HOW TO DISPLAY 3D CONTENT REALISTICALLY

René Klein Gunnewiek and Patrick Vandewalle

Philips Research Laboratories, The Netherlands

ABSTRACT

A rapidly growing number of movie productions are released in 3D. These productions use a stereo format, primarily intended for eye-wear assisted viewing and aiming at 3D movie theaters. At the same time the Blu-ray Disc Association is developing a standard for storing 3D content on Blu-ray disc. In the consumer environment a large variety of display types, sizes and viewing conditions, e.g. viewing distance, will exist. To see the 3D content realistically on such a large spread of displays and viewing conditions, indicative means for proper rendering of that content are required. These indicative means could for example be the intended display size and viewing distance or an entire depth map indicating for each pixel the distance to the viewer. In this paper we address, for a couple of typical scenarios, how to display the 3D content realistically and which indicative means are needed for that.

1. INTRODUCTION

The entertainment industry is currently getting ready for 3D applications based on stereo. The Blu-ray Disc Association is developing a standard for storing 3D content on Blu-ray disc. Similarly, broadcasters are launching the first 3D TV channels and consumer TV manufacturers are releasing their first stereoscopic 3D television sets. These sets use two views with glasses, typically using a time sequential or a line-interleaved format. While 3D movies already exist since the early 20th century, the interest in 3D has grown rapidly over recent years and many people believe this could be its real breakthrough.

However, capturing and rendering of stereoscopic content is a delicate matter. Various depth cues exist and the human brain merges them to create a depth impression. Multiple conflicting cues will cause headaches. Some of them relate to the current state of technology, such as cross-talk between the two channels in a stereoscopic set-up, or insufficient calibration between cameras in a stereoscopic camera rig. Others are more fundamental to viewing 3D on a flat screen. Among these conflicts are depth distortion, vergence-accommodation and frame violation. Stereoscopic content producers pay special attention to optimally configure the scenes such that conflicts are minimal. Therefore, it is preferable to reproduce the stereoscopic content as close as possible to the original content and viewing conditions. Since a large variety of displays and viewing conditions exists, meta data is needed to reproduce the stereoscopic content as intended. This meta data could be the intended display size and viewing distance, offering a limited flexibility. Full flexibility can be obtained by transmitting depth data as meta data. The use of depth information allows for generating 3D content realistically on a large variety of displays for a wide range of viewpoints.

2. RELATED WORK

The way in which humans observe the world around them differs from how cameras capture and televisions reproduce it. For example, when humans focus on a particular depth, the human eyes are slightly toed in. Since it is not known at which position a human viewer will focus, stereoscopic content is typically shot with two parallel cameras. Which image distortions will appear when a scene is shot with toed-in cameras is nicely depicted by Woods et al. [1]. The visibility of the distortions depends on the scene settings, the configuration of the cameras and the intended display. An overview of these distortions is given in the article by Meesters et al. [2]. Yamanoue describes the relation between the distortions and the camera settings during shooting [3]. The stereographic developer's handbook by Lipton gives guidelines on how to shoot stereoscopic content [4]. These guidelines pose constraints on the capturing and displaying devices. Given these constraints, the scene and the depth effect the cinematographer wants to obtain, he should map the perceived depth to the regions of interest [5]. However, the display size of consumer devices varies significantly. The effect of different screen sizes on the depth perception has been discussed by Kutka [6].

In this paper, it is shown that when stereoscopic content is displayed as is, irrespective of the display size and viewing distance, it will result in a violation of the perspective and depth relation. Henceforth, this relation is referred to as the content integrity. To overcome these issues, a display-size-dependent shift between the two images of the stereo input pair is proposed. The shift is calculated from the ratio of the different screen sizes. It is proven that this shift is

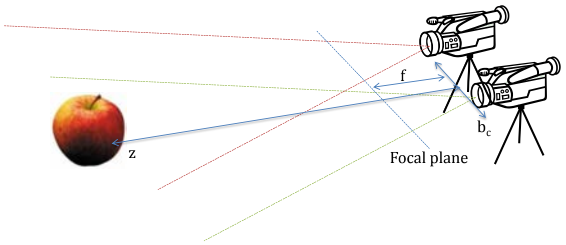

Fig. 1. Stereo capturing setup.

