

# Guidelines for an improved quality of experience in 3-D TV and 3-D mobile displays

Di Xu  
Lino E. Coria  
Panos Nasiopoulos

**Abstract** — Stereoscopic (3-D) movies have become widely popular all over the world. In addition, 3-D TVs and mobile devices have already been introduced to the consumer market. However, while some manufacturers are introducing 3-D cameras and movie studios are using proprietary solutions, there are no guidelines for consistently capturing high-quality stereoscopic content. In this paper, a comprehensive stereoscopic image and video database with content captured at various distances from the camera lenses and under different lighting conditions will be presented. Subjective tests to assess the perceived 3-D quality of these videos and images, which were shown on displays of different sizes, have been conducted. In addition, the horizontal parallax of the content was adjusted to verify via subjective tests whether this change could increase the viewer's quality of experience. Finally, guidelines of acquisition distances between the cameras and the real scene will be published.

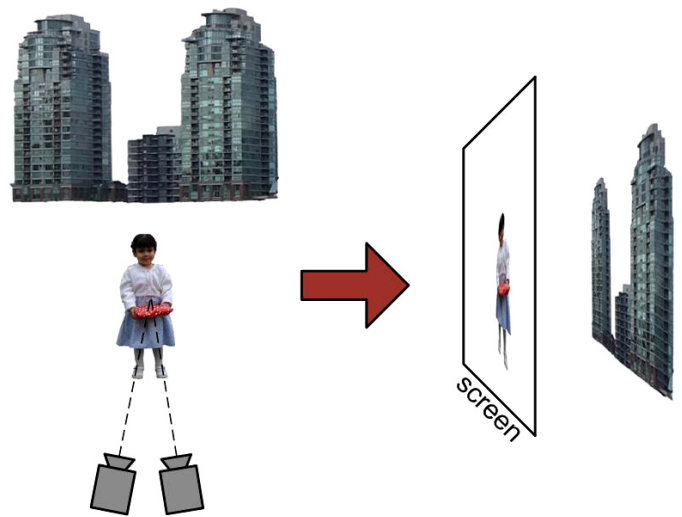
**Keywords** — *Quality of experience, stereoscopic content, capturing guidelines, database, 3-D display.*

DOI # 10.1889/JSID20.7.397

## 1 Introduction

Stereoscopic 3-D technology has become one of the main driving forces in the consumer-electronics market and the entertainment industry.<sup>1</sup> Hollywood studios are releasing most of their high-budget movies in 3-D and there is a vast selection of 3-D TVs available for regular consumers. In addition, other devices capable of displaying stereoscopic content will soon be offered. Many viewers, however, are yet to be convinced of the value of 3-D technology. Some of the criticism expressed is directly related to the headaches and nausea, which are more evident when 3-D content is viewed on home 3-D displays. In order to increase the quality of stereoscopic content for household displays, it is necessary to gain a better understanding of the technical and artistic challenges of this medium. Over the years, stereographers have empirically obtained a few rules of thumb for capturing stereoscopic content.<sup>2-4</sup> Unfortunately, this pragmatic set of recommendations has not been quantified, and there has not been an effort to systematically measure its effectiveness.

One of the main factors for capturing high-quality stereoscopic content is the proper setup of the two cameras because it allows content creators to control the 3-D effect.<sup>4</sup> There are basically two options for setting up the cameras. The first option is having the cameras converge as shown in Fig. 1. For this example, the camera axes converge on the little girl. When the 3-D video is displayed, the image of the girl will appear on the plane of the screen since, for this object, there is no disparity between the left and right views. In the case of the building, however, there will be a difference along the  $x$  (horizontal) axis between the right and left views. This is known as a *positive* horizontal parallax and it



**FIGURE 1** — In order to capture stereoscopic video, the cameras are set up so that their axes converge on a particular object. This object will appear in the plane of the 3-D screen.

makes the building appear to be behind the screen. Changing the angle of convergence between the cameras can be used to control which objects will pop out from the screen and which objects will remain inside. Unfortunately, this camera configuration has shown to have side effects that produce undesirable distortions to the stereoscopic depth.<sup>5</sup> The main distortion caused by this setup, known as the key-stone effect,<sup>6</sup> creates vertical disparities in the four corners of the screen.

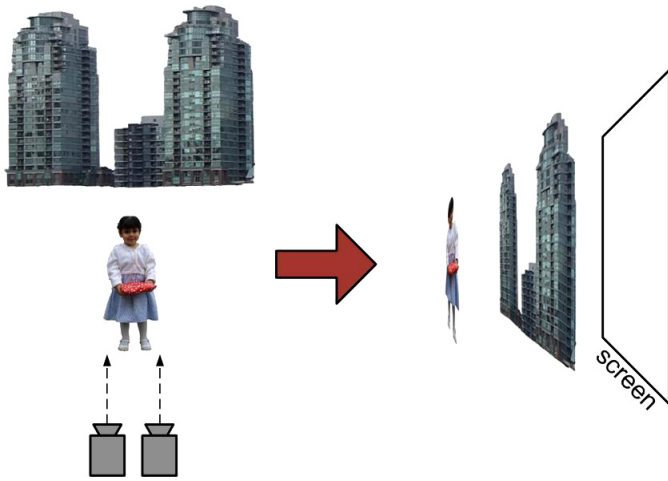
The second option, which seems to be more popular, consists of setting up the two cameras in parallel (see Fig. 2). The cameras converge at infinity and the resulting

Received 01/26/12; accepted 04/25/12.

D. Xu and P. Nasiopoulos are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, V6T 1Z4 Canada; telephone 604/827-4878, e-mail: dixu@ece.ubc.ca

L. E. Coria is with the Institute of Computing, Information, and Cognitive Systems, University of British Columbia, Vancouver, British Columbia, Canada.

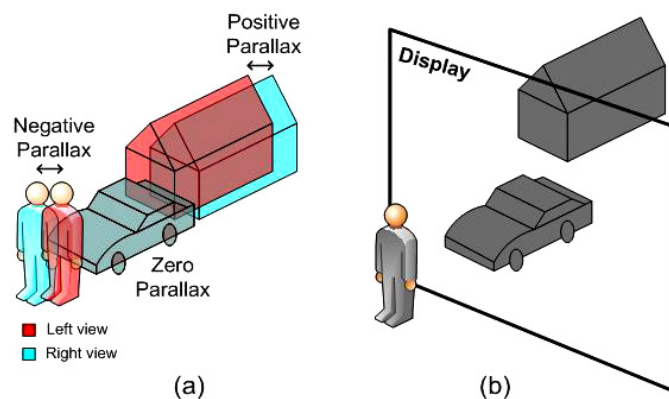
© Copyright 2012 Society for Information Display 1071-0922/12/2007-0397\$1.00.



**FIGURE 2** — Another set-up for capturing stereoscopic video. The cameras are parallel and their axes converge at infinity. All the photographed objects appear to be in front of the screen.

3-D scene appears to be entirely in front of the screen. Each photographed object in this case is known to have a *negative* horizontal parallax. This negative parallax occurs when the left view of an object is located further to the right than the right-view version of the same object. The three different possible types of horizontal parallaxes (negative, zero, and positive) are illustrated in Fig. 3.

