACTIVE CROSSTALK REDUCTION SYSTEM FOR MULTIVIEW AUTOSTEREOSCOPIC DISPLAYS

Philippe Hanhart, Carmelo di Nolfo, and Touradj Ebrahimi

Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

ABSTRACT

Multiview autostereoscopic displays are considered as the future of 3DTV. However, these displays suffer from a high level of crosstalk, which negatively impacts quality of experience (QoE). In this paper, we propose a system to improve 3D QoE on multiview autostereoscopic displays. First, the display is characterized in terms of luminance distribution. Then, the luminance profiles are modeled using a limited set of parameters. A Kinect sensor is used to determine the viewer position in front of the display. Finally, the proposed system performs an intelligent on the fly allocation of the output views to minimize the perceived crosstalk. The user preference between 2D and 3D modes and the proposed system is evaluated. Results show that picture quality is significantly improved when compared to the standard 3D mode, for a similar depth perception and visual comfort.

***Index Terms*—** 3D, multiview autostereoscopic display, crosstalk, viewer tracking, quality of experience

1. INTRODUCTION

Current stereoscopic technologies still require the user to wear bulky glasses. This factor has a significant impact on quality of experience (QoE), especially for users who already wear glasses. Multiview autostereoscopic displays can be the solution to this problem. They provide glasses-free 3D to several viewers simultaneously. Even though this technology is not yet mature enough for a wide acceptance in the consumer market, it is promising. However, multiview autostereoscopic displays suffer from a high crosstalk level between the different views, which is one of the main perceptual factors contributing to image quality and visual comfort [1].

To improve the QoE provided by multiview autostereoscopic displays, researchers have proposed to exploit viewer tracking. Dodgson [2] has analyzed an ideal 3-view display, where only two views are actually displayed, to better deal with the transition of one eye between two adjacent zones. Boev *et al.* [3] have developed a single-viewer system based on user-tracking. The system performs on-the-fly visual optimization to achieve continuous head parallax, i.e., to avoid the repetition effect between the lobes, mitigate crosstalk, and improve brightness. Kooima *et al.* [4] have proposed three techniques to improve the user experience: perspective tracking, channel tracking, and channel reassignment. Nam *et al.* [5] have proposed another approach to actively reduce crosstalk based on the user position. This technique reduces the crosstalk level form 19.1% to only 2.6% for a multiview display using sub-pixel rendering. Recently, advanced multi-user autostereoscopic displays have been developed within the European Union-funded projects MUTED and HELIUM 3D [6]. These displays utilize multi-user head-tracking to provide a proper 3D image to each viewer, based on the eyes position. However, none of these works provides a full description and subjective evaluation of a complete active crosstalk reduction system for current multiview autostereoscopic display technology.

In this paper, we describe and evaluate a system to improve the QoE provided by current and future multiview autostereoscopic display technologies. In particular, our solution aims to reduce the amount of crosstalk perceived by the viewer. The idea is to determine the viewers position, hence the views they can see, and to adjust the different displayed views in real time such that the quality of experience is maximized for each viewer. We implemented our solution considering a single viewer scenario for a 52-inch full HD 28-view Dimenco BDL5231V autostereoscopic display with slanted lenticular sheet. First, the multiview autostereoscopic display was characterized by taking several measurements of the luminance profile of the different views using a DSLR camera. Then, the luminance profiles were modeled using a limited set of parameters. A Kinect sensor was used to determine the viewer position in front of the display using face tracking. Based on this information and the luminance profiles obtained from the display characterization, the views perceived by each eye were determined. Finally, the proposed system performed an intelligent allocation of the output views to minimize the perceived crosstalk in real time. The user preference between 2D and 3D modes and the proposed system was evaluated in terms of image quality, depth quality, and visual comfort through an informal subjective evaluation conducted with five expert viewers. Results show that picture quality is significantly improved when compared to 3D mode, for a similar depth perception and visual comfort.

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1).

