3DTV: one stream for different screens

Keeping perceived scene proportions by adjusting camera parameters

Andrey Norkin and Ivana Girdzijauskas
Multimedia Technologies, Ericsson research
Ericsson
Stockholm, Sweden
{andrey.norkin, ivana.girdzijauskas}@ericsson.com

***Abstract*—Stereo and 3D video material is usually optimized during the production phase for a particular display size and viewing distance. When the content is shown on a display of different size and/or from different viewing distance, the perceived proportions of objects, e.g. object's depth relative to object's size, will be distorted compared to the original viewing conditions. This can make the scene look unnatural and even lead to eye strain and fatigue when observing the content. This paper proposes equations for adapting rendering parameters to new viewing conditions so that the perceived proportions of the objects in 3D scene are exactly the same as for the reference viewing conditions.**

Keywords – 3D video, stereo, multi-view, depth perception, camera and display geometry, display size, viewing conditions, video rendering, view synthesis

I. INTRODUCTION

The area of 3D video and 3DTV has gained momentum in recent years and is now considered the next step in consumer electronics, mobile devices, computers and movie industry. The most common 3D/stereo technology nowadays is to transmit two views corresponding to the left and right eyes of the viewer. The separation of views is usually achieved by means of shutter- or polarized glasses. Moreover, autostereoscopic screens, that do not need glasses, are becoming more and more common. In order to achieve smooth parallax and change of the viewpoint when the user moves in front of the auto-stereoscopic display, a number of views (typically 7-28) are generated. In order to facilitate the use of recent 3D technologies, MPEG has recently issued a Call for Proposals (CfP) on 3D video coding technology [2] and started standardization that should enable both advanced stereoscopic display processing and improved support for auto-stereoscopic multi-view displays. It is suggested to transmit a number of views (usually two or three) that will be shown on stereoscopic or multi-view displays and/or used for generation of intermediate views.

For rendering multiple views at the receiving side additional information may be needed. Typically a depth map, indicating the distance between the objects in the scene and the capturing camera is used for this purpose. Depth maps are represented by gray-scale images and sent together with the 3D/stereo views. At the receiving side, the additional views are then rendered (synthesized) from the texture and depth maps (the intermediate views can also be generated without explicitly sending depth maps to the receiver). One of the popular techniques is depth-image-based rendering (DIBR) [1]. In addition to above-mentioned texture and depth maps, DIBR uses additional parameters such as z_{near} and z_{far} (the closest and the farthest depth values in a depth map), *intrinsic* camera parameters (such as focal length, coordinates of the images principal point and radial distortion) and *extrinsic* camera parameters (camera position and the direction of its optical axis in the chosen real world coordinates).

During the capturing phase of stereo/3D video production, the scene and camera parameters such as camera separation, distance to the scene etc. are carefully chosen in order to provide the best user experience and avoid eye strain when watching the stereo content. Later, during the production/post-processing phase, other parameters such as the sensor (image) shift are adjusted. Capturing and post-processing parameters are usually set up having in mind a target screen size and a viewing setup.

However, when the content is displayed on a screen that has a size different than the reference screen or the actual viewing distance is different from the reference distance, distortions of the perceived depth would appear relative to what has been planned during the production [3]. It has been shown [4][5] that distortion can only partially be compensated by shifting the images with respect to each other, for example, around certain depth. A method that allows keeping the same ratio between the depth in front of and behind the screen was proposed in [7]. Some of these works provide guidelines for 3D-capturing process in order to show the 3D content realistically and achieve visual comfort when displaying it on the target setup [3][5][6].