It is commonly accepted that being exposed for a considerable amount of time to objects that appear to be in front of the screen (*i.e.*, with negative parallax) makes viewers uncomfortable. As described above, when 3-D content has been shot with two parallel cameras, all the objects have negative parallax and, therefore, it is good practice to modify the content in order to reallocate the 3-D effect behind the display. In order to do this, it is necessary to modify the depth information that is produced when the 3-D content is captured. One solution is proposed in Ref. 7, employing an algorithm that modifies horizontal disparity in a non-linear fashion by warping the input video streams. Unfortunately,



**FIGURE 3** — (a) Examples of negative, zero, and positive horizontal parallaxes. The red objects are from the left view and the cyan objects are from the right view; (b) when watching 3-D content, negative parallax results in objects popping out of the screen, zero parallax positions objects on the screen, and positive parallax results in objects appearing behind the screen.

for image regions with frequent and strong changes in disparity, this warping scheme can lead to visible distortions. Another way of changing the depth information is by reducing the negative horizontal parallax of 3-D videos by shifting the left frames towards the left and the right frames towards the right. Although this action introduces black lines on the vertical edges of the frames, this inconvenience can be side-stepped by cropping the content (to match the aspect ratio) and then scaling it up.

Developing a reliable objective quality metric for 3-D content has proven to be very challenging.<sup>8,9</sup> Therefore, researchers have mainly relied on subjective evaluations such as Refs. 10–14 to identify the key factors for producing high-quality stereoscopic content.

In order to successfully assess user experience, the opinion scores must be taken from an adequate sample of typical users carrying out representative tasks in a realistic context of use.<sup>15</sup> Because of this, more meaningful results will be obtained if the media employed for these tests resembles content that is actually being shown on 3-D TVs (*i.e.*, featuring people and objects in ordinary surroundings instead of an artificial lab setting). Testing both images and video sequences is also desirable because spectators might perceive quality differently for different types of content.

In this study, we tested the effect that different distances (measured from the 3-D camera setup to the photographed objects) have on the quality of the stereoscopically captured images and videos when viewed on home 3-D TVs and 3-D mobile devices. We have created our own stereoscopic image and video database that is comprised of scenes depicting people and landscapes with various distances between the cameras and the subjects. Our subjective assessment exercise is comprised by three stages. During the first stage, several viewers of different ages watched and rated the stereoscopic images that we captured using various distances between the cameras and the subjects. For the second stage, the content consisted on stereoscopic video sequences that were shot using the same combination of distances as in the previous test. Viewers were asked again to rate the 3-D quality of the content. Finally, we performed subjective evaluations to verify and quantify the influence in the 3-D quality of experience caused by the adjustment of the horizontal parallax.

The rest of the paper is organized as follows. Section 2 describes the 3-D content acquisition and alignment processes. The subjective evaluation environment and parameters are specified in Sec. 3. In Sec. 4, we present the statistical analysis of the subjective test scores and discuss the findings of the tests. We conclude the paper in Sec. 5.

## 2 Acquisition and alignment

### 2.1 Equipment

In order to capture stereoscopic video and images, we employed two identical HD cameras with the same firmware and settings. These cameras were aligned in parallel and attached



FIGURE 4 — Stereo camera setup consisting of two identical HD camcorders.

to a bar that was specifically made for them. Subsequently, the bar was secured to a tripod as shown in Fig. 4. Since zoom lenses may differ,<sup>11</sup> only an extreme end of the zoom range was used. A single remote control was employed to start both cameras simultaneously and obtain the best possible synchronization.

## 2.2 Image and video capturing

The stereoscopic video- and image-capturing process is illustrated in Fig. 5. Both cameras capture slightly different images of the same event. Each event on our database consists mainly of a person or object standing in front of the camera with a wall or a building as background. For all videos, the object of interest is the one that is closest to the cameras. The camera is always kept still while the people and some objects move moderately. There are four important distances that need to be considered for every stereoscopic image or video pair. They are described in Table 1.

TABLE 1 — Distances considered when capturing stereoscopic images and videos for our database.

Distance	Description	Values
$d_{cam}$	Distance between the two cameras	77 mm
$d_{min}$	The distance between the cameras and the closest point captured in the stereoscopic image or video pair	0.5 m, 1 m, 2 m, 3 m
$d_{obj}$	The distance between the cameras and the main object (usually a person). In most cases, $d_{obj} = d_{min}$	0.5 m, 1 m, 2 m, 3 m
$d_{max}$	The distance between the cameras and the furthest background. If the sky is visible, then $d_{max}$ is considered to be infinity	5 m, 10 m, 50 m, infinity

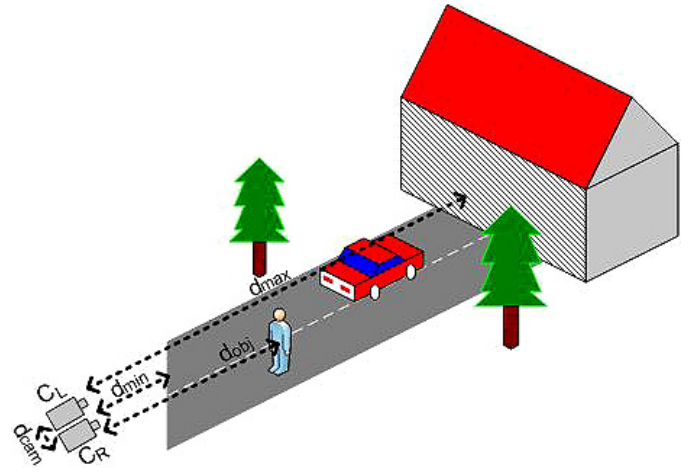


FIGURE 5 — Capturing a live-action event with two parallel cameras  $C_L$  and  $C_R$ ;  $d_{cam}$  is the distance between the cameras,  $d_{min}$  is the distance from the cameras to the closest point,  $d_{obj}$  is the distance from the cameras to the main object (usually a person), and  $d_{max}$  is the distance from the cameras to the background.

## 2.3 Temporal synchronization

Three out of 30 videos were unsynchronized by a few frames and were manually synchronized before further processing.

## 2.4 3-D content alignment

Even though the cameras are carefully lined up, a small amount of vertical disparity between the left and the right views of a stereoscopic image/video is unavoidable. Therefore, it is usually necessary to vertically align the left and right views. This alignment process is performed for every stereoscopic image and video pair in our database.

In addition, for the third stage of our quality-assessment process, we also eliminate the negative horizontal parallax so that the photographed objects do not appear to pop out of the screen. This alignment process involves horizontal frame shifting. By eliminating all negative disparities we avoid stereoscopic window violations<sup>3</sup> (*i.e.*, when objects that pop out of the screen are only partially shown, thus providing the brain with two conflicting depth cues). However, eliminating negative disparities through shifting may result in large positive-disparity regions in the background of a scene, which cause eye divergence. Since the background is usually not the point of interest, especially in the case of video, such divergence does not have a strong negative impact on the quality of the content. This is later verified by the results of our subjective tests in Sec. 4.5.