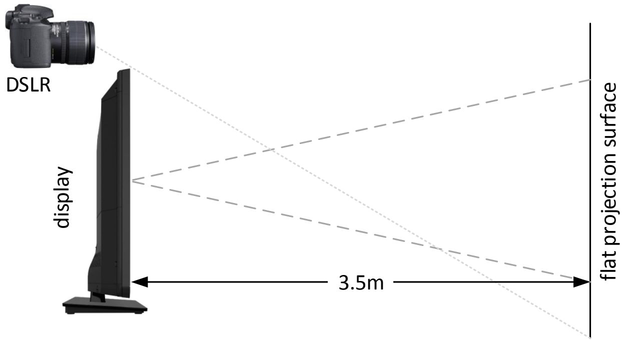

Fig. 1: Schematic of the display characterization setup.

The remainder of the paper is organized as follows. The display characterization and proposed system are described in Sec. 2 and 3. In Sec. 4, the evaluation methodology is described. Results are presented and analyzed in Sec. 5. Finally, concluding remarks and future work are given in Sec. 6.

2. DISPLAY CHARACTERIZATION

The characterization of multiview autostereoscopic displays is usually performed by measuring the luminance emitted by each view at different positions in front of the monitor, which is commonly known as luminance profiles. Significant efforts have been devoted to multiview autostereoscopic display characterization over the recent years. The International Committee for Display Metrology has recently proposed a standardized way to measure crosstalk at a given point in space [7]. However, this approach is time-consuming and expensive, as dedicated measurement devices, e.g., photometers mounted on a rotating stage, high resolution conoscopic cameras, luminance meters, and Fourier Optics [8] are often required. Consequently, we adopted a simpler yet effective approach, which was already used to characterize mobile autostereoscopic display [9]. The main idea is to display a specific test pattern and acquire an estimation of the luminance profiles at a given distance using a DSLR camera. In this paper, a 52-inch full HD 28-view Dimenco BDL5231V autostereoscopic display with slanted lenticular sheet was used.

2.1. Setup

The luminance was measured on a vertical flat projection surface, which was parallel to the display and placed at a fixed distance of 3.5m from the display. This distance is chosen to be the optimal viewing distance of the display. The measurements were performed in a dark room environment. Since the camera cannot be placed at the center of the display without interfering with the measurements on the projection surface, the camera was placed on top of the monitor and controlled remotely. Figure 1 illustrates the setup. We ensured that the camera was parallel to the 3D display and to the projection screen to minimize any distortion. All camera parameters were kept constant during the experiments. The test patterns displayed on the monitor were generated by setting one particular view to white and all other views to black. This process was repeated for each view to measure the luminance profile of the corresponding view. Figure 2 depicts the resulting luminance at 3.5m from the display. As it can be observed, the luminance distribution consists of five slanted cones, due to the use of a slanted lenticular sheet. The luminance distribution is similar for all views, up to a horizontal shift.

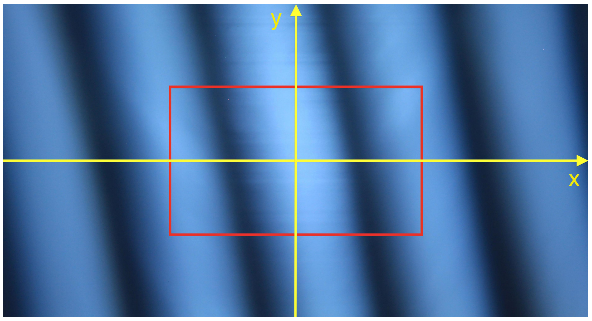

Fig. 2: Resulting luminance at 3.5 m from the display, captured by the camera placed on top of the monitor, when one view is set to white and all other views are set to black. The red box represents the display area.

2.2. Luminance Extraction

For each view, four images were captured and averaged to reduce noise. The averaged images were further cropped to the region of interest (the area of the projection surface). For computation ease, only 30% of the initial size of the picture was kept and a median filter of size 10×10 pixels was applied to further reduce artifacts due to noise. The luminance information was then extracted by converting the gamma encoded sRGB to linear XYZ values and by keeping only the Y channel. Note that the luminance values are defined up to a scale factor, as no reference luminance value was measured.

