Optimal Pixel Aspect Ratio for Stereoscopic 3D Displays under Practical Viewing Conditions

Hossein Azari, Irene Cheng, and Anup Basu

Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada
EMAIL: {hazari, lin, anup}@cs.ualberta.ca

ABSTRACT

In multiview 3DTVs, the original 3D-scene is reconstructed based on the corresponding pixels of the adjacent 2D views. For conventional 2D display the highest image quality is usually achieved by uniform distribution of pixels. However, recent studies on the 3D reconstruction process show that for a given total resolution, a non-uniform horizontally-finer resolution yields better visual experience on 3D displays. Unfortunately, none of these studies explicitly model practical viewing conditions, such as the role of the 3D display as a medium and behavior of the human eyes. In this paper the previous models are extended by incorporating these factors into the optimization process. Based on this extended formulation the optimal ratios are calculated for a few typical viewing configurations. Some supporting subjective studies are presented as well.

1. INTRODUCTION

The output of a multiview 3D display is constructed based on two or more different views of the scene. More specifically, the 3D points of the reconstructed virtual 3D scene are formed based on the corresponding points of the adjacent 2D views. Instead of real-valued continuous 2D projections, discretized pixels are involved in this process. Thus, an original 3D point location is estimated with some error. In the context of stereo vision, this error is usually known as *discretization error* [2], [4],[5]. In 3D displays, this notion is closely related to the concept of 3D resolution or *stereoscopic resolution* which is defined as the precision of discriminating 3D-point locations in a comfortable viewing range of the 3D display [6]. The discretization error (or stereoscopic resolution) on each 3D coordinate component is related to the precision of discretization across the horizontal and vertical axes of the 3D display. Therefore, for a given total resolution, it is reasonable to look for an optimal horizontal *vs.* vertical discretization, or equivalently an optimal pixel aspect ratio (PAR), which reduces the original 3D point location estimation error.

Former studies of this problem show that in general, given a total resolution, a horizontally finer discretization improves the overall picture quality for 3D displays [2], [5], [8], and [9]. In these studies the human eyes are modeled as a standalone stereo imaging system without including the role of the 3D display as a medium, the viewing distance, and the actual behavior of the eyes when they are watching a 3DTV. In this paper, we extend previous works by incorporating these factors into the optimization process, and establish formulations which relate the optimal PAR to the practical viewing conditions and parameters such as display size and its distance from the human eyes. These theoretical foundations are described in Sections 2 and 3. In Section 4 we compute typical optimal ratios based on derivations in Section 3. Section 5 discusses supporting subjective studies, and Section 6 is devoted to concluding remarks and future work.

2. 3D ESTIMATION THROUGH A 3D DISPLAY MEDIUM

We will use following notations and definitions in the rest of this paper (see Figures 1 and 2):

- f: focal length.
- (X,Y,Z): a 3D point, and $(\hat{X},\hat{Y},\hat{Z})$: the corresponding estimated 3D point.
- $(x_{l/r}, y_{l/r})$: projection of a 3D point on the left/right image plane.
- $(\hat{x}_{l/r},\hat{y}_{l/r})$: estimation of the projection of a 3D point on the left/right image plane, considering the nearest pixel.
- e_x (e_y): distance between two neighboring pixels in the x (y) direction.
- b_x: baseline of the stereo setup or horizontal displacement between the left and right images on a display screen.
- α (β): eyes (cameras) vergence angle.
- d: viewing distance.
- R: total resolution, *i.e.*, the total number of pixels over the unit square.

We also use subscripts c, D, and h to refer to a stereo imaging system (cameras), 3D display, and human eyes features, respectively. For example f_c stands for the focal length of cameras while f_h means the eyes' focal length.