Both vertical and horizontal alignments involve frame shifting and are performed automatically using features that are common to both left and right videos. The objective is to be able to implement this method on 3-D TV displays and achieve this “correction” in real-time. Our method uses the Scale Invariant Feature Transform (SIFT) algorithm to identify matching features on both views.<sup>16</sup> Having the coordinates to these common features allows us to compute the parallax between the left and right views of the photo-

graphed objects. The following steps describe in detail this algorithm:

1. The first frame of both the left and right videos is downsampled by a factor of 2 (horizontally and vertically) to reduce the number of computations in the next steps and allow real-time implementation.
2. The features of the downsampled left and right frames are obtained using SIFT.
3. The features of the left frame are matched to the features of the right frame. The top 10% of all matching features, whose vertical disparities are considerably different from the median disparity value of all matching features, are detected as outliers. These outlier features are removed to ensure the stability of the algorithm. The Cartesian coordinates of rest of the matching features are stored.
4.  $\Delta y$ , the amount of pixels that each original frame will be shifted vertically, is found by computing the median of all the  $y$  coordinates of matching points between the two frames and then multiplying by 2 (to compensate for the downsampling).
5.  $\Delta x$ , the amount of pixels that each original frame will be shifted horizontally, is computed by finding the largest negative value of all the  $x$  coordinates of matching points between the two frames and then multiplying by 2 (to compensate for the downsampling). The negative number with the largest absolute value of the  $x$  coordinates represents the photographed point in space that is closest to the cameras ( $d_{min}$ ).
6. Finally, the shifted frames are cropped and then enlarged using bicubic interpolation so that they maintain the same size they had before the shifting process ( $1080 \times 1920$  pixels).

An example of a 3-D video frame before and after the shifting algorithm is shown in Figs. 6(a) and 6(b), respectively. For illustration purposes, the stereoscopic frames are displayed in anaglyph (red and cyan) mode. Notice how the superimposed left and right frames are vertically misaligned in Fig. 6(a). This problem has been solved in Fig. 6(b). In addition, the two frames have been horizontally shifted to produce a zero parallax for the closest object (in this case, the photographed individual); the relative positions of the objects behind the subject indicate positive parallax.

### 3 Evaluation environment

#### 3.1 Displays

Our subjective tests were conducted on four different sizes of stereoscopic displays, namely, a 2.8-in. 3-D camera display, a 22-in. 3-D LCD, a 55-in. 3-D LED TV, and a 65-in. 3-D plasma TV. The 2.8-in. display is an autostereoscopic display that can be viewed without glasses, and the other three displays are paired with different 3-D active shutter glasses. The detailed specifications of the four displays are listed in Table 2.



(a)



(b)

**FIGURE 6** — A stereoscopic video frame from an indoor sequence with  $d_{min} = 3$  m,  $d_{max} = 5$  m, and presented in anaglyph mode for illustration purposes: (a) without any vertical and horizontal shifting, (b) after both vertical and horizontal shifting plus cropping and interpolation to preserve the 16:9 aspect ratio.

#### 3.2 Database

The images have a resolution of  $3840 \times 2160$  pixels, and the videos have a resolution of  $1920 \times 1080$  pixels and a frame rate of 30 frames/sec (fps). The database includes 30 images and 30 video sequences with various combinations of  $d_{min}$ ,  $d_{obj}$ , and  $d_{max}$ , where  $d_{min}$  is in  $\{0.5, 1, 2, 3\}$  m,  $d_{obj}$  is in  $\{0.5, 1, 2, 3\}$  m, and  $d_{max}$  is in  $\{5, 10, 50, \text{infinity}\}$  m. We also prepared a different set of 10 images and four videos as training sequences for our test. All images and videos were shot in a natural environment rather than a lab setup. The main objects in the scenes are often people, chairs, toys, and buildings (see a sample in Fig. 7).

#### 3.3 Observers

Nineteen observers participated in the first stage of our test, including six females and 13 males. Their ages ranged from 23 to 59, with an average age of 33. In the second and third

**TABLE 2** — Properties of the 3-D displays used in our test.

Size	Resolution	Refresh Rate	Glasses
2.8"	Approx. 230,000 dots	-- --	No glasses needed
22"	1680 x 1050	120 Hz	3D active shutter glasses
55"	1920 x 1080	240 Hz	3D active shutter glasses
65"	1920 x 1080	600 Hz	3D active shutter glasses

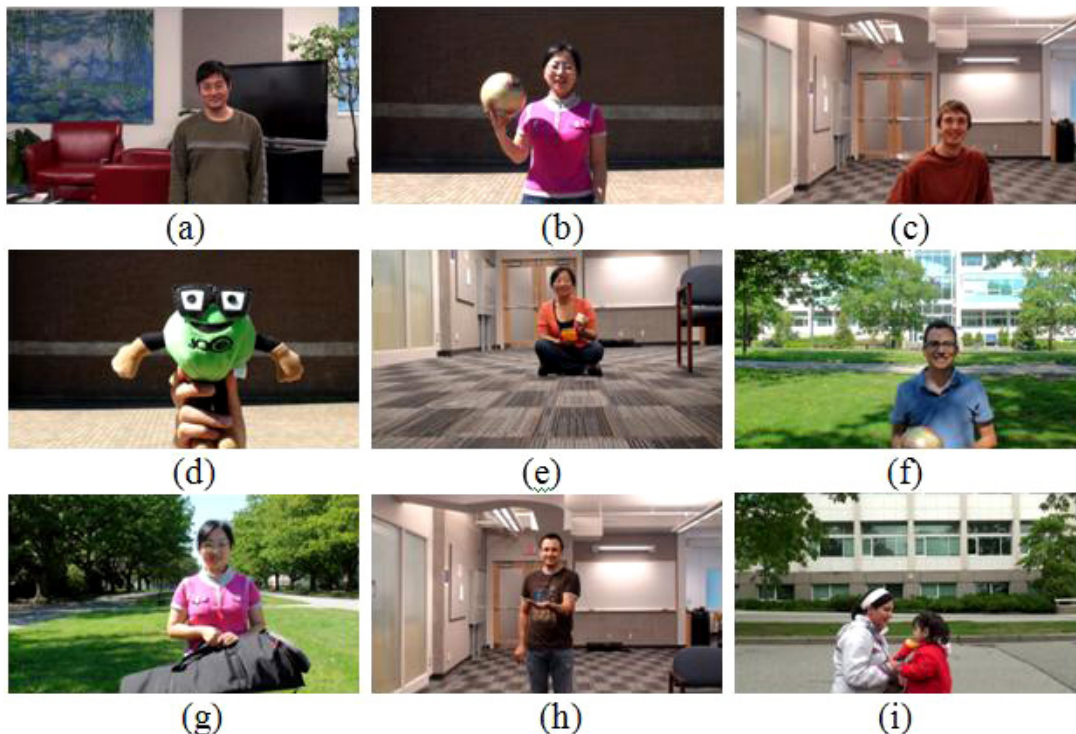


FIGURE 7 — The left view of some images and video frames from our 3-D database.

stages of our test, another 20 subjects participated, including seven females and 13 males. The average age was 32. All observers are non-expert in viewing 3-D images and videos, and they were screened for visual acuity using the Snellen chart and color vision using the Ishihara test.