2.3. Luminance Profile Fitting

Figure 3 depicts the variation of all luminance profiles along the horizontal axis, i.e., the x-axis, at the center of the display ($y = 0$). This corresponds to a cut along the x-axis on Fig. 2, repeated on the luminance distribution generated by each view. As it can be observed, the global intensity is maximum at the center of the display and decreases as the distance from the center of the screen increases. Within the boundaries limited by the display frame (indicated by two red lines on Fig. 3), the global intensity seems to have a Gaussian envelope. For each view, the envelope seems modulated by a squared cosine (since the luminance values are always positive), with five maxima corresponding to the five cones. The

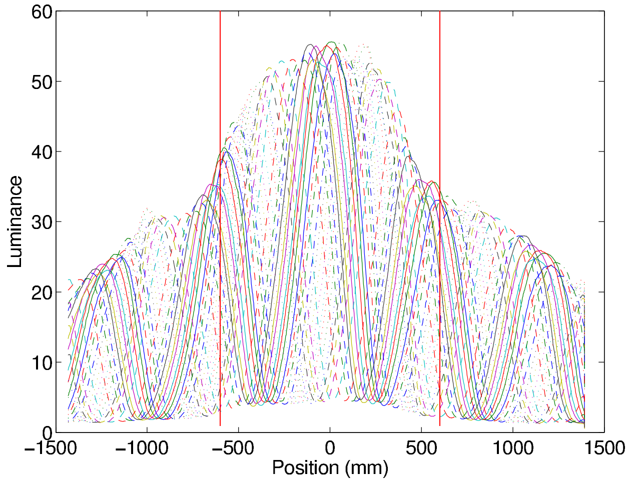

Fig. 3: Variation of all luminance profiles along the horizontal axis at the center of the display ($y = 0$).

luminance profiles of the different views are similar up to a translation, which corresponds to a phase factor in the cosine modulation. A cut along the vertical axis, i.e., the *y*-axis, also reveals a Gaussian shape (not represented here because of the limited space). In this paper, we limited the study of the luminance profiles to an area corresponding to the display area.

Based on the above analysis, the luminance profile, $L(x, y)$, was modeled as a 2D Gaussian envelope modulated by a squared cosine function:

$$\begin{aligned} L(x,y) =& A\cos^2(\omega x + \tau y + \varphi) \\ & \cdot e^{-\left[a(x-x_c)^2 + 2b(x-x_c)(y-y_c) + c(y-y_c)^2\right]} + o \end{aligned} \quad (1)$$

with

$$\begin{aligned} a &= \frac{\cos^2\phi}{2\sigma_x^2} + \frac{\sin^2\phi}{2\sigma_y^2} \\ b &= -\frac{\cos 2\phi}{4\sigma_x^2} + \frac{\sin 2\phi}{4\sigma_y^2} \\ c &= \frac{\sin^2\phi}{2\sigma_x^2} + \frac{\cos^2\phi}{2\sigma_y^2} \end{aligned} \quad (2)$$

where A and o are the amplitude and offset of the 2D Gaussian, respectively, ω represents the frequency of the cosine modulation, τ is phase factor to represent the slanted nature of the luminance distribution, φ is the phase factor representing the translation between the different views, (x_c, y_c) is the center of the 2D Gaussian, σ_x and σ_y represent the horizontal and vertical standard deviations of the 2D Gaussian, respectively, and ϕ is a tilt factor of the 2D Gaussian added to improve the fitting.

2.4. Parameters Reduction

Each luminance profile of the 28 views was fitted independently using Eq. (1), yielding to a total of $28 \times 10 = 280$ parameters. All parameters exhibited small variations, except for φ, which evolved linearly with the view number (up to a period π). These results are in line with the observations reported in Sec. 2.3. Based on these observations, the parameter set was further reduced by computing the average value of the different parameters, except for φ. For the parameter φ, a linear regression was performed

$$\varphi = \alpha v + \beta \quad (3)$$

where v is the view number and α and β are the parameters of the linear regression. The RMSE and coefficient of determination averaged over the 28 views increased from 1.9666 to 2.3113 and decreased from 0.9848 to 0.9789, respectively, which shows that reducing the set of parameters from 280 to 11 parameters had little impact on the error between the measured and fitted values.

3. SYSTEM DESCRIPTION

3.1. User Tracking

The Microsoft Kinect and Face tracking SDKs were used to track the face and face features. In particular, the features corresponding to the left and right corners of each eye were used. The center of the eye was computed as the mid-point between the left and right corners, as this feature is not directly provided by the Face tracking SDK. The face tracking application developed is highly reliable and robust, but sensitive to lightning conditions. The face tracking was performed in real time, with a frame rate varying between 25 and 30 fps, depending on lighting conditions.