This paper does not consider the issues of capturing the 3D content, choosing the best camera parameters for the visual comfort and eliminating the depth distortion relative to the real world scene. The idea is rather to present the 3D content in varying displays setups as close as possible to what the content provider intended to present at the target setup (including varying scene conditions and artistic effects). In order to achieve this goal, this paper presents equations that can be applied at the receiver side in order to find rendering parameters for DIBR or another view synthesis method for the particular viewing conditions at the receiver. When performing view synthesis with the found parameters, the same depth perception for a large variety of screen sizes and viewing distances can be achieved. Namely, it is possible to keep the *perceived* dimensions of the objects in the scene in exactly the same proportion to the screen size as in the target setup. Therefore, the same 3D/stereo video sequence can be used on a large variety of displays and viewing setups while keeping the proportions of the objects in the scene the same as in the setup targeted by the content provider.

The paper is organized as follows. Section II gives a brief overview of camera and display geometry necessary for understanding the rest of the paper. Section III shows how the relative depth impression can be kept constant by changing the camera baseline its z-coordinate. Section IV analyzes special cases followed from derived equations and a numerical example is provided in Section V. The application scenarios are discussed in Section VI. Finally, Section VII concludes the paper.

II. CAMERA AND DISPLAY GEOMETRY

A. Geometry of stereo capturing

Figure 1 shows the so-called parallel sensor-shifted setup, where the convergence plane of cameras is set by the sensor shift h. Let f be the camera focal length, t_c the baseline distance and Z_c the distance to the convergence plane. Suppose the captured object is located at distance (depth) Z from the cameras. The distance between the points in two images that refer to the same captured point is called disparity, denoted by d. Following notation from Figure 1, the following expression for disparity can be derived:

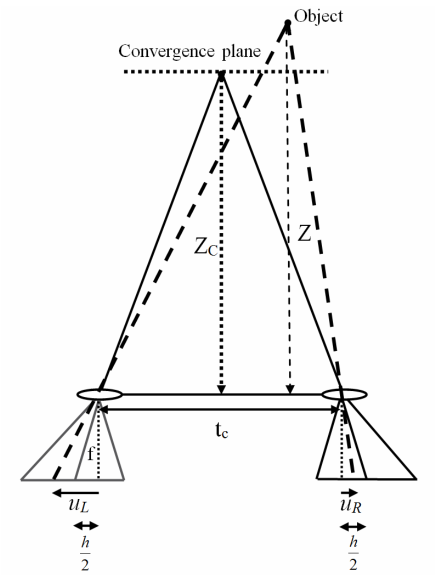

Figure 1 Capturing setup

$$d = h - (u_L + u_R) = h - \frac{t_c f}{Z} = t_c f (\frac{1}{Z_c} - \frac{1}{Z}) \quad (1)$$

We can also see that the objects captured at Z=Z_c have zero disparity, which further gives us:

$$h = \frac{t_c f}{Z_c} \quad (2)$$

B. Geometry of 3D displays

Suppose that the distance between the viewer's eyes is t_e (the so-called inter-ocular distance) and that the viewer sits at the distance Z_D from the screen (so-called viewing distance), as shown in Figure 2. A simple geometry gives us the following expression for the perceived depth:

$$Z_p = \frac{Z_D \cdot t_e}{t_e - P} \quad (3)$$

Here P is the screen parallax, which basically reflects the spatial distance between the points in the left and the right view on the screen. Define the perceived depth *relative to the display* as:

$$Z_{p\_rel} = Z_p - Z_D = \frac{Z_D \cdot P}{t_e - P} \quad (4)$$

Then $Z_{p\_rel} < 0$ means that the object is perceived in front of the screen, while objects behind the screen have $Z_{p\_rel} > 0$. Finally, an object is perceived on the screen if $Z_{p\_rel} = 0$.