### 3.4 Testing procedure

We set up the viewing conditions for the subjective assessment according to Sec. 2.1 of the ITU-R BT.500-11.<sup>17</sup> A single stimulus method has been adopted for the subjective quality evaluation. Before the subjective evaluation on each display, we ran a training session to show to the subjects the quality range of our stereoscopic images and videos, without imposing the quality of the content.

In the first stage of our test, 30 test images were used and were shown in a random order on three displays, *i.e.*, the 2.8-in. 3-D camera display, the 22-in. 3-D display, and the 55-in. 3-D TV. During the test, each stereoscopic image was shown for 5 sec followed by a 5-sec interval of a 2-D mid-gray image with the image index as a grading and relaxation period. One to three observers participated in each viewing session. In the second and third stages of our test, 60 10-sec test video sequences were shown on the 65-in. plasma TV, with a 4-sec break of a 2-D mid-gray image as a grading and relaxation period. The 60 videos are two versions of the 30 videos in our database. One version is processed for vertical alignment, and the other is processed by applying horizontal and vertical shifts so that the closest object is at the depth of the screen. The order of the 60

videos is randomized so that videos with similar capturing parameters are inconsecutive and the two versions of the same video are kept far from each other. Two observers conducted the test in each testing session. In all of our tests, the viewers were seated in line with the center of the display, and at the distances that were recommended by the manufactures of the displays.

## 4 Analysis and results

### 4.1 Detection of the outliers

Before analyzing the scores provided by the observers, we first detect the outliers according to the subjective scores they gave. The screening process is based on the guidelines provided in Sec. 2.3.1 of annex 2 of ITU-R BT.500-11 recommendation.<sup>17</sup> For each image or video, we first determine the normality of its score distribution by computing the kurtosis coefficient, which is defined as the fourth moment about the mean divided by the square of the second moment about the mean of  $-3$ . In other words, the  $j$ -th kurtosis coefficient is:

$$kurtosis_j = \frac{m_4}{(m_2)^2} - 3 = \frac{\frac{1}{M} \sum_{i=1}^M (x_{ij} - \bar{x}_j)^4}{\left( \frac{1}{M} \sum_{i=1}^M (x_{ij} - \bar{x}_j)^2 \right)^2} - 3,$$

where  $x_{ij}$  is the score of the  $j$ -th image or video from the  $i$ -th observer and  $\bar{x}_j$  is the average score of the  $j$ -th image or

video over all  $M$  observers. The score distribution is considered normal if  $-1 \leq \text{kurtosis}_j \leq 1$ , and non-normal otherwise. To check if the  $i$ -th observer is an outlier, we initialize two counters  $P_i$  and  $Q_i$  to zeros. The counter values are then updated based on the score  $x_{ij}$  (for all  $i$ ), as follows:

$$\begin{aligned} \text{if } x_{ij} \geq \bar{x}_j + c_j \sigma_j, & \text{ then } P_i = P_i + 1, \\ \text{if } x_{ij} \leq \bar{x}_j - c_j \sigma_j, & \text{ then } Q_i = Q_i + 1, \end{aligned}$$

where  $c_j = 2$ , if the score distribution of the  $j$ -th image or video is normal; and  $c_j = \sqrt{20}$ , otherwise.  $\sigma_j$  is the standard deviation of the scores of the  $j$ -th image or video. Finally, if  $(P_i + Q_i)/N > 0.05$  and  $|(P_i - Q_i)/(P_i + Q_i)| < 0.3$ , where  $N$  is the number of test images or videos, the observer  $i$  is considered as an outlier.

One out of 19 observers was detected as the outlier in the first stage of our test, and another one out of 20 subjects was detected as the outlier in the second and third stages. All the scores of these outliers were eliminated from the subsequent calculation. Therefore, the data analysis in the three stages is based on the scores provided by 18, 19, and 19 valid observers, respectively.

## 4.2 Score computation

We take the average score across all valid observers for each image or video as the mean opinion score. To assess the credibility of the mean opinion score, we use confidence intervals to indicate the reliability of an estimate. The Student's  $t$ -tests<sup>18</sup> are used to compute confidence intervals with the significance level being 95%.

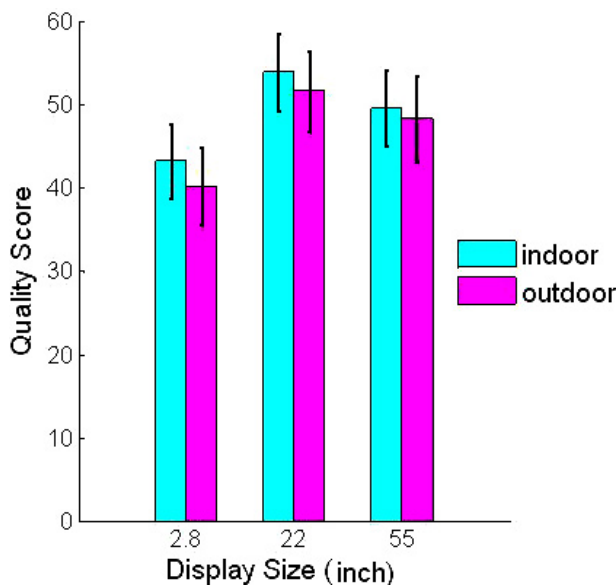


FIGURE 8 — The mean opinion scores and their confidence intervals versus different sizes of 3-D displays under different lighting conditions.

## 4.3 Stage one: Influence of capturing parameters to 3-D image quality on three sizes of displays

We first analyze the quality scores from the first stage of our experiment and reveal how the quality of 3-D images is