3.2. Intelligent View Assignment

Typically, an N-view autostereoscopic system takes $M \ll N$ views as input, due to limitations imposed when using physical cameras. From the limited input views, the missing $N-M$ views are synthesized, for examples by using depth image-based rendering (DIBR). In the most common approach, each view corresponds to a slightly different viewpoint. The reasons behind this approach are multiple: providing a motion parallax effect when the observer moves his/her head in front of the display, coping with different viewing distances, coping with different interpupillary distances, providing 3D effect for different viewers located at different positions, etc. However, this approach might not be optimal in some cases, for example when only one subject is watching the display and standing still, and introduces crosstalk, as the profiles of the different views overlap quite significantly (see Fig. 3).

To reduce perceived crosstalk, our idea consists in performing an intelligent assignment of the different views based on the luminance profiles and the observer's position. Let us assume that a single user is positioned such that his/her left and right eyes see only views 3 and 7, respectively. In this case, the optimal solution would be to assign to views 3 and 7 the content intended for the left and right eyes, respectively. Unfortunately, in a practical scenario, the separation is not

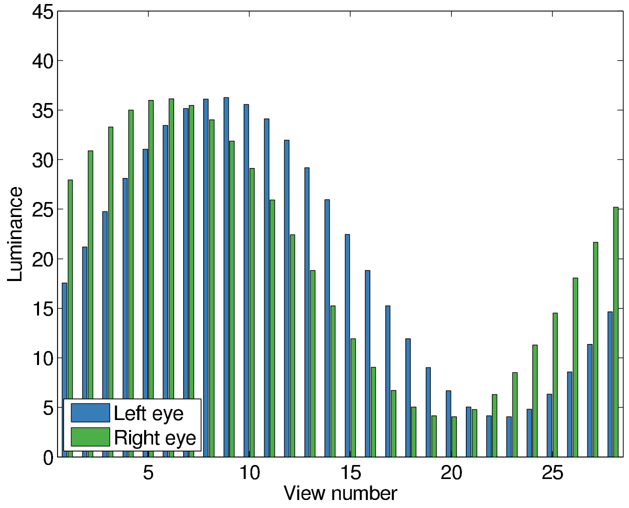

Fig. 4: Luminance perceived by each eye at a given position.

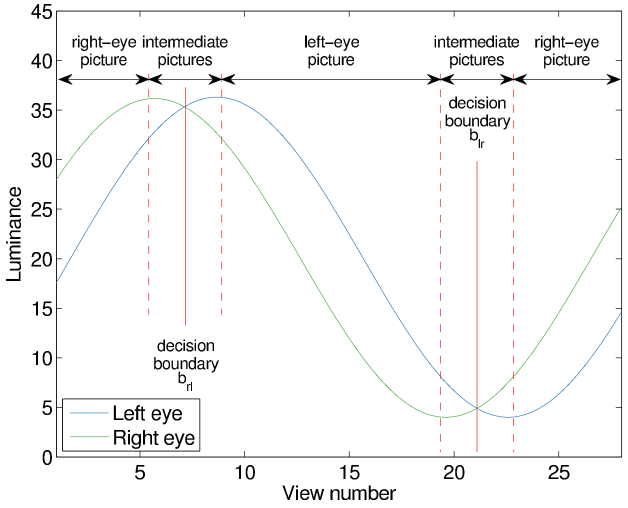

Fig. 5: View assignment for a given position.

that clear and each view is perceived by both eyes, at a different level. However, each view is typically perceived more by one eye than by the other. Therefore, the content that should be assigned to each view can be determined by the eye that sees the most this specific view.