sufficient to reconstruct the correct 3D geometry. However, this restricts the viewpoint to just one viewing distance. For rendering from a new viewpoint, additional data is required. Typically depth data is used to generate a new view from a different viewpoint. A process called Depth-Image-Based Rendering (DIBR) is applied to generate a new viewpoint from depth data. This process in the context of 3D is nicely described by Fehn [7]. In this paper, we will present some scenarios to present stereoscopic content realistically over a wider range of conditions, tailored toward the 3D-Blu-ray disc case and we will show how depth can aid for this matter. Typically, computational resources in a Blu-ray disc player are limited and therefore algorithms with a small footprint are preferred.

3. CONTENT INTEGRITY

First, we will elaborate on the content integrity. Imagine an apple at a distance z is captured by two cameras that are a distance of b_c (baseline between cameras) separated from each other (see Figure 1). The focal length of the camera is given by f. We can then derive the disparity d_c between the two images of the object captured by the stereo camera pair from Figure 1 as:

$$d_c = \frac{b_c f}{z}. \tag{1}$$

It should be noted that these disparities have a metric unit. For a 3D production the stereo pair is given and with that the disparities in pixels. As such, the perceived (metric) distance z will depend on the horizontal display size and the distance of the viewer to the display (v_d). The amount of depth z seen by a viewer is given by

$$z = \frac{v_d}{1 - \frac{d_d}{b_e}}, \tag{2}$$

where b_e is the inter-pupil distance, or the baseline between the eyes, typically 65 millimeters and where d_d is the metric disparity on the display.

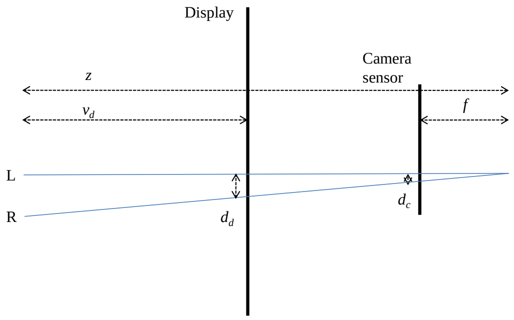

Fig. 2. A scaling factor is required when relating disparities in the camera focal plane to disparities in the display plane.

At the same time, the scene with the apple is captured by a camera, hence the scene is projected onto a plane. We can derive the projection of the scene onto a plane as follows: the (metric) width x_c in an image of an object with width X can be computed as

$$x_c = \frac{Xf}{z}. \tag{3}$$

For the sake of simplicity, we will restrict ourselves to the horizontal axis. The captured images are in turn rendered on a displaying device. The display is the projection plane and therefore we see an object with image width x_d as an object of size

$$X = \frac{x_d z}{v_d}. \tag{4}$$

Merging equations (3) and (4) results in the following condition:

$$v_d = \frac{x_d f}{x_c} = \frac{w_d f}{w_c}, \tag{5}$$

where w_c is the width of the camera sensor and w_d is the width of the display. In other words, the ratio between the focal length of the camera and the width of the sensor should be the same as the ratio between the viewing distance and the width of the display: the field of view of the scene during capturing should be the same as the field of view during rendering.

If we look at the stereo pair and its disparities, we can merge equations (1) and (2). However, we should apply a scaling factor to the disparities in order to relate the disparities in the sensor plane to disparities in the display plane (see Figure 2). The captured pixel disparities in an image on the sensor are scaled to pixels on the display, so the scaling factor is the ratio between the width of the sensor and the width of the display. As such, the normalized disparity $d_{c,d}$ becomes:

$$d_{c,d} = d_c \frac{w_d}{w_c} = \frac{b_c f}{z} \frac{w_d}{w_c}. \tag{6}$$

Merging equations (2) and (6) results in the following relation between the disparity of the camera and the disparity on the display:

$$d_d = \frac{b_e(z - v_d)}{z} = b_e - \frac{w_c v_d b_e}{w_d f b_c} d_{c,d}. \tag{7}$$

The first part of the fraction in equation (7) is equation (5). So, if we satisfy equation (5) and the baseline of the camera is equal to the baseline of the viewer ($b_c = b_e$), content integrity can be obtained by applying a fixed metric shift, equal to the inter-pupil distance b_e, to one of the images in the stereo pair.

