

Subjective evaluation of an active crosstalk reduction system for mobile autostereoscopic displays

Alexandre Chappuis, Martin Rerabek, Philippe Hanhart, and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG), Institute of Electrical Engineering (IEE),
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT

The Quality of Experience (QoE) provided by autostereoscopic 3D displays strongly depends on the user position. For an optimal image quality, the observer should be located at one of the relevant positions, called sweet spots, where artifacts reducing the QoE, such as crosstalk, are minimum. In this paper, we propose and evaluate a complete active crosstalk reduction system running on an HTC EVO 3D smartphone. To determine the crosstalk level at each position, a full display characterization was performed. Based on the user position and crosstalk profile, the system first helps the user to find the sweet spot using visual feedback. If the user moves away from the sweet spot, then the active crosstalk compensation is performed and reverse stereo phenomenon is corrected. The user preference between standard 2D and 3D modes, and the proposed system was evaluated through a subjective quality assessment. Results show that in terms of depth perception, the proposed system clearly outperforms the 3D and 2D modes. In terms of image quality, 2D mode was found to be best, but the proposed system outperforms 3D mode.

Keywords: Autostereoscopic display, 3D, Subjective quality assessment, Crosstalk compensation, Face detection, Display characterization, Android

1. INTRODUCTION

In the past few years, the entertainment industry expanded into mobile technologies. Constant improvement of the processing power and rendering technology enabled the consumption of 3D media on mobile devices via autostereoscopic displays using parallax barrier. Parallax barrier, as a less intrusive technology (glassesless), exploits a special LCD layer as a spatial filter directing two different views towards the eyes of the viewer. There are certain limitations related to parallax barrier technology based systems. The viewing areas are limited by specific viewing zones, and as soon as the viewer moves his head outside the sweet spot, a significant visual impairment is perceived. QoE of such devices is therefore limited as well as the viewing freedom of the user. However, there were no particular attempts to improve the QoE for such devices and no commercially available hardware exists with embedded algorithms improving the visual quality.

When a user watching an autostereoscopic display moves away from the sweet spot along a horizontal axis, the right image progressively leaks into the left eye and vice versa, which is commonly referred to as a crosstalk. Moving further along the horizontal axis reaches a point where the left (right) eye only perceives the right (left) image. This is known as reverse stereo or pseudoscopy. These effects decrease depth and image quality and increase visual discomfort, especially when reverse stereo occurs. To improve the QoE provided by autostereoscopic displays, researchers have proposed to perform active crosstalk reduction based on user position.¹⁻³ The user is tracked using face and eye detection and its position related to sweet spot is evaluated. Outside the sweet spots, crosstalk is canceled by simply switching back to 2D mode² or reduced by performing an intelligent assignment of the pixel values based on the visibility profiles of different views.³ In case of reverse stereo, the left and right images are swapped.³

Further author information: (Send correspondence to Martin Rerabek)

Alexandre Chappuis: E-mail: alexandrechappuis@bluewin.ch

Martin Rerabek: E-mail: martin.rerabek@epfl.ch

Philippe Hanhart: E-mail: philippe.hanhart@epfl.ch

Touradj Ebrahimi: E-mail: touradj.ebrahimi@epfl.ch

Combining face detection and eye tracking with dynamic 3D rendering to improve QoE attracts the attention of researchers in autostereoscopic community. In,^{4,5} authors used the eye tracking systems coupled with autostereoscopic two-view displays. They simulated motion parallax and synthesized the left and right views in real time according to the user position. Other displays,^{6,7} such as the famous Free2C display from Fraunhofer HHI, controls the parallax barrier or the lenticular sheet such that the sweet spot follows the user head. Several works also proposed to combine face/eye tracking with multiview autostereoscopic display. In,⁸ an ideal 3-view display is analyzed, where only two views are actually displayed, to better deal with the transition of one eye between two adjacent zones. Boev et al.¹ presented 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. There are also some autostereoscopic displays* that use head tracking to get better 3D experience. Another autostereoscopic displays[†] use the projected infrared light to improve head tracking. The same approach is used in Kinect device. In all above mentioned cases, a powerful hardware including a pair of cameras used for better tracking is needed.

Similar approach has been rarely addressed for mobile devices due to their limitation in processing power and/or real-time face/eye tracking capabilities. Boev et al.² relied on the Open Multimedia Applications Platform (OMAP) embedded in a smartphone and using one camera for eye tracking. Correction of the pseudoscopy was performed, whereas the crosstalk was not reduced and when the viewer experienced it, the screen was switched to 2D mode. Moreover, the good viewing zones were measured empirically and the viewer was supposed to be at one fixed optimal distance from the screen. In another words, fixed interpupillary distance was assumed. Another special hardware prototype of mobile multiview autostereoscopic display based on a method providing clear images appropriate to the positions of the observer's eyes was proposed by Park et al.³ However, no subjective assessment demonstrating the effectiveness of an active crosstalk reduction system on a mobile device has been yet reported. In this paper, we propose and evaluate a complete active crosstalk reduction system running on an HTC EVO 3D smartphone.

The rest of the paper is organized as follows. In Section 2, the description of the design of the crosstalk reduction system including the real-time face detection and eye tracking, the autostereoscopic display characterization and sweet spot location, camera calibration and user position location, as well as the reduction of the artefacts, is presented. Also, the rendering method exploiting eye information and device characterization is described here. In Section 3, details about subjective assessment experiment are presented and its results are discussed. Finally, concluding remarks are given in Section 4.