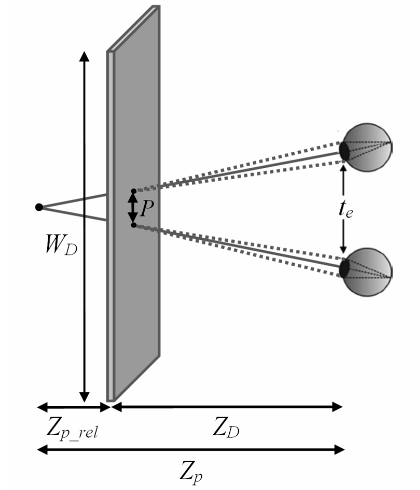

Figure 2 Display setup

C. Combining capturing and display geometry

It is not difficult to see analogy between the capturing and the display side. For example, disparity d on the capturing side relates to screen parallax P on the display side. The so-called magnification factor S_M is defined accordingly as:

$$S_M = \frac{P}{d} = \frac{W_D}{W_S} \quad (5)$$

where W_D is the display width and W_S is the sensor width.

III. MAINTAINING RELATIVE PERCEIVED DEPTH FOR DIFFERENT SCREEN SIZES AND VIEWING DISTANCES

A. Problem formulation

In order to maintain the same (or similar) viewing experience for the users having displays of different sizes and watching them from different distances, it is important to keep the relative perceived depth of the objects Z_{p_rel} proportional to the horizontal and vertical screen size. That means that if the screen width is scaled with factor b, the relative perceived depth of the scene should be scaled with the same factor *b* in order to maintain the same width/depth relation of the object in the video scene. These proportions should be maintained at any viewing distance Z_D.

This problem can be formulated as follows: *Given the reference distance from the display Z_{D1}, the reference display width W_{D1}, the actual distance from the display $Z_{D2} = aZ_{D1}$ and the actual display width $W_{D2}=bW_{D1}$, what are the rendering parameters such that the relative perceived depth scales proportionally to the screen width, i.e., $Z_{p\_rel2} = bZ_{p\_rel1}$, for an arbitrary choice of parameters a and b?*

In order to accommodate for changes of the screen width and the viewing distance, we allow changing the baseline distance and virtual cameras shift over *z* coordinate. Changing *z* coordinate of the cameras would therefore change Z_c and Z. In order to account for these changes, denote $Z_{c2} = \alpha\, Z_{c1}$ and $t_{c2} = \beta\, t_{c1}$. In the following, we derive parameters α and β such that the relation $Z_{p\_rel2} = bZ_{p\_rel1}$ holds for all Z_{p_rel1}.

B. Formula derivation

Since we would like to keep the same ratio between the screen width and the perceived depth relative to the display position, the following equality should hold:

$$\frac{Z_{p\_rel2}}{W_{D2}} = \frac{Z_{p\_rel1}}{W_{D1}} \tag{6}$$

Thus, by using (4) and combining it with (6) we get:

$$\frac{a}{P_1} - \frac{b}{P_2} = \frac{1}{t_e}(a-b) \tag{7}$$

When changing Z_c, one should also change the focal distance of the camera in order to avoid scaling of the objects size. Images of the objects located at the convergence distance should have the same size relative to the sensor width and to the screen size (in other words one should keep the same "virtual screen" in the camera space). This requires changing focal length with the same scaling factor as the convergence distance, i.e. $f_2 = \alpha f_1$.

Let $Z_r = Z_1 - Z_{c1}$ be the relative depth in the camera space. Similarly, it can be written $Z_2 = \alpha Z_{c1} + Z_r$. When substituting the relative depth values into (1), the following expressions for disparities are obtained:

$$d_1 = t_{c1} f_1 \left(\frac{1}{Z_{c1}} - \frac{1}{Z_{c1} + Z_r}\right) \tag{8}$$

$$d_2 = \beta t_{c1}\, \alpha\, f_1 \left(\frac{1}{\alpha Z_{c1}} - \frac{1}{\alpha Z_{c1} + Z_r}\right) \tag{9}$$

Having in mind (5) and after substituting (8) and (9) into (7), the following expression can be obtained:

$$\left(a - \frac{\alpha}{\beta}\right) Z_{c1}^2 + \left(a - \frac{1}{\beta}\right) Z_{c1} Z_r = Z_r \frac{W_{D1} t_{c1} f_1}{t_e W_S}(a-b) \tag{10}$$