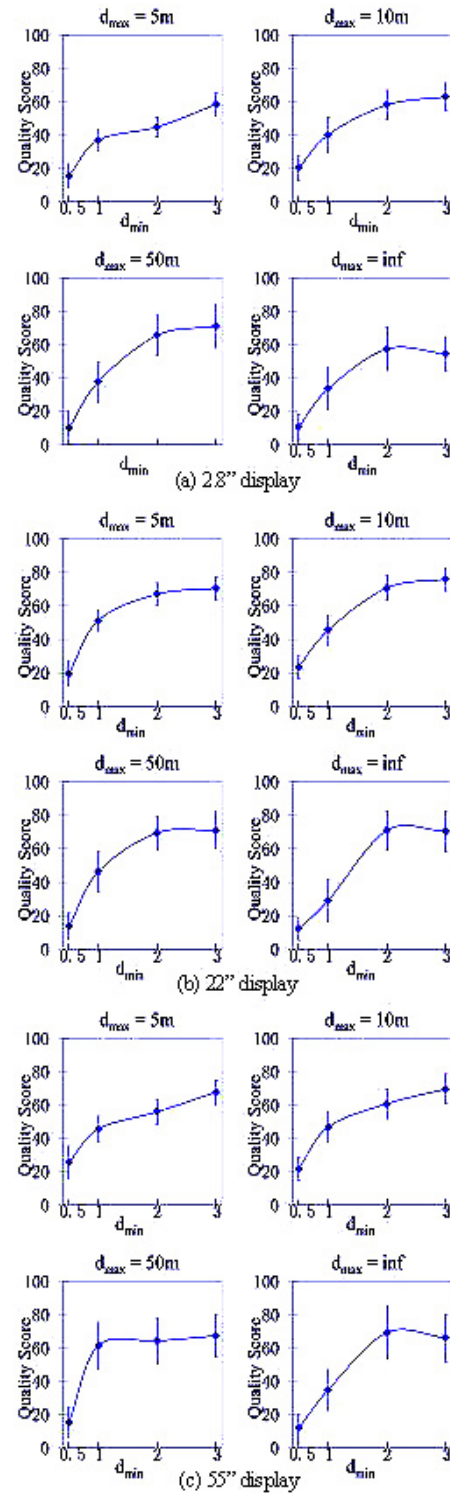
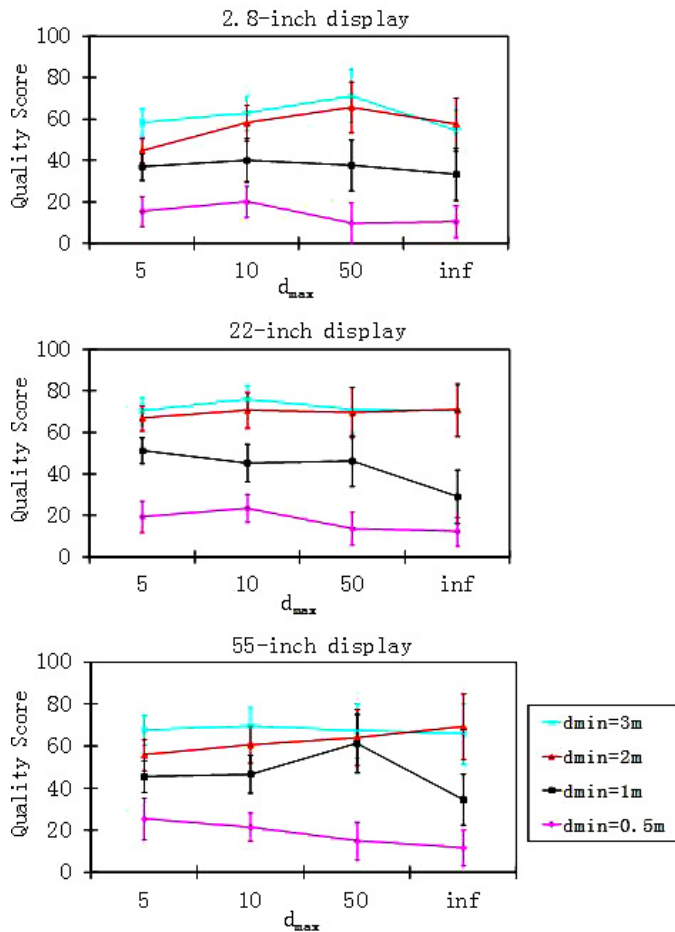


FIGURE 9 — The mean opinion scores and their confidence intervals on different sizes of displays at various  $d_{min}$  (0.5 m, 1 m, 2 m, and 3 m). Parts (a), (b), and (c) show the results associated with the 2.8-, 22-, and 55-in. displays, respectively. The four subplots correspond to the cases when  $d_{max}$  are 5 m, 10 m, 50 m, and infinity.



**FIGURE 10** — The mean opinion scores and their confidence intervals with various  $d_{max}$  (that is, 5 m, 10 m, 50 m, and infinity) at different  $d_{min}$  on three sizes of displays.

affected by the lighting condition,  $d_{min}$ ,  $d_{max}$ , and  $d_{obj}$  not being the foreground object.

### (1) Influence of lighting condition to image quality

Our 3-D image database includes eight sets of images. Each set was captured with the same distance parameters (*i.e.*,  $d_{min}$ ,  $d_{obj}$ , and  $d_{max}$ ) but under two different lighting conditions (outdoors on a sunny day and indoors). For every image, our cameras provided the best exposure parameters. However, indoor images tend to have a more uniform light distribution whereas outdoor images have some bright regions that contrast with some dark ones. We grouped these images by their lighting conditions, and the mean opinion scores of the indoor images and outdoor images are compared in Fig. 8. Indoor lighting results in slightly higher 3-D quality than the outdoor lighting for all three sizes of displays. The quality difference between these lighting conditions, however, is insignificant. Therefore, in the following subsections, the mean opinion score of each capturing-parameter set is the average score over the indoor and outdoor scenes.

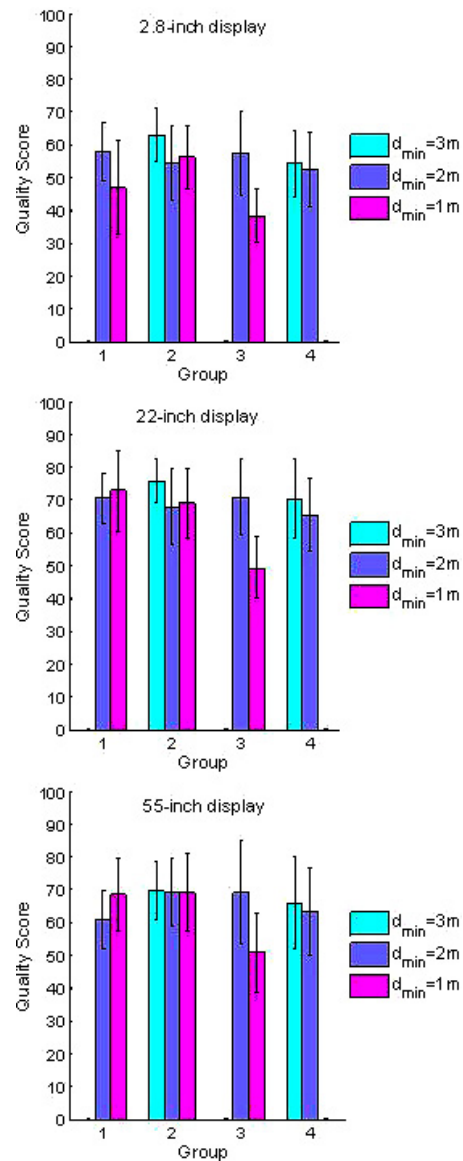
### (2) Influence of $d_{min}$ to image quality

We compared the subjective quality between images taken at different  $d_{min}$  when  $d_{min}$  is the same as  $d_{obj}$ . Figure 9 shows the mean opinion scores and confidence intervals *vs.*

$d_{min}$  at different  $d_{max}$  distances. The figure indicates that for the same  $d_{max}$ , the image quality increases with  $d_{min}$  and levels off when  $d_{min}$  is beyond 2 m. The confidence intervals when  $d_{min}$  is 0.5 m are smaller than those when  $d_{min}$  is large. In other words, the observers consistently provided low scores when the closest object was very close to the cameras. Figure 9 shows the results based on scores from the three displays. The quality trend affected by  $d_{min}$  is the same for all three sizes of displays.