2. DESIGN AND DESCRIPTION OF AN ACTIVE CROSSTALK REDUCTION SYSTEM FOR AUTOSTEREOSCOPIC MOBILE DEVICES

In this paper, we propose an implementation in form of a mobile application for parallax barrier based autostereoscopic devices, improving 3D user experience. The core principle of this application is based on an active crosstalk reduction dependent on user position. The actual user position is tracked using face and eye detection and an efficiently designed pixel-wise image operation is performed based on the luminance profiles of each view. To determine the concrete luminance profiles, characterization of the mobile display was performed. Final application subsequently exploits both, the information of eye position and the display characterization, and applies the rendering methods enhancing the QoE. Moreover, the following requirements were taken into account for the final application development:

- Real time face detection and eye tracking.
- Real time crosstalk reduction and pseudoscopy correction.
- Intuitive, user-friendly feedback for the user.
- User-oriented, allowing the parameters change according to user characteristics.

*<http://www.advancedtechnologykorea.com/6340/>

†<http://3dvision-blog.com/6505-review-of-the-toshiba-qosmio-f750-autostereoscopic-3d-laptop/>

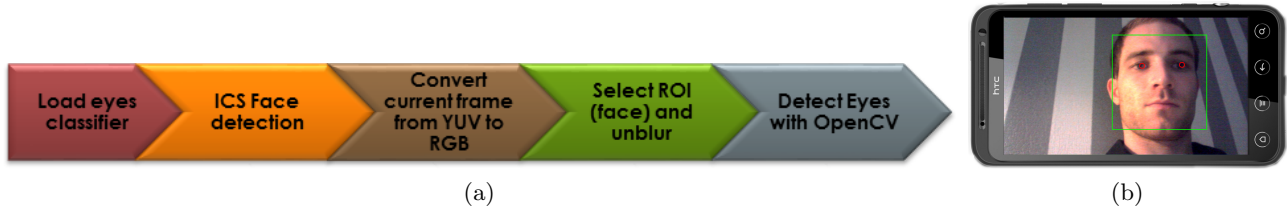


Figure 1: (a) General diagram of eye detection., (b) Preview of face and eye detection.

- Low battery/CPU consumption.

Because 3D imaging is quite a recent topic for mobile displays, there are only two 3D-capable Android phones available, the LG Optimus 3D and the HTC Evo 3D. Both use a 3D display with parallax-barrier that can be switched off to get 2D mode, but differ in terms of CPU, memory, software version and 3D API (*Real3D*[‡] for the Optimus 3D and *S3D*[§] for the Evo 3D). For our development, the HTC Evo 3D with the latest available version of Android[¶] system was used assuming the landscape displaying mode. The main parameters related to display and front camera are:

- screen size of the smart phone = 960×540 pixels.
- front camera preview size = 960×544 pixels.
- front camera picture size = 1280×720 pixels.

In this section, technical aspects of above mentioned individual parts, as well as of the entire mobile application, are described.

2.1 Face detection and eye tracking

There are several different face detection algorithms running on Android devices: the Android built-in face detector algorithm^{*}, the OpenCV face detection algorithm[†], and the CamShift algorithm[‡]. In order to get as real-time face detection feedback as possible, the Android built-in algorithm *FaceDetectionListener* was used together with OpenCV and HAAR cascade classifiers for eye detection[§].

The general diagram of the eye tracking algorithm is summarized in Figure 1(a). First, the cascade of HAAR classifiers for eye detection is stored in an XML file and loaded when application boots. Each frame is processed independently and once a face is detected, the current frame is converted from YUV (Android camera format) to RGB using OpenCV conversion functions. Converted frame is then cropped to the region of interest (the upper half of the face) and the result is sharpened by subtracting a blurry (Gaussian blur is assumed) image from the cropped region. Cropping significantly decreases the computational time. Eye detection is performed on the cropped image using the OpenCV library only if a face is previously found.

The frame rate of the front-facing camera varies a lot according to ambient light. When the environment around the user is too dark, the frame rate of the camera decreases and face detection becomes slow. For good lighting conditions, the real-time face detection achieves a frame rate of $15fps$, which is the maximal frame rate of the front-facing camera, whereas eye detection is performed at maximum of $10fps$. Figure 1(b) shows an example of camera preview with detected face and eyes.

[‡]<http://developer.lge.com/main/Intro.dev>

[§]<http://www.htcdev.com/devcenter>

[¶]Android Ice Cream Sandwich (ICS)

^{*}used as a black box since no detailed information is provided by developers and/or vendors

[†]<http://opencv.org/platforms/android.html>

[‡]http://docs.opencv.org/trunk/doc/py_tutorials/py_video/py_meanshift/py_meanshift.html#facetracker