Let us now assume we have stereoscopic content intended to be displayed on a reference display with size w_r. If we want to display this content on a display with actual size w_d, we need to apply an offset s to one of the two views given by

$$s = \frac{w_r}{w_d} d_d - d_r, \tag{8}$$

where d_r and d_d are the disparities on the reference display and the actual display, respectively. This can be rewritten using the first part of equation (7) and the equality in field of view ($v_d/v_r = w_d/w_r$) as

$$s = \frac{w_r}{w_d} \frac{b_e(z - v_d)}{z} - \frac{b_e(z - v_d \frac{w_r}{w_d})}{z} = b_e \left(\frac{w_r}{w_d} - 1 \right). \tag{9}$$

3.1. Floating Windows

When an object at the border of the screen is shown at a depth in front of the screen, this causes a window violation: although the object appears to be in front of the screen (and should therefore be visible to both eyes, occluding the screen edge in one view), it is visible with only one eye (in one of the two views). To prevent this issue, content manufacturers place black bars over the content that appears in front of the screen and is just visible in one view. Typically, this results in a black bar at the left side for the left view and a black bar to the right of the right view. An object appearing in front of this (virtual) border can now be shown in both views. It will be displayed over the virtual border in one of the views.

In Figure 3 we schematically depict the width of the floating window that is required to prevent window violation for an object that appears at depth z using two displays: a reference and the actual display.

Our method described above for correct viewing implies a shift to the stereoscopic content. We will now analyze how that affects these side bars. The required width of the black bar is defined by:

$$F_d = \frac{v_d b_e}{z} - b_e. \tag{10}$$

In order to relate the offsets for both cases, a scaling factor is required, since the display size scales as well. As can be seen from equation (5), the ratio between display sizes is equal to the ratio between the respective viewing distances. We can then compute the difference in floating window between the reference and the (scaled) actual display as

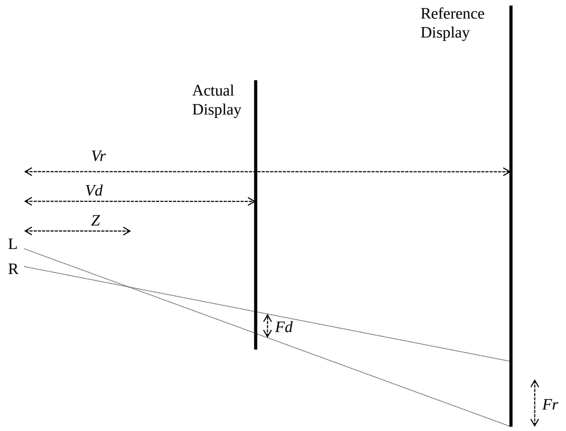

Fig. 3. Schematic illustration of floating windows used to prevent window violations.

$$\begin{aligned} F_r - \frac{v_r}{v_d} F_d &= \frac{v_r b_e}{z} - b_e - \frac{v_r}{v_d} \frac{v_d b_e}{z} + \frac{v_r}{v_d} b_e \\ &= b_e \left(\frac{v_r}{v_d} - 1 \right) \\ &= b_e \left(\frac{w_r}{w_d} - 1 \right) \\ &= s. \end{aligned} \tag{11}$$

In order to keep the floating windows at the position where they were, and therefore not introducing any additional window violations, an offset is required that is identical to the offset s computed in equation (8). In other words, the proposed shifting method does not introduce any additional window violation. It keeps the floating windows at the distance where they were, until the distance to the actual display is smaller than the distance to the floating window. In that case floating windows will disappear, which is natural, as objects will then disappear behind the actual display as well.

3.2. Analysis

The above described scenario and the proposed shifts can intuitively be depicted in the graphs below. Figure 4 illustrates the relation between the disparity and the perceived depth. It nicely shows the $1/x$ relation between disparity and perceived depth. In this case, the reference viewing distance for the stereoscopic content is 6 meters.