Figure 2 illustrates the process of capturing, displaying and watching stereo images. Two different stereo systems are involved in this process: stereo capturing and human stereo vision. As showed in this figure, either of these two systems can have its own configuration (parallel or with vergence) independent of the other one. Thus, four different scenarios are possible in this process. In the simplest

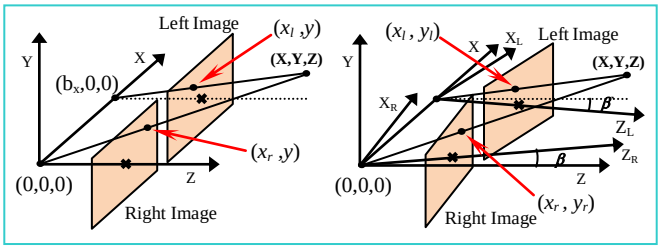

Figure 1: Left: parallel-stereo. Right: vergenced-stereo.

scenario we may consider parallel geometry for both capturing and viewing sides. Then, assuming a pinhole camera model [11], the projection of a 3D point (*X*,*Y*,*Z*) on left and right camera image planes are given by:

$$x_{rc} = \frac{f_c X}{Z}, x_{lc} = \frac{f_c (X - b_{xc})}{Z}, y_c = y_{lc} = y_{rc} = \frac{f_c Y}{Z} \quad (1)$$

The images captured by the stereo cameras are scaled by a factor *S* and presented on the display. Thus, the corresponding 2D point coordinates on the display screen can be computed as:

$$x_{rD} = Sx_{rc}, \quad x_{lD} = Sx_{lc}, \quad y_D = Sy_c \quad (2)$$

Finally, the 3D point projections on the eyes through a display medium placed at distance *d* are obtained as:

$$x_{rh} = \frac{f_h x_{rD}}{(f_h + d)}, x_{lh} = \frac{f_h (x_{lD} + b_{xD} - b_{xh})}{(f_h + d)}, y_h = \frac{f_h y_D}{(f_h + d)} \quad (3)$$

From the formulae (3) the 3D point reconstructed by human eyes is theoretically given by:

$$Z_h = \frac{f_h b_{xh}}{x_{rh} - x_{lh}} = \frac{(f_h + d) b_{xh}}{(x_{rD} - x_{lD} - b_{xD} + b_{xh})}$$
$$X_h = Z_h \frac{x_{rh}}{f_h} = Z_h \frac{x_{rD}}{(f_h + d)}, \; Y_h = Z_h \frac{y_h}{f_h} = Z_h \frac{y_D}{(f_h + d)} \quad (4)$$

In a more realistic scenario we may assume that there is a small vergence angle α acting on the human eyes when watching a 3DTV. From the formulations established in [4] for the vergence-stereo configuration, the 3D point reconstructed by the eyes in this case is given by:

$$Z_h = \frac{b_{xh} AB}{CD + EF}, X_h = Z_h \frac{D}{B}, Y_h = Z_h \frac{y_{rh}}{F} \quad (5)$$

where

$$A = (f_h \cos\alpha + x_{lh} \sin\alpha), B = (f_h \cos\alpha - x_{rh} \sin\alpha)$$
$$C = (f_h \cos\alpha + x_{lh} \sin\alpha), D = (f_h \sin\alpha + x_{rh} \cos\alpha)$$
$$E = (f_h \sin\alpha - x_{lh} \cos\alpha), F = (f_h \cos\alpha - x_{rh} \sin\alpha)$$

and x_{rh}, x_{lh}, and y_{rh} are given by:

$$x_{rh} = \frac{f_h (x_{rD} \cos\alpha - (f_h + d)\sin\alpha)}{(f_h + d)\cos\alpha + x_{rD} \sin\alpha}$$
$$x_{lh} = \frac{f_h ((x_{lD} + b_{xD} - b_{xh})\cos\alpha + (f_h + d)\sin\alpha)}{(f_h + d)\cos\alpha - (x_{lD} + b_{xD} - b_{xh})\sin\alpha}$$
$$y_{rh} = \frac{f_h y_D}{x_{rD} \sin\alpha + (f_h + d)\cos\alpha} \quad (6)$$

If we apply the formulae in (6) into (5) to obtain the 3D point estimation in terms of display coordinates, after some simplification, the formulae in (4) are obtained again (details are skipped here). This means that if the stereo images presented on the display screen are captured under parallel configuration (Figure 1-left), then the 3D-scene