[§]<http://www-personal.umich.edu/~shameem/>

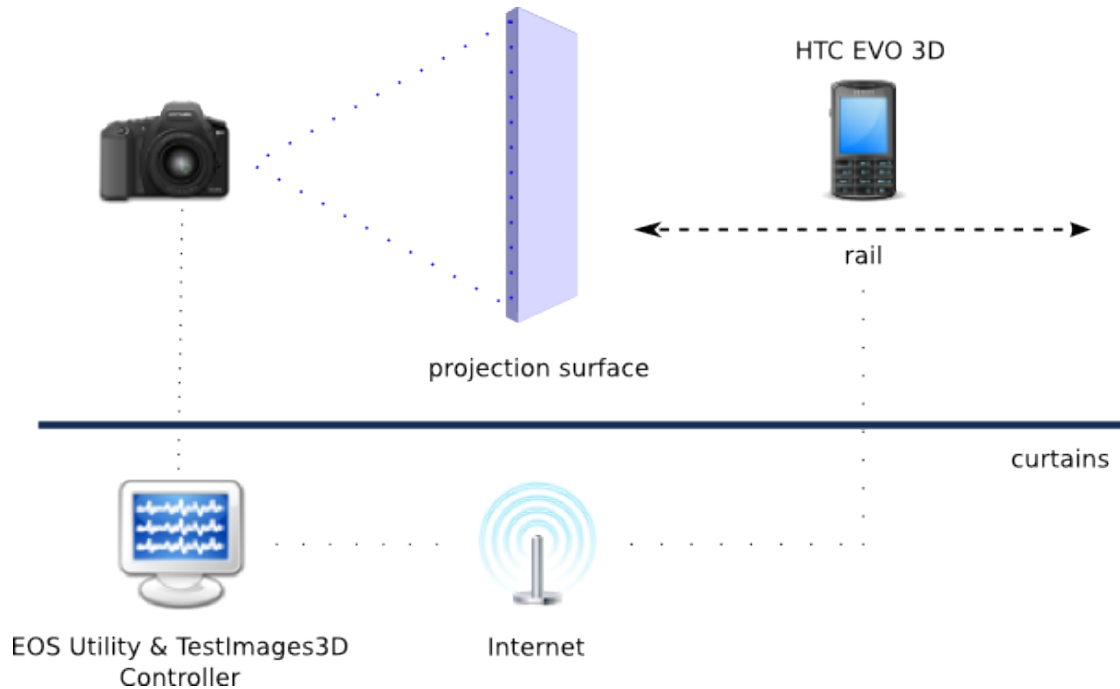


Figure 2: Display characterization setup.

2.2 Characterization of mobile display

The main goal of the display characterization is to define the sweet spots in terms of viewing angle and optimal viewing distance. Knowing the technical details of the autostereoscopic display, such as barrier slit or pixel-barrier separation, allows to exploit simple optical laws in order to characterize the display. In common practice, display technical details are not provided and more complex approach must be used.

The autostereoscopic display characterization is usually performed as a measure of crosstalk at different positions in space providing 3D crosstalk map. Significant effort has been devoted to autostereoscopic display characterization over past few years. The International Committee for Display Metrology has recently proposed a standardized way to measure crosstalk at a given point in space.⁹ This approach is, however, time-consuming and expensive as dedicated hardware, such as luminance meters and Fourier Optics,¹⁰⁻¹² is often required. Another simple yet effective approach, which we adopted for our purposes, was proposed by Sykora et al.¹³ and Hong.¹⁴ The main idea is to project a predefined pattern rendered by the autostereoscopic display and acquire a crosstalk map at a given distance using a DSLR camera. The measurement setup, assuming a semi-transparent projection surface is shown in Figure 2.

The mobile device with the autostereoscopic display was placed on a rail on one side of the projection surface, whereas the DSLR camera was placed on the other. Both devices, i.e., mobile phone and DSLR, were controlled remotely by a computer which assures the setting of shooting parameters as well as the user distance (distance between display and projection surface) and the changes of test pattern. The predefined patterns black-black (KK), white-black (WK)(see Figure 3(a)) and black-white (KW), were displayed and recorded at different user distances and the 3D crosstalk map was then reconstructed. In our experiments, the distance of the mobile phone varied from 220mm to 420mm with a step of 5mm. A dark room environment, totally isolated from any external light source, was needed to assure the high contrast of the projected pattern. A sheet of tracing paper placed between two glass plates providing efficient flatness, transparency, uniformity, and light scattering was used as a projection surface.

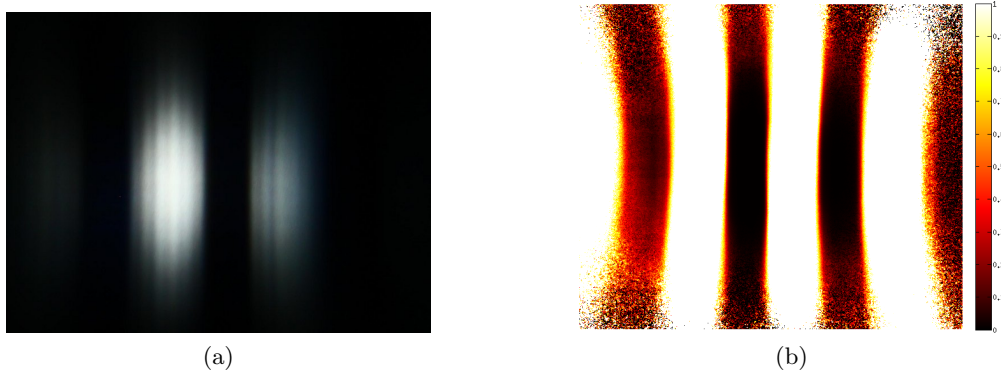


Figure 3: (a) Example of the projected image of the WK test pattern. (b) Crosstalk profile in the middle of the display at distance $310mm$.

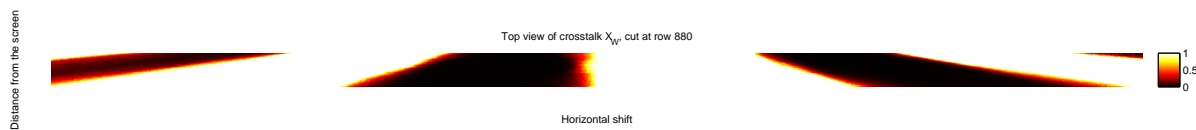


Figure 4: Top view of the crosstalk map for the left eye. Captures of the projection surface were made in the range $[220, 420]mm$ in steps of $5mm$.

2.2.1 Computing the crosstalk

The first step towards crosstalk computation is the preprocessing of the DSLR JPEG images. For each test patterns and distance, five images were captured and averaged in order to suppress the noise. Averaged images were further cropped to the region of interest (the area of the projection surface). The luminance information was then extracted by using sRGB to XYZ conversion matrix as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{linear} \\ G_{linear} \\ B_{linear} \end{bmatrix}. \quad (1)$$

The linear sRGB tristimulus values were extracted by using the following gamma correction formula:

$$C_{linear} = \begin{cases} \frac{C_{srgb}}{12.92} & \text{if } C_{srgb} \leq 0.04045 \\ \left(\frac{C_{srgb} + \alpha}{1 + \alpha} \right)^{2.4} & \text{if } C_{srgb} > 0.04045, \end{cases} \quad (2)$$

where C_{linear} is one of the three linear sRGB tristimulus values (R_{linear} , G_{linear} , B_{linear}), C_{srgb} is one of the RGB values of input image in the range $[0, 1]$, and $\alpha = 0.055$. The final crosstalk value is computed by using the luminance Y value only. The crosstalk X_L for the left eye is

$$X_L = \frac{\max(KW - KK, 0)}{\max(WK - KK, 0)}, \quad (3)$$

where KK is the luminance of the black-black pattern, KW the luminance of the black-white pattern, and WK the luminance of the white-black pattern. The maximum function is used to avoid negative luminance values. Right eye crosstalk is simply the inverse of X_L .