### (3) Influence of $d_{max}$ to image quality

We compared the quality scores between images with different  $d_{max}$  while keeping the same  $d_{min}$ . No clear trend is observed from Fig. 10. Thus, we conclude that  $d_{max}$  does



**FIGURE 11** — Comparison of the mean opinion scores and confidence intervals for four groups of content. Each group of content was captured at the same  $d_{max}$  and  $d_{obj}$  with different  $d_{min}$ . Group 1 was captured at  $d_{max} = 10$  m and  $d_{obj} = 2$  m, group 2 was captured at  $d_{max} = 10$  m and  $d_{obj} = 3$  m, group 3 was captured at  $d_{max} = \text{infinity}$  and  $d_{obj} = 2$  m, and group 4 was captured at  $d_{max} = \text{infinity}$  and  $d_{obj} = 3$  m.

not strongly affect the quality of 3-D content. Again, the same conclusion can be drawn for different sizes of displays.

#### (4) Influence of $d_{obj}$ not being the foreground object to image quality

We tested a few images where the object of interest is not the closest object in the image, that is, when  $d_{obj}$  is greater than  $d_{min}$ . We compared image sets with the same  $d_{obj}$  and the same  $d_{max}$ , but various  $d_{min}$ . The mean opinion scores and confidence intervals associated with different sizes of displays are shown in Fig. 11. In each group of images, the left most bar is associated with the image where  $d_{min}$  equals  $d_{obj}$ , and the other bars are associated with images where  $d_{min}$  is less than  $d_{obj}$ . Having compared the four sets of images, we observe that the quality of most images is impacted to certain extent when some foreground objects, such as floor and ceiling, appear closer to the cameras than the object of interest.

### 4.4 Stage two: Influence of capturing parameters to 3-D video quality

Having discussed the influence of capturing parameters to 3-D image quality, we studied the influence of the same parameters to 3-D video quality in the second stage of our experiment. All 3-D videos in our database are with moderate motion. The subjective test was performed on a 65-in. 3-DTV.

#### (1) Influence of lighting condition to video quality

The statistical results of video quality were compared between the indoor and outdoor videos with the same set of param-

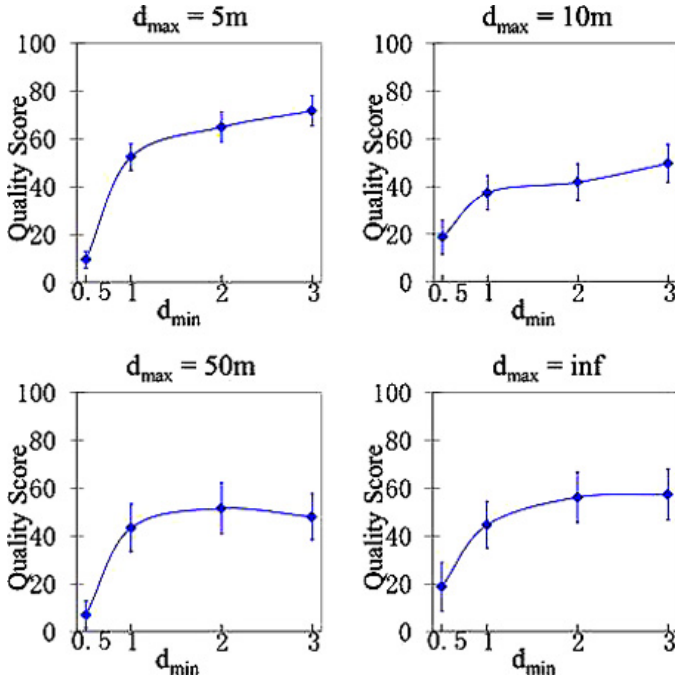


FIGURE 12 — The mean opinion scores and their confidence intervals at various  $d_{min}$  (that is, 0.5 m, 1 m, 2 m, and 3 m). The four subplots correspond to the cases when  $d_{max}$  are 5 m, 10 m, 50 m, and infinity.

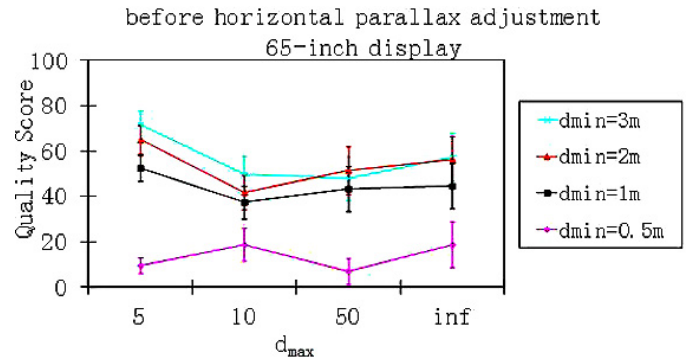


FIGURE 13 — Comparison of the mean opinion scores at  $d_{max}$  equal to 5 m, 10 m, 50 m, and infinity, with different  $d_{min}$ .

ters. No significant difference is found between videos taken under these lighting conditions. The conclusion is consistent with that of stage one.

#### (2) Influence of $d_{min}$ to video quality

In this comparison, we chose videos where  $d_{obj}$  equals  $d_{min}$ . These videos were divided into four groups according to the value of  $d_{max}$ . The subjective quality between videos taken at different  $d_{min}$  is compared in each group. Figure 12 shows the quality score versus  $d_{min}$  at different  $d_{max}$  distances. For the same  $d_{max}$ , the video quality increases with  $d_{min}$  and levels off when  $d_{min}$  is greater than 2 m. The same trend was found in stage one of our test.

#### (3) Influence of $d_{max}$ to video quality

We examined the influence of  $d_{max}$  to the video quality scores. Figure 13 shows the mean opinion scores of videos with different  $d_{max}$  while keeping the same  $d_{min}$ . No clear trend is found based on Fig. 13, although the quality scores vary with  $d_{max}$ . The result is also consistent with that from stage one.

#### (4) Influence of $d_{obj}$ not being the foreground object to video quality

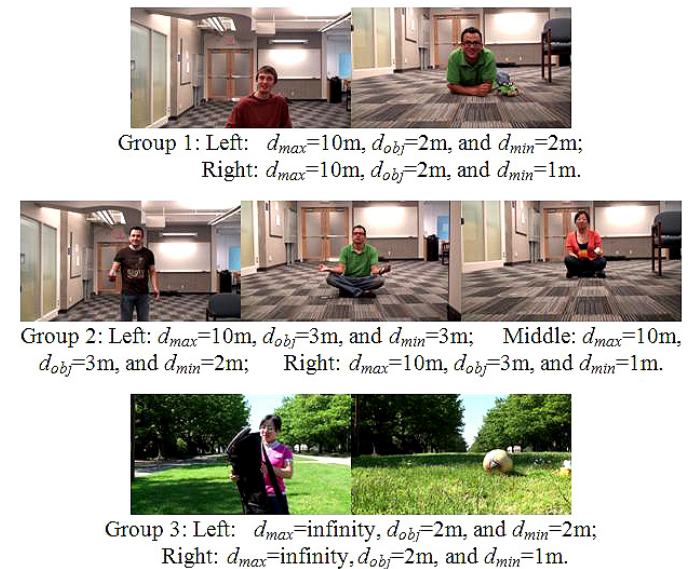
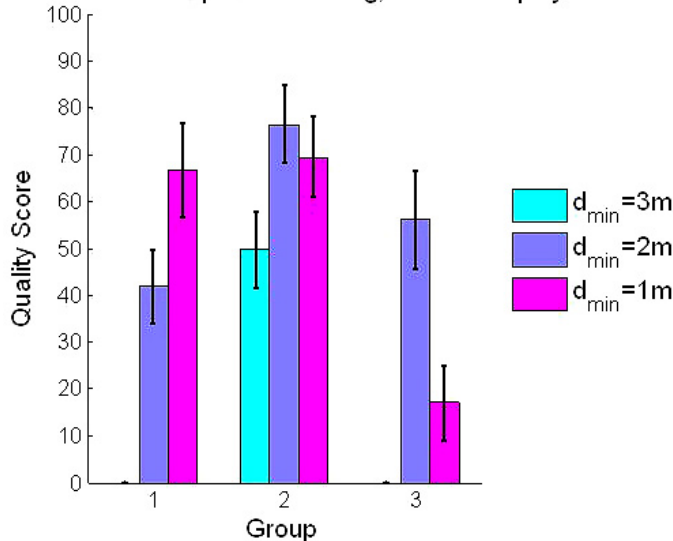