In Figure 5, the depth disparity relations are shown for the case that this content is displayed on a consumer display

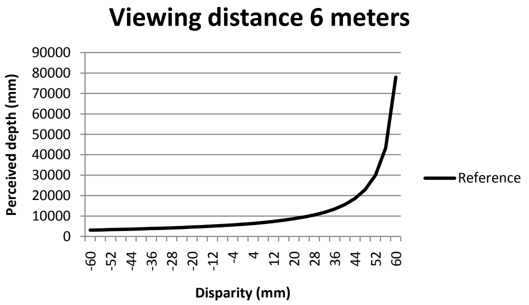

Fig. 4. Perceived depth as a function of the amount of disparity for a reference display at a viewing distance of 6 m.

that has half the horizontal size, as is (no shift) and with an offset s applied as described above (shift). This shows that by applying this shift, the same curve can be obtained as depicted in Figure 4. Thus, the depth perception remains the same for both cases.

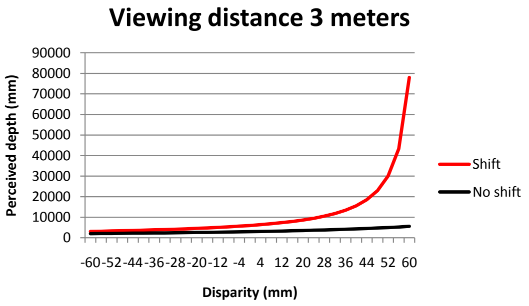

Fig. 5. Perceived depth as a function of the amount of disparity for a display half the size of the reference display, viewed at a distance of 3m, with and without shift.

4. ISSUES WITH CONTENT INTEGRITY

In a practical scenario, the stereoscopic content will be shot and distributed having a particular display size and viewing conditions in mind. As a result, a shift (which can even be scene dependent) is applied to the stereoscopic pair. We propose to transmit the intended or reference display size w_r and preferred viewing distance v_r along with the stereoscopic content, such that the optimal settings, obeying the content integrity, can be applied. The optimal viewing configuration can then be obtained by applying equations (5) and (9).

There are, however, perceptual issues with the above presented equations. The equations above imply that if cinema content is shown on a consumer display, almost all content will be positioned behind the screen. This could result in eye fatigue due to the vergence-accommodation conflict: the eyes focus at the display depth, while vergence is at the perceived depth behind the screen [8]. This is especially a problem for somewhat smaller displays, since, according to equation (5), the viewing distance should be relatively small (close to the display). This could be (partially) overcome by putting a small positive lens (e.g. +0.5 diopter) in the glasses needed to watch the stereoscopic content as also suggested by Hoffman et al. [8]. Due to this small positive lens, it seems for your eyes that the display is a few meters further to the back. This reduces eye fatigue significantly, yet the distortion due to the extra lens is marginal, since the strength of the positive lens is negligible with respect to the strength of the lens from the eye, which is about 40 diopter.

A second perceptual issue with the proposed shift is that a fixed viewing distance is not always desirable from a consumer point of view. The constellation of the living room does not always allow for a seat at that position or the viewer does not want to sit so close to the screen. Therefore, we have tried to come up with a solution that approximates the content integrity as much as possible. The content is distributed typically with the plane of interest around screen depth for the intended display size. Therefore, we try to keep the content integrity correct around this depth. Suppose the new viewing position in front of the screen with size w_d is v_n, whereas the viewing distance should be v_d according to equation (5). This will change the actual projection x_n of an object with size X at distance z as follows:

$$x_n = \frac{X v_n}{z + v_n - v_d}. \tag{12}$$

The field of view changes accordingly.