Figure 3(b) illustrates the profile of computed crosstalk at distance $310mm$. Zero means no crosstalk, whereas 1 means maximum crosstalk. Values higher than 100% have been set to 1. The regions with high crosstalk are clearly separated from regions with no crosstalk. Crosstalk appears as bent stripes whose widths increase with

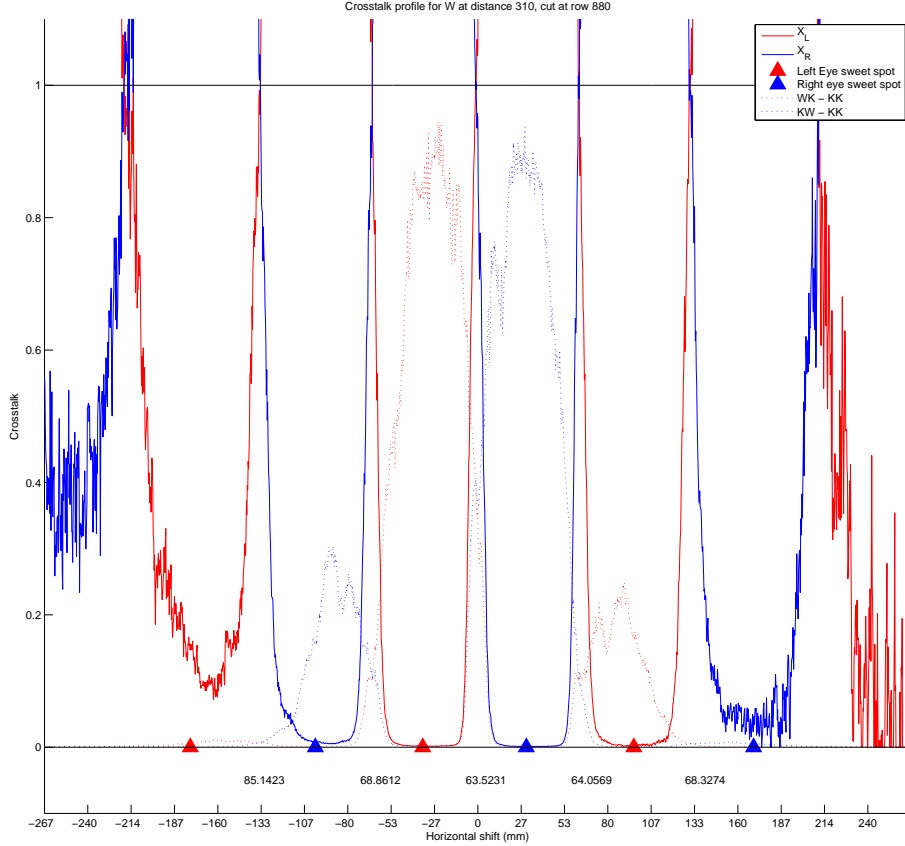


Figure 5: Crosstalk profile in the middle of the screen, at distance 310mm .

user distance as can be seen in Figure 4. Optimal eye locations, defined as sweet spots, are those areas where the crosstalk is minimum. Sweet spots can be found by a thorough analysis of crosstalk profiles at a given user distance for the horizontal center of the display (see Figure 5). Crosstalk values higher than 1 are ignored, and intensities of $WK - KK$ and $KW - KK$ representing the luminance information are shown with a dashed line. So, when one eye sees as much from the left view as from the right view, the crosstalk value is 1. The sweet spot position is determined with respect to a tolerated error ϵ as a middle point between crosstalk values $X_L = X_R = 1 \pm \epsilon$. The left and right eye sweet spots are marked by red and blue triangles, respectively. Having the crosstalk profile for a whole range of distances, the relation between Interpupillary Distance (IPD) and Optimal Viewing Distance (OVD) can be estimated.¹⁵ To achieve the best viewing condition, the distance between the sweet spots for left and right eye should correspond to the IPD. Figure 6(a) illustrates the mutual dependence between IPD and OVD, which can be approximated by a linear regression

$$OVD [mm] = 5.848203 * IPD [mm] - 67.143329. \quad (4)$$

For a standard male IPD (65mm) the OVD corresponds to 313mm . Although the crosstalk stripes are bent (see Figure 3(b)), the variation corresponding to possible user movements along the vertical axis is relatively small. Therefore, an assumption of spatial invariance of crosstalk along vertical axis can be made.

Considering certain crosstalk threshold, assuring good viewing quality, the viewing freedom in horizontal axis for a given user distance can be determined. For the autostereoscopic displays, the viewing freedom is limited by