From the eyes position determined by the user tracking and luminance profiles, it is possible to determine for each view the luminance perceived by each eye. Figure 4 illustrates the luminance perceived by each eye at a given position, as a function of the view number. These values are obtained by evaluating Eq. (1) at the eyes positions for each view independently. From this information, the eye for which the luminance is maximum would determine the content assigned to each view. In this case, a direct comparison of the luminance perceived by each eye would be performed for each view. However, Fig. 4 can be seen as the sampled version at fixed integer positions, corresponding to the view numbers, of a continuous function, as if the view numbering was continuous instead of discrete. Since the luminance profiles were fitted with a limited set of parameters where only φ was depending on the view number, Eq. (1) can be evaluated at non-integer view numbers. Figure 5 illustrates the luminance perceived by each eye, as a continuous function of the view number. Views for which the luminance curve corresponding to the right eye lies above the luminance curve corresponding to the left eye should display the right eye picture, and vice-versa. The decision boundaries can easily be determined by computing the two points at which the curves intersect.

Assigning only two different images, i.e., the left and right eye pictures, following the methodology described here above did not look very pleasant on the display for two reasons. First, the luminance profiles have a significant overlap: the view that maximizes the luminance perceived by the right eye leaks quite significantly into the left eye, and vice-versa (see Fig. 4). Second, the edges at the objects' boundaries corresponding to sharp depth transitions did not look very pleasant because of the sub-pixel interlacing. This effect does not appear in standard 3D mode, because the multiple views contain somewhat similar information, which tends to smooth out the depth transitions and blur objects' boundaries. To overcome these issues, three intermediate pictures, corresponding to equidistant viewpoints, located in between the left and right eye pictures, were used for the views near the decision boundary (see Fig. 5). For example, near the right eye picture to left eye picture decision boundary (b_{rl}), the center-right (p_{cr}), center (p_c), and center-left (p_{cl}) intermediate pictures are assigned as

$$v_n = \begin{cases} p_{cr} & \text{if } n \in [b_{rl} - \delta_e, b_{rl} - \delta_c[\\ p_c & \text{if } n \in [b_{rl} - \delta_c, b_{rl} + \delta_c] \\ p_{cl} & \text{if } n \in]b_{rl} + \delta_c, b_{rl} + \delta_e] \end{cases} \tag{4}$$

where v_n is the *n*-th view. Therefore, 5 pictures were assigned to the 28 views, but with a different spacing for each picture, whereas 28 different pictures are used in the 3D mode with regular spacing, as each view uses a different picture. The number of intermediate pictures and parameters ($\delta_c = 1$ and $\delta_e = 4$) were determined empirically to achieve the best rendering. This solution smooths the image and enhances the visual comfort when compared to using only two pictures.

3.3. Multiview Shuffling

The Dimenco BDL5231V monitor uses an LCD panel composed of 1920 × 1080 pixels. However, the shuffling of the 28 views is done at a sub-pixel level. Dimenco provides a software for shuffling 28 full HD video sequences corresponding to the 28 views into a single full HD video to be displayed on the monitor. This tool was reversed engineered, by using simple input patterns, to determine the sub-pixel arrangement, i.e., to determine which sub-pixel corresponds to which view. This information allows us to perform the multiview shuffling in our application, which is also much faster than the original software provided by Dimenco.

3.4. Final System and Implementation

To reduce the impact of the user tracking imprecision and increase visual comfort, small head movements (less than 2 cm in any direction) were discarded. Additionally, to avoid flickering when a new picture is assigned, fading was performed between two successive renderings. The fading was performed by computing three intermediate images using weighted addition of the old and new pictures to display, whose weights increased gradually in favor of the new picture. For our experiments, the system was implemented in C++ using the OpenCV library and achieved a rendering at about 30 fps.

4. SUBJECTIVE EVALUATION

To evaluate the performance of the proposed system over the 2D and 3D modes of the display, an informal subjective evaluation was performed with five expert viewers.

4.1. Dataset

Four multiview video plus depth (MVD) contents were used in the experiments: *GT Fly*, *Poznan Street*, *Shark*, and *Undo Dancer*. These contents are used by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of VCEQ and MPEG [10]. *Poznan Street* is a real scene with estimated depth maps, whereas the three remaining contents are computer-generated scenes with ground truth depth maps. One key frame, which maximizes the amount of depth, was selected for each content.

4.2. Test Methodology

The paired comparison methodology [11] was chosen as judging the quality of different 2D and 3D rendering systems individually may be quite difficult. Pairs of images, A and B, which resulted from different rendering systems, were presented in succession order on the display. Subjects were asked to judge which image in a pair (A or B) is preferred in terms of picture quality, depth quality, and visual comfort [11]. The option same was also included to avoid random preference selections. For each of the 4 test contents, all the possible combinations of the 3 conditions (2D mode, 3D mode, and proposed system) were considered, leading to a total of $4 \times \binom{3}{2} = 12$ paired comparisons.