FIGURE 14 — Frames from three groups of videos used to examine the influence of  $d_{obj}$  not being the foreground object.



before horizontal parallax shifting, 65-inch display



**FIGURE 15** — Comparison of the mean opinion scores and confidence intervals for three groups of videos before horizontal parallax adjustment on the 65-in. display. Each group of content was captured at the same  $d_{max}$  and  $d_{obj}$  with different  $d_{min}$ . Group 1 was captured at  $d_{max} = 10$  m and  $d_{obj} = 2$  m, group 2 was captured at  $d_{max} = 10$  m and  $d_{obj} = 3$  m, and group 3 was captured at  $d_{max} = \text{infinity}$  and  $d_{obj} = 2$  m.

We chose three groups of videos. Within each group, the videos are captured with the same  $d_{obj}$ , the same  $d_{max}$ , and different  $d_{min}$ . Sample frames of some of these videos are shown in Fig. 14. The mean opinion scores and confidence intervals of each group are shown in Fig. 15. Based on the ratings of videos in group 3, we note that the grass being the foreground is annoying to the observers. The patterned floor in groups 1 and 2, however, increases 3-D quality. Most observers prefer to watch videos with the patterned floor popping out of the screen and appearing in front of the object of interest.

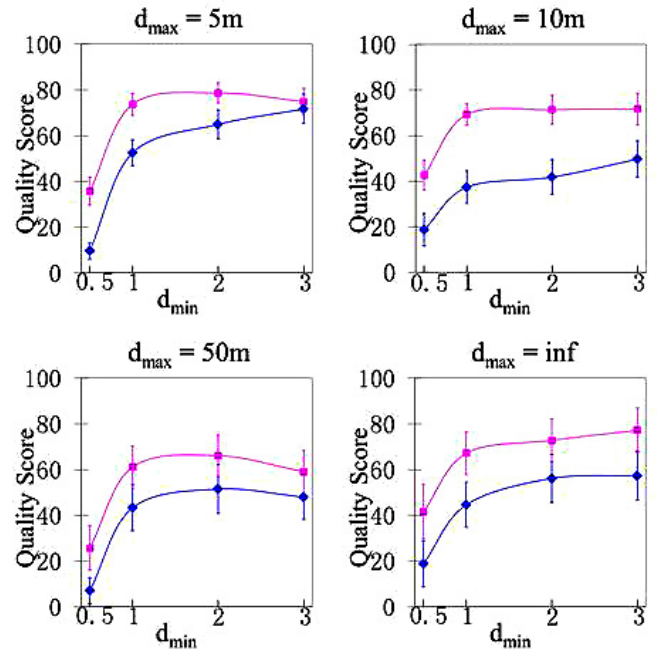
Since the results presented in stage one and stage two are mainly consistent, we conclude that the influence of capturing parameters to 3-D image quality and video quality are the same, providing no very fast motion is included in the video. Therefore, in stage three, we will focus only on 3-D videos.

#### 4.5 Stage three: Influence of capturing parameters to 3-D video quality after horizontal parallax adjustment

The horizontal parallax adjustment was proposed to improve the 3-D quality using simple post-processing. In this section, we reveal how the parallax adjustment described in Sec. 2.4 affects the quality of 3-D content.

##### (1) Influence of lighting condition to video quality after horizontal parallax adjustment

Despite the quality changes brought by the horizontal parallax shifting, the influence of lighting conditions to 3-D quality remains the same. In other words, there is still no significant quality difference between indoor scene with



◆ before horizontal parallax adjustment  
 ■ after horizontal parallax adjustment

**FIGURE 16** — The mean opinion scores and their confidence intervals at various  $d_{min}$  (that is, 0.5 m, 1 m, 2 m, and 3 m). In reading order, the four subplots correspond to the cases when  $d_{max}$  are 5 m, 10 m, 50 m, and infinity. The blue lines represent data before horizontal parallax adjustment and the pink lines show the data after horizontal parallax adjustment.

uniform artificial lighting and outdoor scene under direct sunlight.

##### (2) Influence of $d_{min}$ to video quality after horizontal parallax adjustment

The quality of 3-D content after parallax adjustment is again significantly affected by  $d_{min}$ . Figure 16 shows the quality scores for videos with various  $d_{min}$  before and after horizontal parallax adjustment. The quality increases with  $d_{min}$  and it levels off at  $d_{min}$  equals 2 m for videos both before parallax adjustment and after. The quality improvement by parallax adjustment is significant, with an exception of one video where  $d_{max} = 5$  m and  $d_{min} = d_{obj} = 3$  m. This exception is due to the fact that in this specific setup the depth bracket (*i.e.*, the amount of 3-D space used in a shot or a sequence) is very small, hence there is little room for quality improvement by parallax adjustment.

We can also conclude that although the horizontal parallax adjustment has greatly improved the quality of 3-D content, very small  $d_{min}$  still leads to unsatisfactory quality (*i.e.*, the mean opinion score is below 50 on a 0–100 rating scale in all four subplots in Fig. 16), and, hence, needs to be avoided in the capturing process.

We computed the mean opinion scores of all 25 test videos that  $d_{min}$  equals  $d_{obj}$ , before and after horizontal parallax adjustment. The scores of all 25 video sequences indicate that viewers perceive horizontally adjusted videos to possess higher quality than non-adjusted ones, with an average quality-score gain of 19.86%.

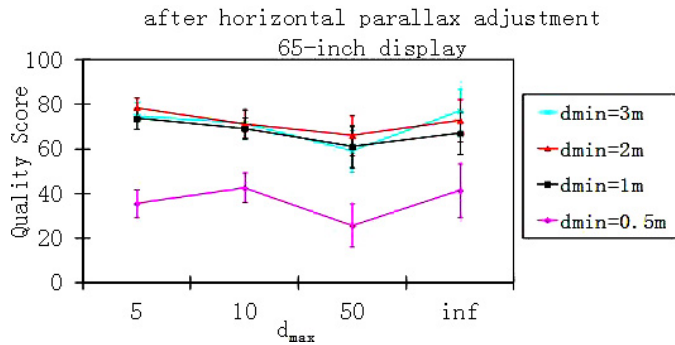


FIGURE 17 — Comparison of the mean opinion scores at  $d_{max}$  equal to 5 m, 10 m, 50 m, and infinity, with different  $d_{min}$ .