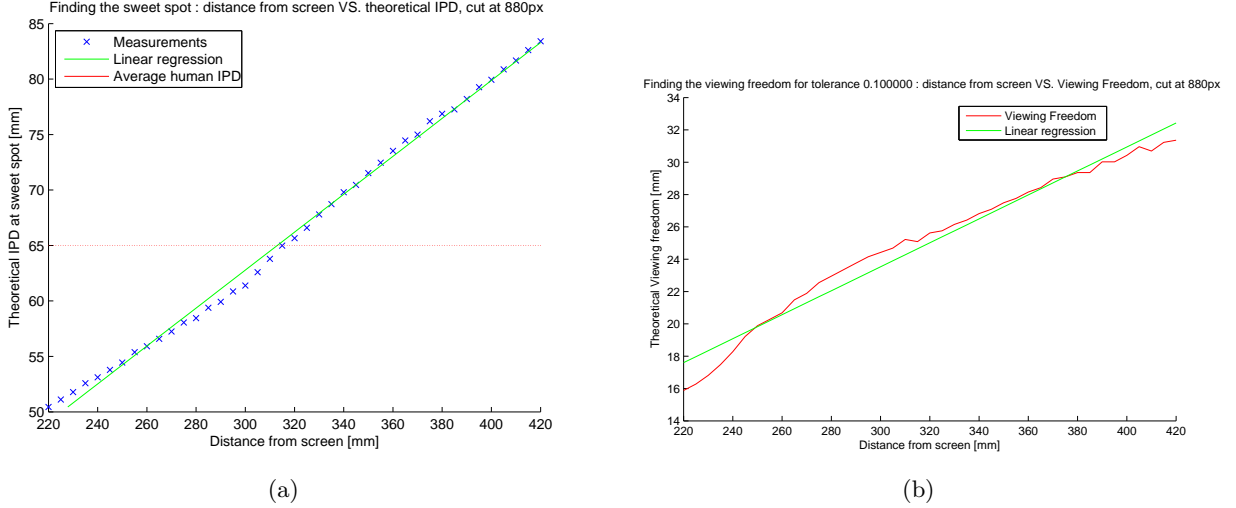


Figure 6: (a) Sweet spot for a given IPD., (b) Viewing freedom vs. user distance.

diamond-shaped regions. However, as the measurements in our experiment were done only within a certain range from 220mm to 420mm , the relationship between user distance and viewing freedom must be linear. Moreover, when the crosstalk tolerance decreases, so does the viewing freedom. An example of graph with 10% crosstalk tolerance is illustrated in Figure 6(b). A linear regression is used for fitting purposes, regression models for different crosstalk tolerances are as follows:

Crosstalk tolerance	→	Linear regression
5%	→	$VF[mm] = 0.068600 * distance[mm] + 1.314246$
10%	→	$VF[mm] = 0.074119 * distance[mm] + 1.295925$
15%	→	$VF[mm] = 0.078044 * distance[mm] + 1.120717$
20%	→	$VF[mm] = 0.080536 * distance[mm] + 1.195488$
25%	→	$VF[mm] = 0.081833 * distance[mm] + 1.447605$

Note that the viewing freedom is assumed to be constant along the vertical axis.

2.3 Camera calibration

In order to find the distance of the user with respect to the autostereoscopic display, and to determine the sweet spot location, calibration of the front camera was performed. An ideal pinhole camera model, providing the approximation of the relationship between a scene and its projection on a sensor, is assumed. The real-world scene is projected on the image plane, which lies at a focal distance f measured along the optical axis. The focal length f is divided in two components, f_x and f_y , because usual CMOS sensors have rectangular dimensions. The point of intersection between the optical axis and the image plane is called the principal point and, in an ideal case, it corresponds to the center of the image plane. However, this is not the case for most camera models, so two additional parameters, c_x and c_y , related to directional shifts must be introduced. These four parameters were estimated using well known algorithms and a calibration pattern.¹⁶ Distortion parameters of the camera were ignored as they don't significantly influence the user position in space.

Once the parameters f_x and f_y are found, one can easily compute the distance of the user with respect to the display. Assuming, that the positions of left (y_1) and right eye (y_2) on the horizontal axis of the image plane are known, then

$$y_1 = f_y \frac{Y_1}{Z} + c_y y_2 = f_y \frac{Y_2}{Z} + c_y, \quad (6)$$

and their difference, representing the known eye distance D in pixels in the image plane is

$$D = y_1 - y_2 = f_y \frac{Y_1 - Y_2}{Z}, \quad (7)$$

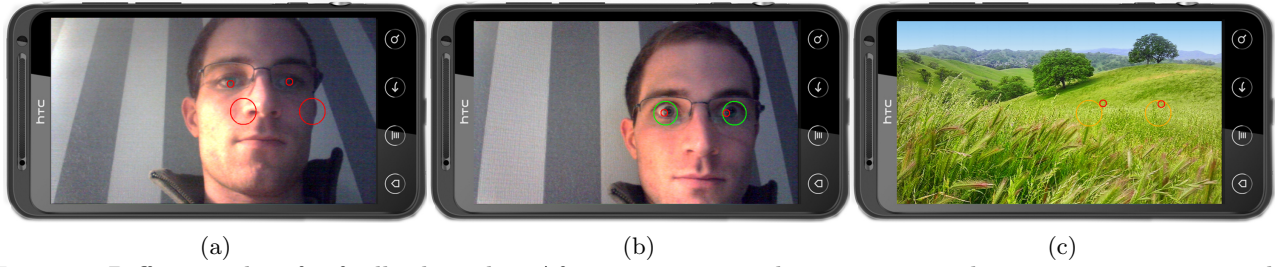


Figure 7: Different colors for feedback circles. After some time in the sweet spots, the camera preview is made invisible and 3D content is shown.

where the difference $Y_1 - Y_2$ represents the IPD of the user in mm. The only unknown variable in (7) is Z , which is the user distance with respect to the image plane.

The front camera is located in the upper left corner of the phone, when used in landscape orientation. Therefore, an offset depending on the user position correcting the face and eye position with respect to the horizontal center of the display must be set. Actual offset value for given user distance can be computed according to following linear regression

$$\text{offset [pixels]} = \max(0, -0.410263 * \text{user distance [mm]} + 318.707571), \quad (8)$$

which was found empirically for seventeen different user distances. Moreover, the ratio between the image size (720×1280) and the screen dimension (540×960) is taken into account, thus the offset value corresponds to an accurate shift on the screen. For the unrealistic user distances bigger than 776.8mm , the offset is set to zero. In the final system, the horizontal offset was computed only for the OVD, whereas the vertical offset was ignored.