Viewers were allowed to move freely (within a range defined by the monitor frame) along a line parallel to the display, at the optimal viewing distance of 3.5 m, which corresponded to the measurement distance (see Sec. 2.1).

4.3. Analysis of the Results

First, the winning frequency w_{ij} of stimulus i against stimulus j and tie frequency t_{ij} between the two stimuli were computed from the individual ratings. Note that $t_{ij} = t_{ji}$ and $w_{ij} + w_{ji} + t_{ij} = N$, where N is the number of subjects.

Then, the Bradley-Terry-Luce model [12] was used to convert the winning frequencies to continuous-scale quality scores, which are equivalent to mean opinion scores (MOS). In this model, the empirical probability P_{ij} of choosing stimulus i is defined as

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \tag{5}$$

where π_i satisfying $\pi_i \geq 0$ and $\sum_i \pi_i = 1$ can be considered as the quality score for stimulus i and can be obtained via maximum likelihood estimation. Ties were considered as half way between the two preference options and equally distributed between P_{ij} and P_{ji} [12]. The confidence intervals (CI) for the maximum likelihood estimates of the scores were obtained using the Hessian matrix of the log-likelihood function. The results were normalized to the range $[0, 100]$ for a better representation.

5. RESULTS AND DISCUSSION

Figure 6 left shows the preference and tie probabilities obtained over all test images for picture quality, depth quality, and visual comfort. As it can be observed, the proposed system significantly improves picture quality when compared to the 3D mode, as it has a preference probability of 70%, whereas the 3D mode has a preference probability of only 20%. With the proposed system, less crosstalk was perceptible and there was no unpleasant transition between the different viewing cones. The 2D mode and proposed system were perceived as similar in 60% of the test stimuli, which shows that the proposed system provided a picture quality comparable to that of the 2D mode.

Regarding depth quality, the 3D mode showed a clear advantage over the 2D mode. Results show a slight preference for the 3D mode over the proposed system, with a preference probability of 45%. Nevertheless, the depth quality of the proposed system is still much better than that of the 2D mode, despite the absence of motion parallax depth cues when compared to the 3D mode. In terms of visual comfort, 2D mode is preferred most of the time. The proposed system also improves visual comfort when compared to the 3D mode, as it is preferred in 55% of the test stimuli, whereas the 3D mode is preferred in only 30% of the test stimuli. From the comments of the viewers, this can be explained by the fact that they had some difficulties to predict the behavior of the system as they moved when compared to the 3D mode, where they could find a predictable and fixed sweet-spot.