### (3) Influence of $d_{max}$ to video quality after horizontal parallax adjustment

Figure 17 shows the influence of  $d_{max}$  to the quality of 3-D after horizontal parallax adjustment. It is interesting to know that the shape of Fig. 17 is very similar to Fig. 13, except that the quality level of each curve is raised.

## 5 Conclusion and future work

We conducted comprehensive subjective tests to determine the influence of a few capturing parameters (*i.e.*, lighting conditions,  $d_{obj}$ ,  $d_{max}$ , and  $d_{obj}$  not being the foreground object) to the quality of 3-D images, quality of 3-D videos, and quality of 3-D videos after horizontal parallax adjustment when viewed on 3-D TVs and 3-D mobile devices. The influences of these parameters are consistent over different sizes of displays and over images and videos before and after horizontal parallax adjustment. Parameter  $d_{min}$  is the main factor affecting the 3-D quality. Having  $d_{obj}$  not being the foreground object slightly degrades the 3-D quality, whereas the lighting conditions and  $d_{max}$  do not severely affect the 3-D quality. The automatic horizontal parallax adjustment algorithm that we implemented has greatly improved the 3-D quality of experience by 19.86%. Despite such quality improvement, very small  $d_{min}$  still leads to poor quality and hence needs to be avoided in the capturing process.

Our future work includes developing a smart parallax adjustment method which incorporates a Visual Attention Model (VAM) tailored for stereoscopic 3-D videos. This VAM will use information from the 3-D content such as motion, texture, color, and depth. This information will be employed to adjust the parallax so that the important 3-D content is within the comfort zone. We will also take into consideration the sizes and aspect ratios of 3-D displays when developing the algorithm. Although it is now common to use short video sequences for subjective quality tests, we plan to use longer video sequences. By doing this, we expect that viewers will be able to provide more useful information about the quality of their experience, especially regarding visual fatigue.

## Acknowledgment

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

- 1 P. Merkle *et al.*, “3-D video: acquisition, coding, and display,” *IEEE Trans. Consumer Electron.* **56**, No. 2, 946–950 (May 2010).
- 2 G. P. Herbig, “The three golden rules of stereography (in German),” *Stereo J.* **65** (March 2002).
- 3 F. Zilly *et al.*, “The stereoscopic analyzer – An image-based assistance tool for stereo shooting and 3-D production,” *IEEE Intl. Conf. Image Processing (ICIP)*, 4029–4032, 26–29 (Sept. 2010).
- 4 B. Mendiburu, *3-D Movie Making: Stereoscopic Digital Cinema from Script to Screen* (Elsevier – Focal Press, 2009).
- 5 R. S. Allison, “The camera convergence problem revisited,” *Stereoscopic Displays and Virtual Reality Systems XI, Proc. SPIE* **5291**, 167–178 (2004).
- 6 A. J. Woods *et al.*, “Image distortions in stereoscopic video systems,” *Stereoscopic Displays and Applications IV, Proc. SPIE* **1915**, 36–48 (1993).
- 7 M. Lang *et al.*, “Nonlinear disparity mapping for stereoscopic 3-D,” in *ACM SIGGRAPH 2010*, Hugues Hoppe (ed.). ACM, New York, NY, USA, Article 75 (2010), pp. 10.
- 8 S. L. P. Yasakethu *et al.*, “Analyzing perceptual attributes of 3-D video,” *IEEE Trans. Consumer Electron.* **55**, No. 2, 864–872 (May 2009).
- 9 S. L. P. Yasakethu *et al.*, “Quality analysis for 3-D video using 2-D video quality models,” *IEEE Trans. Consumer Electron.* **54**, No. 4, 1969–1976 (November 2008).
- 10 L. D. Pockett and M. P. Salmimaa, “Methods for improving the quality of user created stereoscopic content,” *Stereoscopic Displays and Applications XIX, Proc. SPIE-IS&T Electronic Imaging*, SPIE **6803**, 680306 (2008).
- 11 G. Lutz *et al.*, “Impact of acquisition distortion on the quality of stereoscopic images,” *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, Arizona, USA, January 13–15, 2010.
- 12 J. Hakkinen *et al.*, “Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width,” *J. Soc. Info. Display* **17**, 433 (2009).
- 13 M. Polonen *et al.*, “Subjective measures of presence and discomfort in viewers of color-separation-based stereoscopic cinema,” *J. Soc. Info. Display* **17**, 459 (2009).
- 14 S. Marja *et al.*, “Effect of number of views to the viewing experience with autostereoscopic 3-D displays,” *J. Soc. Info. Display* **17**, No. 5, 449–458 (2009).
- 15 N. Bevan, “Classifying and selecting UX and usability measures,” *Proc. Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, 5th COST294-MAUSE Open Workshop, Reykjavik, Iceland, June 2008, pp. 13–18.
- 16 D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Intl. J. Computer Vision* **60**, No. 2, 91–110 (November 2004).
- 17 ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R, Tech. Rep. BT.500-11 (2002).
- 18 S. Boslaugh and P. A. Watters, *Statistics in a Nutshell* (O’Reilly Media, Inc., 2008), pp. 151–168.



**Di Xu** received her B.S. degree in electrical engineering from Beijing Normal University, Beijing, China, in 2003, and her M.A.Sc. degree in electrical and computer engineering from the University of Victoria, Victoria, BC, Canada, in 2007. She is currently working toward her Ph.D. degree in the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, BC, Canada. Her research interests are focused on the fields of digital image and video processing. These include 3-D quality of experience, high-dynamic-range image and video generation, video quality metric, digital video transcoding, and image coding.



**Lino E. Coria** received his B.S. degree in electronics engineering from the Instituto Tecnológico de Morelia, Morelia, Mich., Mexico, in 1996, his M.Eng. degree in electrical engineering from McMaster University, Hamilton, Ontario, Canada, in 1998, and his Ph.D. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada, in 2008. Currently, he is a Research Engineer at the Institute for Computing, Information and Cognitive Systems with the University of British Columbia. His interests include quality of experience, stereoscopic video, digital watermarking, and high-dynamic-range imaging. He was awarded two scholarships for his graduate studies (1996–1998 and 2003–2007) by the Mexican Council for Science and Technology (CONACYT). He is a member of the Mexican Researchers System (SNI), a member of ACM, and a senior member of IEEE.



**Panos Nasiopoulos** received his B.S. degree in physics from the Aristotle University of Thessaloniki (1980), Greece, and his B.S. (1985), M.S. (1988), and Ph.D. (1994) degrees in electrical and computer engineering from the University of British Columbia, Canada. He is presently the Director of the Institute for Computing, Information, and Cognitive Systems (160 faculty members and 800 graduate students) at the University of British Columbia (UBC). He is also a Professor with the UBC Department of Electrical and Computer Engineering, the Inaugural Holder of the Dolby Professorship in Digital Multimedia, and the current Director of the Master of Software Systems Program at UBC. Before joining UBC, he was the President of Daikin Comtec US (founder of DVD) and Executive Vice-President of Sonic Solutions. He is a registered professional engineer in British Columbia, the Associate Editor for *IEEE Signal Processing Letters*, and has been an active member of the Standards Council of Canada, the IEEE, and the Association for Computing Machinery (ACM).