An intuitive and non-intrusive feedback helping the user to find a sweet spot was implemented in the final system and its mechanisms are illustrated in Figure 7. First, the circles denoting a position where the viewer's left and right eyes should be placed are shown together with camera preview. The position and radius of the circles are based on the iterative combination of four equations:

- Equation (4) is used to compute OVD according to the user IPD (internal parameter of the system).
- Equation (8) is used to shift the circles with an adequate offset corresponding to the OVD.
- The OVD is then mapped to a theoretical eye distance in pixels by using the camera calibration equation (7).
- The diameter of the circles depends on the viewing freedom, which is computed using (5) according to viewer preference for crosstalk tolerance.

The color of the circles is changed from red to green when the eyes of the viewer enter in the sweet spot feedback circles. When good viewing conditions are detected for several consequent frames, feedback circles disappear and 3D content is displayed.

2.4 Artifacts correction

Relative user position with respect to the screen determines the particular corrections of the visual information. Corrections of crosstalk visual artefact are carried out in three different ways: pseudoscopy correction, crosstalk compensation and 2D/3D switching, as illustrated in Figure 8. Whereas main sweet spot position implies no processing of the visual information, in one of the side sweet spot position the pseudoscopy must be corrected. For the position between sweet spots, the crosstalk compensation is performed depending on the crosstalk tolerance threshold. At the edge positions, when the viewing angle is too large, the display mode is switched back to 2D.

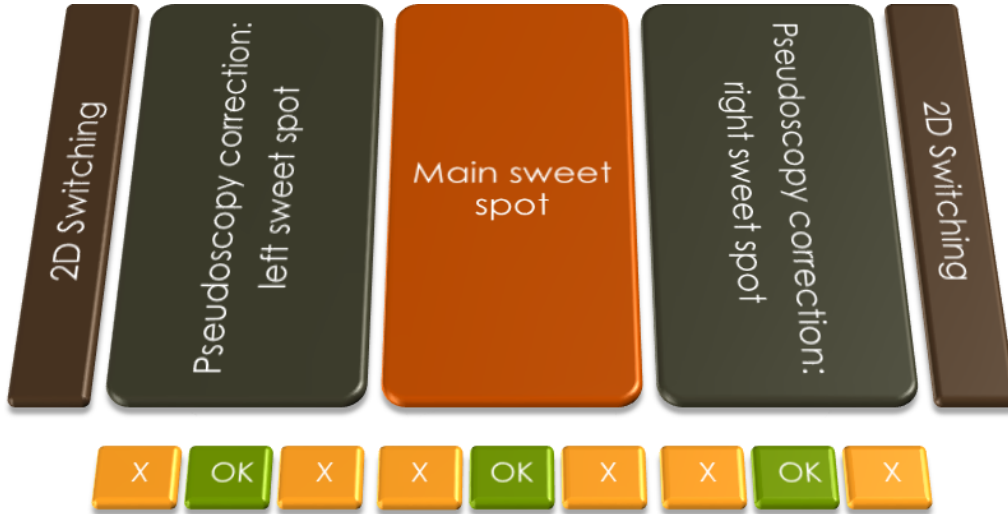


Figure 8: Overview of artefacts correction.



Figure 9: Active crosstalk compensation process.

Pseudoscopy is corrected by swapping left and right views when the user is in a “pseudoscopic” sweet spot, *i.e.* when the left eye sees the image intended for the right eye and vice versa. More specifically, the left and right views are swapped when the crosstalk for left or right eye is higher than 50%. Pseudoscopy correction considers four central sweet spots illustrated as blue and red triangles in Figure 5.

The simple yet effective method for crosstalk reduction has been implemented according to,¹⁷ and it is illustrated in Figure 9. First, the crosstalk values X_L and X_R are retrieved from the crosstalk map for the left and right eye. Subsequently, the unintended visual information X_LR and X_RL is subtracted from the left L and right view R , respectively. Each RGB channel is finally normalized independently into the range $[0, 255]$. All operations are carried out on matrices, instead of bitmaps, in 16 bits representation to allow negative values and normalization. The average time to process a pair of views whose size is 540×480 pixels is $200ms$, which is slower than real-time ($15fps \rightarrow 67ms$), however still sufficient to improve the 3D rendering.

Switching to 2D mode has been implemented when crosstalk information is not available for a user position. To get a real-time response and to keep the same level of the display luminosity when switching from 2D to 3D rendering mode, the 3D device capability is always enabled and only one image (left or right) is displayed for both views. In 2D mode, left image was chosen to be displayed for both views by default.

3. EXPERIMENT

In this section, the proposed crosstalk reduction system is evaluated through a subjective assessment. More specifically, the user preference between two standard modes (2D and 3D) of the mobile phone and the proposed system is analyzed in terms of two different aspects, picture quality and depth perception quality.



Figure 10: Left view of the images used for subjective evaluation, image (a) was used for training.

3.1 Image dataset

The dataset consists of six JPS images in side by side stereo format with different disparities and colors. Resolution of the content was set to 540×960 pixels (resolution of display) for efficiency reasons. One content was used for training (see Figure 10(a)) and the rest for the test purposes (see Figure 10(b)-(f)).

3.2 Test methodology

Several methods have been proposed for the quality evaluation of 2D¹⁸ and 3D¹⁹ images and video sequences. Since judging the quality of different 2D and 3D restitution techniques individually may be quite difficult, the stimulus comparison method seems to be the most suitable method for the subjective test.

The subjective test has been performed as follows. A pair of test stimuli “A” and “B” was shown sequentially on the mobile phone to the viewer using the developed Android application. After comparing the two test stimuli, the subject is asked to choose his preference (“A” or “B”) in terms of picture quality and depth perception quality as suggested in ITU recommendation.¹⁹ The option “same” was also included to avoid random preference selections. For each of the 5 test images (Cactus, Dog, Flowers, Landscape, and Mountains) all the possible combinations of the 3 test conditions (2D, 3D and the proposed “new3D”) were considered. This resulted in a test set with $5 \times \binom{3}{2} = 15$ paired comparisons. The two asked questions for each test condition were closely specified as:

- Picture quality: Tilt the phone a little bit! Which scenario leads to the best image quality?
- Depth quality perception: Which scenario gives better depth or 3D perception?