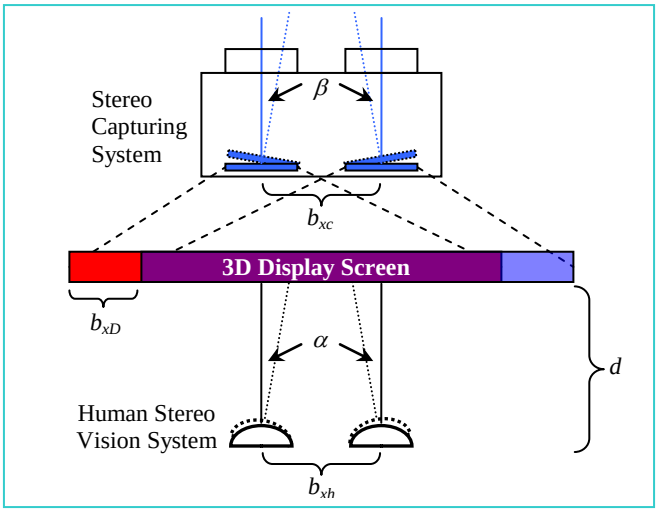

Figure 2: Process of capturing, display and watching stereo.

reconstructed by human eyes, theoretically does not depend on the amount of vergence of the eyes.

In the other two scenarios, the stereo images are captured under vergence (Figure 1-right), therefore assuming a vergence angle β and again using formulations given in [4] the 3D point projections on the camera image planes are obtained as:

$$x_{rc} = f_c \frac{X \cos\beta - Z \sin\beta}{Z \cos\beta + X \sin\beta}, y_{rc} = f_c \frac{Y}{X \sin\beta + Z \cos\beta} \quad (7)$$
$$x_{lc} = f_c \frac{(X - b_{xc})\cos\beta + Z \sin\beta}{Z \cos\beta - (X - b_{xc})\sin\beta}, y_{lc} = \frac{f_c Y}{Z \cos\alpha - (X - b_{xc})\sin\beta}$$

Thus, the coordinates of these projections after representing on the display by a scale factor *S* are:

$$x_{rD} = Sx_{rc}, \quad y_{rD} = Sy_{rc}, \quad x_{lD} = Sx_{lc}, \quad y_{lD} = Sy_{lc} \quad (8)$$

Contrary to the first and second scenarios, here the corresponding projections do not locate on the same raster line. Although, it is still easy to calculate image projections on the eyes, due to differences between the orientations of the eyes and the cameras, it is difficult to find out how the corresponding points are projected back to the 3D space by human vision. Here, we may assume that eyes simply compensate for vertical differences of the corresponding points (*i.e.* differences between y_{rD} and y_{lD}). This may be the reason for some undesired effects of non-parallel capturing like improper scale or shape happen to the reconstructed objects [10]. Nevertheless, our experiments show that this assumption fairly describes the actual behavior of the eyes (see Section 5). Thus, we can say that in all the above-mentioned scenarios, Equation (4) can be used as a good approximate model for 3D point estimation by human eyes via stereo images presented on a 3D display.

3. Optimizing 3D Estimation

In practice, due to the discretized nature of the display screen the actual projections are rounded off to the nearest pixel, therefore the location of a 3D-point is determined using $(\hat{x}_{rD}, \hat{y}_D)$ and $(\hat{x}_{lD}, \hat{y}_D)$. This implies that the 3D-point

(X_h,Y_h,Z_h) is estimated as:

$$\hat{X}_h = \hat{Z}_h \frac{\hat{x}_{rD}}{(f_h+d)}, \hat{Y}_h = \hat{Z}_h \frac{\hat{y}_D}{(f_h+d)}, \hat{Z}_h = \frac{(f_h+d)b_{xh}}{(\hat{x}_{rD}-\hat{x}_{lD}-b_{xD}+b_{xh})} \quad (9)$$

The discretized pixels are within half a pixel from their continuous projected values. Thus, in the worst case:

$$\hat{x}_{rD} = x_{rD} \pm (e_x/2), \quad \hat{x}_{lD} = x_{lD} \pm (e_x/2)$$

$$\hat{y}_D = \hat{y}_{rD} = \hat{y}_{lD} = y_D \pm (e_y/2) \quad (10)$$

Following the procedure mentioned in [5] and considering the worst-case error in 3D estimation, $\hat{Z}_h$ can be rewritten as:

$$\hat{Z}_h = \frac{(f_h+d)b_{xh}}{(x_{rD}-x_{lD})\pm e_x} = Z_h\left(1 \pm e_x \frac{Z_h}{(f_h+d)b_{xh}}\right)^{-1} \quad (11)$$

Ignoring higher order terms in the Taylor expansion of (11) we have:

$$\hat{Z}_h \cong Z_h\left(1 \pm \frac{e_x Z_h}{(f_h+d)b_{xh}}\right) \quad (12)$$