The ratio between what the projection actually is at the new viewpoint and what the projection should be, is given by

$$\alpha = \frac{x_n}{x_d} = \frac{v_n}{z + v_n - v_d}\frac{z}{v_d}. \tag{13}$$

The amount of depth observed by a viewer is given by equation (2) for the reference setting:

$$z_r = \frac{b_e v_r}{b_e - d_r}, \tag{14}$$

where d_r is the disparity in the reference view. For the new viewpoint, we will apply an offset δ to the stereoscopic content to display it as realistically as possible. We can write this as

$$z_n = \frac{b_e v_n}{b_e + \delta - \frac{w_d}{w_r} d_r}. \tag{15}$$

We will now derive the optimal offset δ. Since we cannot keep the content integrity for the entire depth range at this new viewpoint, we will keep the content integrity for a particular depth position. In order to preserve the content integrity around a particular depth position, we want to keep the depth relation around that depth the same. To achieve that, we have to keep the depth gradient the same as for the original content:

$$\begin{aligned} z'_r &= -\frac{b_e v_r}{(b_e - d_r)^2} \\ z'_n &= -\frac{b_e v_n \frac{w_d}{w_r}}{\left(b_e + \delta - \frac{w_d}{w_r} d_r\right)^2}, \end{aligned} \tag{16}$$

with z'_r and z'_n indicating the derivative of z_r and z_n, respectively. The depth values of the stereoscopic content are typically distributed around screen depth of the intended display, hence at the viewing distance. Therefore we try to preserve content integrity around the reference screen depth v_r of the intended display size; hence the disparity is zero at that depth for the original content.

In this case, the projection distortion becomes

$$\alpha_0 = \frac{v_n}{v_r + v_n - v_d} \frac{v_r}{v_d}, \tag{17}$$

and the derivatives

$$\begin{aligned} z'_{r,0} &= -\frac{b_e v_r}{b_e^2} \\ z'_{n,0} &= -\frac{b_e v_n \frac{w_d}{w_r}}{(b_e + \delta)^2}. \end{aligned} \tag{18}$$

For content integrity, the derivative of the amount of depth observed at screen depth for the original sequence should be equal to the derivative of the depth at that position for the new viewpoint, scaled with the ratio in projection distortion:

$$z'_{r,0} = \alpha_0 z'_{n,0}. \tag{19}$$

Using the above equations (17) and (18), we obtain

$$\frac{b_e v_r}{b_e^2} = \frac{v_n}{v_r + v_n - v_d} \frac{v_r}{v_d} \frac{b_e v_n \frac{w_d}{w_r}}{(b_e + \delta)^2} \tag{20}$$

or

$$\frac{(b_e + \delta)^2}{b_e^2} = \frac{v_n}{v_r + v_n - v_d} \frac{v_n \frac{w_d}{w_r}}{v_d}. \tag{21}$$

This results in the following offset:

$$\delta = b_e \left(\frac{w_d}{w_r} \beta - 1 \right) \text{ with } \beta = \frac{v_n}{v_d} \sqrt{\frac{v_r}{v_r + v_d - v_n}}. \tag{22}$$

When the new viewing distance is equal to the intended distance, the parameter $\beta = 1$ and equation (22) reduces to equation (8). Equation (22) implies that we apply a (small) extra offset when the viewer is closer than the preferred viewing distance and, likewise, a somewhat smaller offset when the viewer is further away than the preferred distance.

This approach works for a small range of depth values and for a limited deviation from the preferred viewing point. For larger deviations, new viewpoint rendering should be applied, for which a depth map is typically required.

5. CONTENT INTEGRITY WITH DEPTH MAPS

Having a depth map offers flexibility. For example, it allows variation in the baseline for the stereo pair, it makes high quality rendering possible on other than stereoscopic displays, it gives full control of the window violation issue and it facilitates new viewpoint rendering. The latter option allows the viewer to have the stereoscopic content rendered around screen depth for any display size without sacrificing content integrity. By modifying the baseline of existing stereo content, it can be adapted to youngsters (having a smaller inter-pupil distance) and to the preference of the user. The process of rendering a new view by means of a depth map is referred to as Depth-Image-Based Rendering (DIBR) and comes at various complexity levels. Adaptation of the stereo baseline of the stereoscopic pair is rather simple. As can be seen from equation (3), the projection on the plane remains the same, hence only depth-dependent horizontal pixel shifts are required. Such a rendering process has a relatively small complexity, as depicted by Berretty and Ernst [9]. Rendering a new viewpoint closer to or further from the scene along the line of sight does require a depth dependent re-projection of the pixels. The new pixel coordinates (x', y') are determined by:

$$\begin{aligned} x' &= \frac{xz}{z - \Delta} \\ y' &= \frac{yz}{z - \Delta}, \end{aligned} \tag{23}$$

where Δ is the absolute distance the viewpoint becomes closer to the scene, keeping the field of view (focal length) the same. This process has also a rather small complexity. However, for rendering a new viewpoint away from the line of sight, the entire DIBR process should be applied. This process consists of the following two steps: first, the depth data is used to render the texture data into a 3D space. Then a new viewpoint is chosen and from that viewpoint the 3D space is projected on a plane with respect to the new viewpoint. For more details on DIBR, see for example the work by Fehn [7].

6. RESULTS

The above described scenarios have been tested on a variety of displays and for a variety of stereoscopic content. Experiments have shown that the liveliness of a sequence increases with the proposed additional shift for screen size compensation. If no shift is applied, it seems, especially on smaller displays, that the depth at infinity is about 50 centimeters behind the screen rather than being at infinity. Furthermore, the aspect ratio seems distorted, which is best observed on humans. After applying the proposed shift, the scene becomes more natural. As an illustration, we have included anaglyph snapshots of Pinocchio. In Figure 6a we show the

(a) Without size compensation.

(b) With size compensation.

Fig. 6. Pinnocchio snapshot with and without compensation for display size.

sequence as is, whereas in Figure 6b we show the sequence on which screen size correction is applied[1]. After screen size compensation, the scene looks more natural, creating a more immersive experience.

7. CONCLUSIONS

Within the Blu-ray Disc Association, discussions are ongoing to standardize a format for storing 3D stereoscopic content on a Blu-ray disc. In this paper we have presented methods to display this stereoscopic content realistically with some additional data. By shifting the left and the right view with respect to each other, content integrity can be maintained for a particular viewing distance. For other viewing distances, methods to render this stereoscopic content optimally by applying a shift to the stereo pair have been presented as well. Finally, full flexibility with respect to a preferred baseline, viewpoint and viewing conditions can be achieved when a depth map is available, at the cost of a depth based rendering process. Perceptual studies to measure the amount of discomfort from viewing content on different display sizes and at varying distances, as well as the effect of the proposed compensations would be interesting topics for further research.

8. ACKNOWLEDGEMENTS

We would like to thank our 3D colleagues within Philips for providing a stimulating collaborative working environment and for their input to this paper.

9. REFERENCES

[1] A. Woods, T. Docherty, and R. Koch, "Image distortion in stereoscopic video systems," in *Proc. SPIE Stereoscopic Displays and Applications IV*, 1993, vol. 1915.

[2] L. Meesters, W. A. IJsselsteijn, and P. Seuntiëns, "Survey of perceptual quality issues in three-dimensional television systems," in *Proc. SPIE Stereoscopic Displays and Applications XIV*, 2003, vol. 5006.

[3] H. Yamanoue, "The relation between size distortion and shooting conditions for stereoscopic images," *Journal of the SMPTE*, 1997.

[4] L. Lipton, "The stereographics developer's handbook," http://www.cs.unc.edu/Research/stc/FAQs/Stereo/stereo-handbook.pdf, 1997.

[5] N. Holliman, "Mapping perceived depth to regions of interest in stereoscopic images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291.

[6] R. Kutka, "Reconstruction of correct 3-D perception on screens viewed at different distances," *IEEE Transactions on Communication*, vol. 42, no. 1, 1994.

[7] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)," in *Proc. of VIIP 03*, 2003.

[8] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 1–30, 2008, Article 33.

[9] R.-P. Berretty and F. Ernst, "High quality images from 2.5d video," in *Short Presentations, Eurographics*, 2003, pp. 255–262.

---

[1] These snapshots are typically not well preserved on paper due to color mismatches; the intended width of these images is about 15 inches. Therefore, these images should be viewed on a computer monitor. Computer monitors generally represent the colors better, although there is also a large spread in color performance between displays. It can be beneficial to try to watch the content on various displays. Finally, remark that since we apply a shift that places the shot usually further behind the screen, good separation of the left and right view becomes more important.