Eighteen naive subjects (16 male and 2 female) participated in our subjective experiment. They reported normal or corrected to normal vision according to ITU recommendation.¹⁸ Moreover, their stereo acuity was tested using the Randot test. Before the test, oral and written instructions were given to the participants to explain their tasks and different aspects of the evaluation, such as short description of 3D technology and common artefacts. Then, the subject’s IPD was measured and set as a parameter of the application manually. Additionally, a training session was organized to allow participants to familiarize with the assessment procedure. The experiment was performed in normal daylight conditions. The crosstalk tolerance threshold for the proposed 3D rendering scheme was set to 5%.

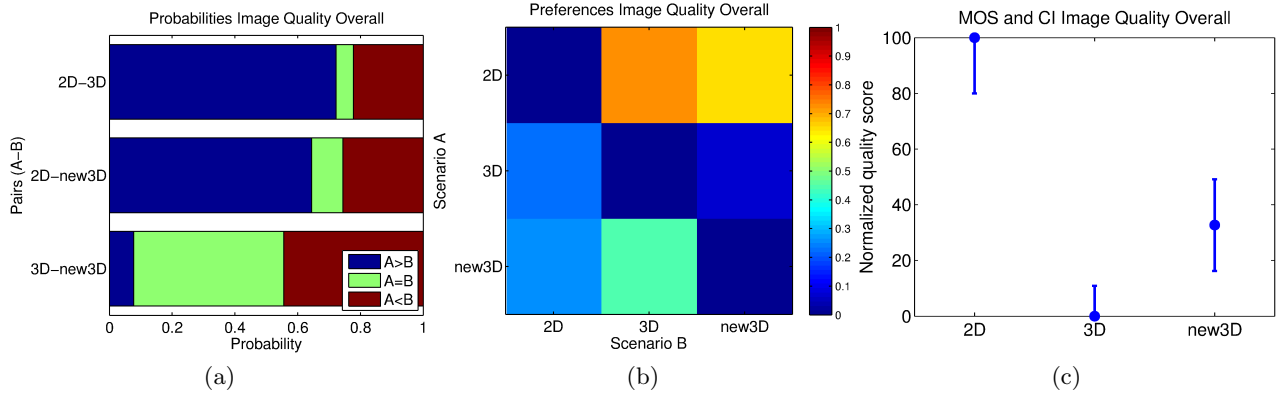


Figure 11: Image quality evaluation results.

3.3 Results and analysis

After collecting the preference ratings from the individual users, statistical tools were applied to analyze the preferences for the different scenes and test conditions. To analyze a set of paired comparisons, the distribution of the votes over the different categorical levels (“A”, “same”, “B”) and their normalization by the number of subjects was computed. Figure 11(a) and Figure 12(a) show the resulting probabilities, averaged over all image content, for both evaluated aspects. In terms of overall picture quality, the proposed 3D rendering outperforms the 3D mode with a ratio of preference probability 5 : 1. The general preference of 2D mode over proposed 3D rendering and 3D mode is again demonstrated with a preference probability of 65% and 75%, respectively. The results for the perceived depth quality are, as expected, in favor of both 3D modes. The proposed 3D rendering is much better when compared to the normal 3D mode. This can be explained by the fact that the crosstalk and pseudoscopy reduce perceived depth in general, as showed by Tsirlin et al.²⁰

For a more detailed analysis of the individual test conditions and their performance with respect to each other, a preference matrix from the individual paired comparisons by discarding all the ties can be constructed. It provides the preference probabilities of a test condition A versus another test condition B along the rows. Figure 11(b) and Figure 12(b) show the preference probability matrices, averaged over all the image content, for both evaluated aspects. Analysing the overall matrices shows that the performance of the proposed 3D rendering is the best in terms of depth perception quality. Moreover, the proposed method wins against 3D mode with respect to the overall quality.

Continuous quality scores can be obtained by applying the Bradley-Terry-Luce (BTL) model²¹ on the preference probability matrix. In this model the preference probability P_{ij} of choosing i over j can be represented as

$$P_{ij} = \frac{\pi_i}{\pi_i + \pi_j}. \quad (9)$$

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} .²¹ 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 11(c) and Figure 12(c) show the obtained overall MOS (Mean Opinion Score) and CI averaged over all the image content for both evaluated aspects. The comparison of depth perception for the different rendering modes shows that the proposed 3D rendering clearly outperforms the others. Moreover, it offers significant improvement in terms of picture quality in comparison to normal 3D rendering mode.

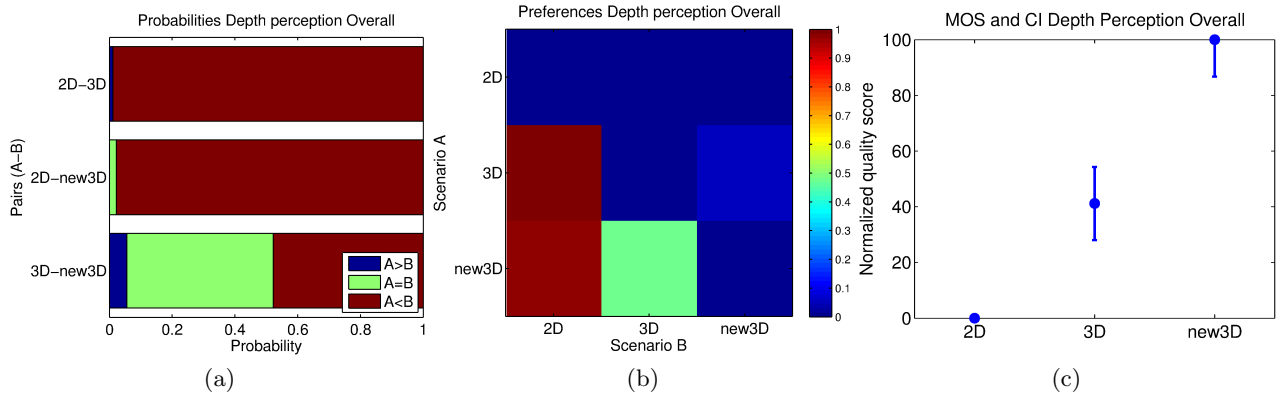


Figure 12: Depth quality perception results.