Both sides of (10) are linear functions of Z_r. In order for (10) to hold for all relative depth values Z_r, which can take any values in the range $(Z_{near} - Z_{c1},\ Z_{far} - Z_{c1})$, it is necessary that the coefficients of linear functions at the left and the right sides of the equations are equal, which means:

$$\begin{cases} \left(a - \dfrac{\alpha}{\beta}\right) Z_{c1}^2 = 0 \\ \left(a - \dfrac{1}{\beta}\right) Z_{c1} = \dfrac{W_{D1} t_{c1} f_1}{t_e W_S}(a-b) \end{cases} \tag{11}$$

After solving (11), the following scaling factors α and β are obtained for Z_c and t_c respectively.

$$\alpha = \frac{1}{1 - \dfrac{W_{D1} t_{c1} f_1}{W_s t_e Z_{c1}}(1 - \dfrac{b}{a})} = \frac{1}{1 - S_{M1}\dfrac{h_1}{t_e}(1 - \dfrac{b}{a})} \tag{12}$$

$$\beta = \frac{\alpha}{a} = \frac{1}{a} \frac{1}{\left[1 - S_{M1}\dfrac{h_1}{t_e}(1 - \dfrac{b}{a})\right]} \tag{13}$$

Finally, the offset to be applied to virtual cameras' *z*-coordinate is obtained as:

$$Z_{offset} = Z_2 - Z_1 = (\alpha - 1) Z_{c1} = \frac{(\alpha - 1) t_{c1} f_1}{h_1} \tag{14}$$

and the new sensor shift as:

$$h_2 = \frac{t_{c2} f_2}{Z_{c2}} = \beta h_1 \tag{15}$$

IV. DISCUSSION ON SPECIAL CASES

So far, we have assumed a general case where the screen size and viewing distance are scaled with different factors. In the following some special cases are analyzed.

A. Viewing distance changes proportionally to screen size

The first special case is when the viewing distance and the screen size change with the same factor, i.e. $a = b$. This is an important case since the optimal viewing distance is often equal to one of the screen dimensions multiplied with a constant factor (at least, for the same class of display resolutions) [6]. Moreover, if the information about the viewing distance is not provided, the default viewing distance can be implied, which corresponds to the same scaling of the viewing distance as the screen width. It now follows from (12) and (13) that:

$$\alpha = 1; \quad \beta = \frac{1}{a}; \quad h_2 = \frac{h_1}{a}. \tag{16}$$

This means that the virtual cameras should stay at the same distance from the scene (virtual screen). Therefore, their Z coordinates remain the same. The baseline will change with the factor inversely proportional to the screen scaling factor; the same applies to the sensor shift.

B. *Only viewing distance changes*

Another important special case is when the viewing distance changes but the screen size stays the same. This happens for example when an observer is moving towards or away from the screen. The equations above suggest that, in this case, both the convergence distance and the camera baseline should change in order to keep the proportions of the object constant.

C. *Only screen size changes*

In this case only the screen size changes, but the viewing distance remains the same. For example, users can watch the content on mobile devices with stereoscopic screens of different sizes, while observing the screen from approximately the same distance. In this case, the derived equations give $\beta = \alpha$, which means that the baseline distance should change with the same factor as the convergence distance (the distance from camera to the "virtual screen").

The obtained results indicate that when the ratio between the screen width and the viewing distance is kept constant it is possible to keep the same geometrical relations of the 3D video scene by changing the baseline distance and sensor shift. This work also confirms that a simple shift of left and right images cannot accommodate for changes of the screen size and the viewing distance. However, if the change in convergence distance, baseline, focal distance and sensor shift is allowed, it is possible to maintain the exact proportions of the objects in the scene over a wide range of display sizes and viewing distances.