Bounds on error in estimating Y_h can be obtained as follow:

$$\hat{Y}_h = \hat{Z}_h \frac{\hat{y}_D}{(f_h+d)} \cong \frac{Z_h}{(f_h+d)}\left(1 \pm \frac{e_x Z_h}{(f_h+d)b_{xh}}\right)\left(y_D \pm \frac{e_y}{2}\right)$$

or,

$$\left|\frac{\hat{Y}_h - Y_h}{Y_h Z_h}\right| \le f(e_x) = \left\{\frac{e_x}{(f_h+d)b_{xh}} + \frac{e_y}{2|y_D|Z_h} + \frac{e_x e_y}{2(f_h+d)|y_D|b_{xh}}\right\} \quad (13)$$

Considering a unit display area:

$$\left(1/e_x\right)\left(1/e_y\right) = R \text{ or } e_y = 1/e_x R \quad (14)$$

from Equations (13) and (14) we have:

$$f(e_x) = \left\{\left(\frac{1}{(f_h+d)b_{xh}}\right)e_x + \left(\frac{1}{2R|y_D|Z_h}\right)\frac{1}{e_x} + \frac{1}{2(f_h+d)|y_D|Rb_{xh}}\right\} \quad (15)$$

The best solution which minimizes the estimation error of X_h or Z_h is to have e_x as small as possible, but this is the worst possible solution for estimating Y_h. As a compromise, Equation (15) which relates the relative maximum estimation error of Y_h to e_x, can be used to find the optimal PAR. This can be formally stated as the following theorem.

Theorem 1: The optimal display discretization in terms of the relative error in estimating Y_h for a single 3D-point is given by:

$$e_x = \frac{1}{\sqrt{R}}\sqrt{\frac{(f_h+d)b_{xh}}{2|y_D|Z_h}}, \quad e_y = \frac{1}{\sqrt{R}}\sqrt{\frac{2|y_D|Z_h}{(f_h+d)b_{xh}}}. \quad (16)$$

Proof:

$$f'(e_x) = \left\{\frac{1}{(f_h+d)b_{xh}} - \frac{1}{e_x^2}\left(\frac{1}{2R|y_D|Z_h}\right)\right\} = 0 \quad (17)$$

Solving Equation (16) in terms of e_x and considering (14) gives the results. □

Theorem 1 gives the optimal discretization in terms of a single 3D point. Instead, we need to consider minimizing an appropriate error metric over the viewing volume formed in front of the viewer over a depth range $[Z_{min}, Z_{max}]$. Here we use following MSE metric obtained from Equation (17):

$$E(Z_h, y_D) = \left(\frac{(f_h+d)b_{xh}}{Z_h} - |y_D|2e_x^2 R\right)^2 \quad (18)$$

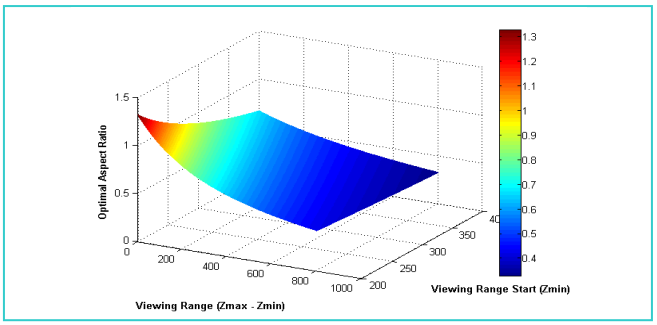

Figure 3: Optimal pixel aspect ratio changes for 14.1" display.

and optimize the following function w.r.t. e_x :

$$F(e_x) = \int_{Z_{min}}^{Z_{max}} \int_{-y_{max}}^{y_{max}} \int_{X_0 - \frac{\Delta X}{2}\frac{(Z_h - Z_0)}{(Z_{max}-Z_0)}}^{X_0 + \frac{\Delta X}{2}\frac{(Z_h - Z_0)}{(Z_{max}-Z_0)}} E(Z_h, y_D) dX_h dy_D dZ_h \quad (19)$$

$$y_{max} = \max(y_D), x_{max} = \max(x_{rD})$$

$$X_0 = \frac{b_{xh}}{2}, Z_0 = \frac{X_0(f_h+d)}{x_{max}}, \Delta X = 2\left(X_0 - Z_{max}\frac{x_{max}}{(f_h+d)}\right)$$