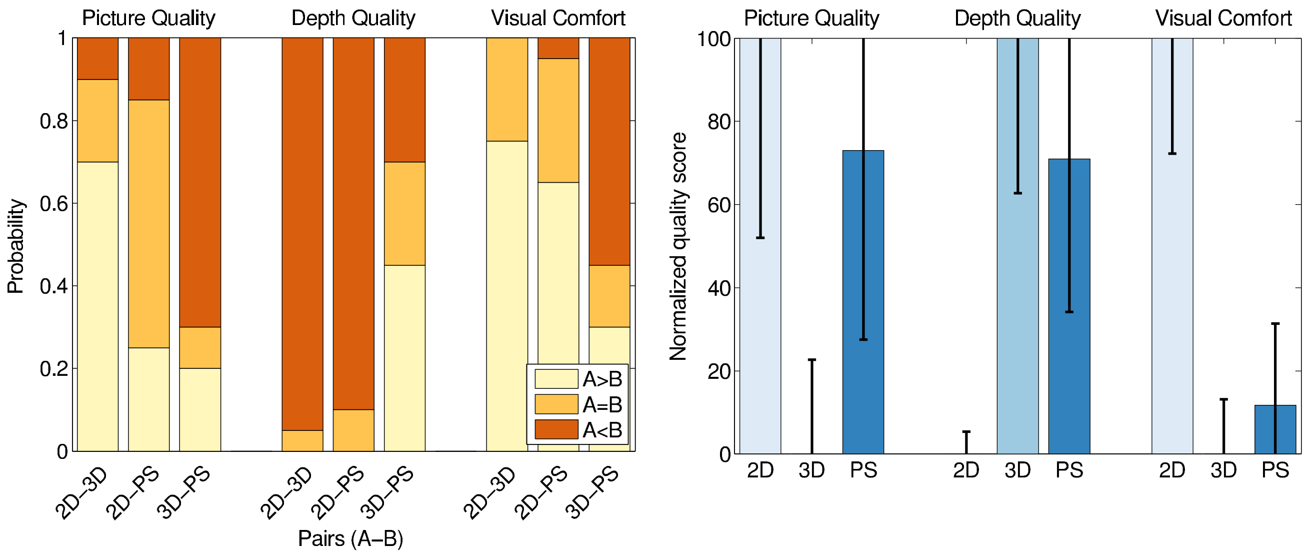

Fig. 6: Preference probabilities (left) and normalized quality scores (right) for 2D mode, 3D mode, and proposed system (PS).

Figure 6 right shows the MOS and CI obtained over all test images for picture quality, depth quality, and visual comfort. As it can be observed, the proposed system significantly enhances picture quality when compared to the 3D mode and provides similar depth perception, as the CIs overlap significantly. However, the improvement in terms of visual comfort is not significant.

6. CONCLUSION

In this paper, we proposed a system to improve 3D QoE on multiview autostereoscopic displays. The proposed system relies on display characterization and viewer tracking to perform an intelligent allocation of the output views on the fly to minimize perceived crosstalk. The system was implemented to improve 3D QoE on a 52-inch full HD 28-view Dimenco BDL5231V autostereoscopic display with slanted lenticular sheet. A Kinect sensor was used to track the viewer and a simple display characterization was performed using a DSLR camera. The user preference between standard 2D and 3D modes and the proposed system was evaluated. Results showed that picture quality is significantly improved when compared to 3D mode, for a similar depth perception and visual comfort.

In a future study, we plan to improve the system and to perform a complete subjective evaluation of the improved version with naïve viewers. Improvements include better assignment of the views, especially near the decision boundary, better fading, and better filtering of the user position. We also plan to extend the measurements and luminance model for different viewing distances to allow the user to move back and forth. The final goal is to develop a system that can handle several viewers, located at different positions.

7. REFERENCES

[1] L. Meesters, W. IJsselsteijn, and P. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[2] N. A. Dodgson, "On the number of viewing zones required for head-tracked autostereoscopic display," in *Stereoscopic Displays and Virtual Reality Systems XIII*, Jan. 2006.

[3] A. Boev, M. Georgiev, A. Gotchev, and K. Egiazarian, "Optimized single-viewer mode of multiview autostereoscopic display," in *16th European Signal Conference*, Aug. 2008.

[4] R. Kooima, A. Prudhomme, J. Schulze, D. Sandin, and T. DeFanti, "A multi-viewer tiled autostereoscopic virtual reality display," in *17th Symposium on Virtual Reality Software and Technology*, Nov. 2010.

[5] D. Nam, J. Park, D. Park, and C. Y. Kim, "Autostereoscopic 3D - How can we move to the next step?" in *10th Euro-American Workshop on Information Optics*, June 2011.

[6] P. Surman, R. Brar, I. Sexton, and K. Hopf, "MUTED and HELIUM3D autostereoscopic displays," in *International Conference on Multimedia and Expo*, July 2010.

[7] International Committee for Display Metrology, "Information Display Measurements Standard," version 1.03a, June 2012.

[8] P. Boher, T. Leroux, T. Bignon, and V. Collomb-Patton, "Optical characterization of different types of 3D displays," in *Advances in Display Technologies II*, Feb. 2012.

[9] A. Chappuis, M. Rerabek, P. Hanhart, and T. Ebrahimi, "Subjective evaluation of an active crosstalk reduction system for mobile autostereoscopic displays," in *Stereoscopic Displays and Applications XXV*, Feb. 2014.

[10] ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, "Common Test Conditions of 3DV Core Experiments," Doc. JCT3V-D1100, Incheon, Korea, Apr. 2013.

[11] ITU-R BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," Aug. 2012.

[12] 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.