4. CONCLUSION

In this paper, a complete active crosstalk reduction system running on an HTC EVO 3D smartphone was proposed, implemented and evaluated. A special Android application was implemented to track the user face and eyes, and to correct artefacts in real-time according to his/her position. Artefact correction is based on autostereoscopic display characterization and each view visibility profiles determination. Furthermore, the localization of sweet spot and computation of the viewing freedom was performed. Moreover, the proposed system was designed in the way that it first helps user to find the sweet spot and then it corrects the crosstalk artefact and/or pseudoscopy. This means that the performance of the proposed system is limited by the face recognition and eyes tracking capabilities depending on lighting condition. The evaluation of the Android application showed that there are also limitations in terms of processing speed and power usage. In order to compare the capabilities and performance of the conventional and proposed system, the subjective assessment was conducted. It was demonstrated that the new 3D restitution method significantly improves normal 3D in terms of image quality and depth perception.

This work could be extended in the future in many directions to achieve better QoE. An automatic IPD computation, eye tracking improvement and transparency of the feedback might be the possible aspects in one direction. Implementation of more complex crosstalk compensation algorithm while keeping the computational complexity low is another challenge for the future work as well as the extension of this approach for n-view autostereoscopic display by using a combination of sub-pixels interlacing and motion parallax phenomenon.

ACKNOWLEDGMENTS

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1) and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET.

REFERENCES

- [1] Boev, A., Georgiev, M., Gotchev, A., and Egiazarian, K., "Optimized single-viewer mode of multiview autostereoscopic display," in *[16th European Signal Conference, EUSIPCO]*, (2008).
- [2] Boev, A., Georgiev, M., Gotchev, A., Daskalov, N., and Egiazarian, K., "Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display," in *[17th European Signal Conference, EUSIPCO]*, (2009).
- [3] Park, J., Nam, D., Sung, G., Kim, Y., Park, D., and Kim, C., "61.4: Active crosstalk reduction on multi-view displays using eye detection," *SID Symposium Digest of Technical Papers* **42**(1), 920–923 (2011).
- [4] Chen, Y.-S., Su, C.-H., Chen, J.-H., Chen, C.-S., Hung, Y.-P., and Fuh, C.-S., "Video-based eye tracking for autostereoscopic displays," *Optical Engineering* **40**(12), 2726–2734 (2001).

- [5] Yang, J.-C., Wu, C.-S., Hsiao, C.-H., Tsai, R.-Y., and Hung, Y.-P., “Evaluation of an Eye Tracking Technology for 3D Display Applications,” in [*3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*], 345–348, IEEE (2008).
- [6] Chae, H.-B., You, Y.-R., Kwon, S.-C., and Lee, S.-H., “Three-dimensional display system using a variable parallax barrier and eye tracking,” *Optical Engineering* **50**(8), 087401–087401 (2011).
- [7] Yi, S.-Y., Chae, H.-B., and Lee, S.-H., “Moving parallax barrier design for eye-tracking autostereoscopic displays,” in [*3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*], 165–168, IEEE (2008).
- [8] Dodgson, N. A., “On the number of viewing zones required for head-tracked autostereoscopic display,” in [*Electronic Imaging 2006*], 60550Q–60550Q, International Society for Optics and Photonics (2006).
- [9] International Committee for Display Metrology - IMDS, “Information display measurements standard,” (June 2012).
- [10] Abileah, A., “3-D displays — Technologies and testing methods,” *Journal of the Society for Information Display* **19**(11), 749–763 (2011).
- [11] Boher, P., Leroux, T., Bignon, T., and Collomb-Patton, V., “A new way to characterize autostereoscopic 3D displays using Fourier optics instrument,” in [*Stereoscopic Displays and Applications XX*], *Proc. SPIE* **7237**, 72370Z–72370Z–12 (2009).
- [12] Boher, P., Leroux, T., Bignon, T., and Collomb-Patton, V., “Optical characterization of different types of 3D displays,” in [*Advances in Display Technologies II*], *Proc. SPIE* **8280**, 82800A–82800A–16 (2012).
- [13] Sykora, M., Schultz, J., and Brott, R., “Optical characterization of autostereoscopic 3D displays,” in [*Stereoscopic Displays and Applications XXII*], *Proc. SPIE* **7863**, 78630V–78630V–8 (2011).
- [14] Hong, H., “Simple method of characterizing the spatial luminance distribution at the user position for autostereoscopic 3-D display,” *Journal of the Society for Information Display* **20**(2), 118–122 (2012).
- [15] Salmimaa, M. and Järvenpää, T., “Optical characterization and measurements of autostereoscopic 3D displays,” in [*Photonics in Multimedia II*], *Proc. SPIE* **7001**, 700102–700102–9 (2008).
- [16] Bradski, G. and Kaehler, A., [*Learning OpenCV*], O’Reilly Media Inc. (2008).
- [17] Daly, S., Held, R., and Hoffman, D., “Perceptual issues in stereoscopic signal processing,” *Broadcasting, IEEE Transactions on* **57**(2), 347–361 (2011).
- [18] ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures.” International Telecommunication Union (January 2012).
- [19] ITU-R BT.2021, “Subjective methods for the assessment of stereoscopic 3DTV systems.” International Telecommunication Union (August 2012).
- [20] Tsirlin, I., Allison, R. S., and Wilcox, L. M., “Crosstalk reduces the amount of depth seen in 3D images of natural scenes,” in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], **8288** (Feb. 2012).
- [21] Glickman, M. E., “Parameter estimation in large dynamic paired comparison experiments,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **48**(3), 377–394 (1999).