See details of calculating integral (19) boundaries in [2]. This gives us the following theorem:

Theorem 2: The optimal display discretization with respect to the average relative error in the estimation of Y_h over a viewing volume defined by a depth range is given by:

$$e_x = \left(-\frac{(f_h+d)b_{xh}}{2R}\frac{Z_0 I_1 - I_2}{I_3 - Z_0 I_4}\right)^{\frac{1}{2}} \quad (20)$$

where,

$$I_1 = y_{max}^2 \ln\left(Z_{max}/Z_{min}\right), I_2 = y_{max}^2(Z_{max} - Z_{min})$$

$$I_3 = \frac{y_{max}^3}{3}\left(Z_{max}^2 - Z_{min}^2\right), I_4 = \frac{2y_{max}^3}{3}\left(Z_{max} - Z_{min}\right).$$

Proof: Follows the same steps as those used in [2]. □

4. EXPERIMENTAL SETTING

Table 1 shows a typical set of values used for calculating optimal PARs (e_x/e_y). Values of x_{max}, y_{max}, and R are obtained assuming that a 14.1" display with 1280x800 resolution is used. The focal length and baseline are selected close to the human visual parameters. Figure 3 shows the computational results for values mentioned in Table 1 based on Theorem 2 formulations. Figure 3 shows that for closer minimum depth (Z_{min}) with smaller ranges (Z_{max}-Z_{min}), a larger ratio is obtained, but if the minimum depth is far or the viewing-range is large enough a smaller ratio is calculated. Specifically, for the range 150-600 mm, which is a practical range for this configuration, the optimal ratio is calculated as 0.6621 or approximately 2:3 (3 horizontal *vs.* 2

Table 1: Values used for calculating optimal PAR.

Parameter	**Value** (mm)	**Parameter**	**Value**
f_h	17	R	17.76 pixel/mm^2
b_{xh}	65	d	500 mm
x_{max}	303.7021	Z_{min}	200 - 400 mm
y_{max}	189.8138	Z_{range} (Z_{max}- Z_{min})	1 - 800 mm

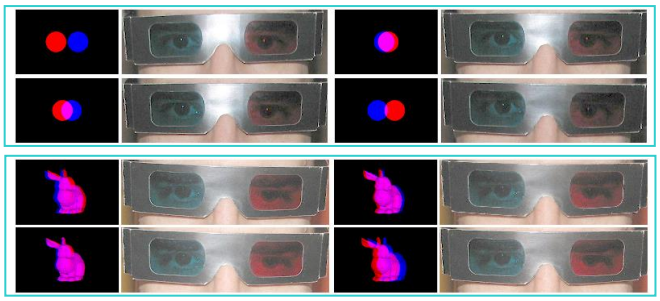

Figure 4: Orientation of eyes in response to different disparities (top) and different stereo capturing vergences (bottom).

vertical pixels). Similar results are obtained for 15" and 17" displays if the viewing distance *d* is proportionally adjusted (larger displays typically viewed from farther distances).

5. CONDUCTING USER TESTS

To understand the human eyes behaviour, we tracked the eyes' reaction to the changes in disparities and also changes in stereo capturing vergence. Figure 4 shows a subset of red-blue images used for test. The two upper rows are some sample of disparity variations and corresponding images of the eyes. The images roughly show that the eyes' orientation is almost independent of the amount of disparity. The two bottom rows are the stereo pairs generated from Bunny 3D mesh under different vergence angles using the virtual stereo imaging system we have established for this purpose. Again, we can see that, in consistent with our assumption in Section 3, the eyes reveal almost the same behaviour in dealing with the stereo pairs generated under different vergence configurations.

We have also conducted some subjective studies to compare users' 3D experience w.r.t. to different PARs. On the capturing side, we used our imaging system to generate stereo pairs with different PARs. On the display side we used a group of neighboring pixels of a conventional display to simulate different PARs. Figure 5 shows a test set generated from a synthetic scene composed of two ellipsoids. Based on the 3D visual experience model proposed in [7], more than 15 viewers were asked to rank these images based on sense-of-depth, picture-quality (regardless of the depth affects), and overall sense. Figure 6, shows the final results of this study. These results show that PAR has meaningful direct relationship with sense-of-depth, but reverse relationship with picture-quality. The picture quality of images with non-uniform discretization decreases because of the image degradation happening in the vertical due to the coarser discretization. However, this degradation is compensated with finer horizontal resolution and improved depth estimations which in turn yields better overall perception of these images. In summary, consistent with our theoretical results, image 2 with PAR 2:3 is the best choice for most (about 42%) of the viewers.