V. NUMERICAL EXAMPLE

Table 1 shows how the camera parameters should change when changing the display size and viewing distance. The inter-ocular distance is assumed equal to 6 cm. The example is provided for the *Kendo* sequence having the following parameters W_s = 1024, f = 2241.25607, h = 12, t_c = 5, Z_c = 933.8567. Suppose, that the sequence is optimized for the screen width W_D = 40 cm and the viewing distance Z_d = 90 cm.

Table 1. Sequence Kendo. Changing rendering parameters after change of Z_D and W_D

Z_D (cm)	W_D (cm)	α (Z_c, f)	β (t_c, h)	Z_{offset}
90	*40*	*1*	*1*	*0*
60	40	0.9624	1.4436	–35.1074
180	40	1.0407	0.5203	37.9617
90	80	0.9275	0.9275	–67.6708
90	20	1.0407	1.0407	37.9617
180	80	1	0.5	0
45	20	1	2	0
120	30	1.0354	0.7765	33.0485
60	50	0.9360	1.4040	–59.7532

One can see examples of special cases in Table 1. When Z_D and W_D change proportionally, only baseline distance and sensor shift are modified. When only the display size changes, $\alpha = \beta$.

VI. APPLICATIONS OF APPROACH

The proposed algorithm can be used for both stereoscopic and auto-stereoscopic displays. For stereoscopic displays, two views can be rendered (or one view rendered and one transmitted directly) using parameters obtained from equations (12)–(15). In case of multi-view auto-stereoscopic screens, knowing the viewing distance it is easy to find which views are seen by observer's eyes assuming a normal (average) inter-ocular distance. Then the baseline distance and virtual camera positions can be derived using the proposed formulas, and the other virtual camera positions can be found accordingly.

With the proposed equations it is possible to obtain the same proportions of the objects in 3D-scene on a large variety of screen sizes and viewing setups from the same stereo/3D stream. It is only needed to transmit the following reference parameters (W_{D_ref}, Z_{D_ref}, and t_e) in addition to reference camera parameters, derive rendering parameters for the actual viewing setup using equations (12)–(15), and perform view synthesis with new camera parameters. Therefore, with the proposed approach it is possible to use the same coded stereo/3D stream on a large range of display sizes and viewing distances.

VII. CONCLUSIONS

This paper proposes equations for deriving such rendering parameters for new viewing conditions that the perceived proportions of the objects in a 3D scene are exactly the same as for the reference viewing conditions. The adjusted parameters are derived from camera parameters and the reference parameters sent to the receiver. This makes it possible to avoid distortion in perceived depth relative to the target setup that would occur if the same 3D content was shown on different screens regardless of the change in viewing conditions. Consequently, the same coded content can be rendered for a large variety of different 3D/stereo displays while the eye strain and distortion of depth are avoided.

REFERENCES

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

[2] "Call for Proposals on 3D Video Coding Technology", ISO/IEC/JTC1/SC29/WG11, MPEG2011/W12036, Geneva, Switzerland, March 2011.

[3] F. Zilly, J. Kluger and P. Kauff: "Production rules for stereo acquisition", *Proceedings of the IEEE*, Vol. 99, No. 4, pp. 590–606, April 2011

[4] L. Chauvier, K.Murray, S. Parnall, R. Taylor, and J.Walker, "Does size matter? The impact of screen size on stereoscopic 3DTV," *IBC Conf.*, Amsterdam, Netherlands, Sep. 9–14, 2010.

[5] R. Klein Gunnewiek and P. Vandewalle: "How to display 3D content realistically", *Proceedings of Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, Arizona, U.S.A., Jan. 2010.

[6] W.J. Tam, F. Speranza, S. Yano, K. Shimono, H. Ono, "Stereoscopic 3D-TV: Visual Comfort", in *IEEE Transactions on Broadcasting*, Vol. 57, No. 2, pp. 335–346, June 2011.

[7] G. Jones, D. Lee, N. Holliman, D. Ezra, "Controlling perceived depth in stereoscopic Images", *Proceedings of SPIE: Stereoscopic displays and applications XII:2001,* pp. 24–53, 2011.