6. CONCLUSION AND FUTURE WORK

In this paper we showed that the behavior of the human

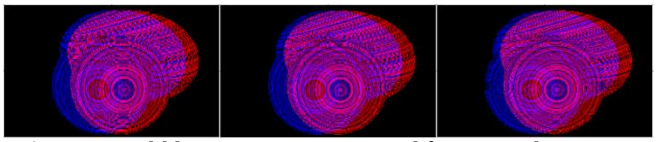

Figure 5: Red-blue stereo pairs generated from a synthetic scene with different PARs. Left 3:8, middle 2:3, and right 1:1.

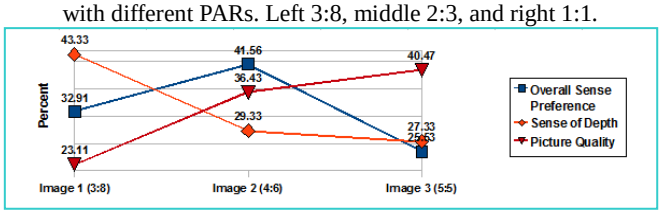

Figure 6: Subjective study results on stereo pairs showed in Fig 5.

eyes in viewing stereo content is almost independent of the imaging system configuration (vergence vs. non-vergence) generating the content. Taking this observation into account, we described a formal method for obtaining the optimal vertical *vs.* horizontal resolution for a 3D display with a specific total resolution. We inferred optimal PAR for typical practical conditions and through subjective evaluations showed that the optimal setting actually improves the 3D-display output quality. However, it will be interesting to extend the results by considering the concept of Just Noticeable Difference (JND) for 3D perception [3] and other human perception factors [1]. Moreover, the effects of non-parallel capturing and other factors such as cross-talk and correlation errors still need further investigation.

Acknowledgments

The authors thank iCORE and NSERC for financial support of our research, and the Stanford Computer Graphics Laboratory for providing some of the 3D models used in our experiments.

References

[1] B.T. Backus, D.J. Fleet, A.J. Parker and D.J. Heeger, "Human Cortical Activity Correlates With Stereoscopic Depth Perception," The Journal of Neurophysiology, 2054-2068, October 2001.

[2] I. Cheng and A. Basu, "Optimal aspect ratio for 3D TV," IEEE 3DTV Conference, 4 pages, KOS, Greece, May 2007.

[3] I. Cheng and P. Boulanger, "A 3D Perceptual Metric using Just-Noticeable-Difference", EUROGRAPHICS, 2005, Ireland.

[4] H. Sahabi and A. Basu, "Analysis of error in depth perception with vergence and spatially varying sensing", Computer Vision and Image Understanding, vol. 63, no. 3, pp. 447-461, 1996.

[5] A. Basu, "Optimal discretization for stereo reconstruction", Pattern Recognition Letters, vol. 13, no. 11, pp 813-820, 1992.

[6] N. Holliman, "3D display systems", Department of Computer Science, University of Durham, Science Laboratories, 2005, http://www.dur.ac.uk/n.s.holliman/Presentations/3dv3-0.pdf.

[7] P. J. H. Seuntins, "Visual experience of 3D TV", Eindhoven: Technische Universiteit Eindhoven, 2006, Proefschrift. http://alexandria.tue.nl/extra2/200610884.pdf.

[8] I. Cheng, K. Daniilidis, and A Basu, "Optimal Aspect Ratio under Vergence for 3D TV", 3DTV Conference, 2008.

[9] A. Basu, and H. Sahabi, "Optimal Non-uniform discretization for Stereo Reconstruction", Pattern Recognition, 1996, Proceedings of the 13th Inter. Conf. on, vol.1, pp.755-759 vol.1, 1996.

[10] B. Javidi, F. Okano, "Three-dimensional Television, Video, and Display Technologies", Chapter 1, Springer, 2002.

[11] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2003.