VISUAL QUALITY ASSESSMENT OF LENTICULAR BASED 3D-DISPLAYS

Ralph Braspenning[1], *Eric Brouwer*[2], *and Gerard de Haan*[1,2]

[1] Department Video Processing, Philips Research Eindhoven
Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands
phone: + (31) 40 27 44141, fax: + (31) 40 27 42630, email: ralph.braspenning@philips.com
[2] Dept. EE: Information and Communication Systems group, Eindhoven University of Technology, The Netherlands

ABSTRACT

The picture quality of prototype 3D TVs based on lenticular displays is currently not at the level of regular 2D displays. The frequency reproduction properties of such a 3D display are analyzed and an experiment is presented to validate the conclusions. Furthermore, we show options, like crosstalk reduction, based on the new knowledge to improve the picture quality.

1. INTRODUCTION

3D Displays that are capable of presenting an illusion of a real 3-dimensional scene to the viewer have been a topic of research for a long time. Common approaches include displays that require the viewer to wear glasses in order to separate the left and right image from the mixed light output of the display. Well-known are the red-green or polarized glasses used for viewing 3D movies in amusement parks.

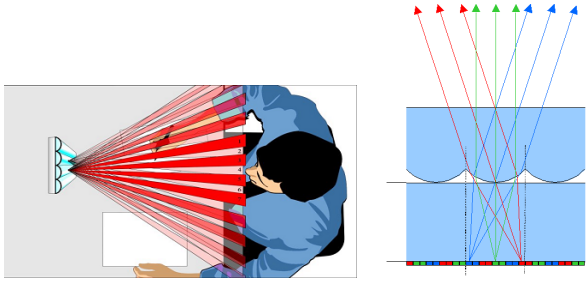
However, wearing (non-customized) glasses has the disadvantage of giving rise to discomfort, especially for longer viewing periods, as would be the case for 3D TV applications. Therefore, so-called auto-stereoscopic displays that do not require the viewer to wear glasses, have been researched. The two main techniques use, either barriers, or lenses (lenticular sheets). In this paper, we will focus on the latter technique. A more elaborate description of this technique will be given in Section 2. Because of the rapid cost-down of flat panel displays, like LCD and plasma, these 3D displays become affordable for 3D TV applications.

Over the years, the picture quality of regular (2D) television sets has steadily improved. An important part of this is due to improved video processing algorithms in these TVs. Some picture quality improvements equally apply to 2D and 3D TV. However, some new issues arise related to the characteristics of the 3D display. This calls for an analysis of the important display characteristics, in order to design algorithms to improve the picture quality.

2. LENTICULAR BASED 3D DISPLAYS

A lenticular based 3D display directs the light of neighboring sub-pixels into different directions by means of small lenses placed immediately in front of the sub-pixels. In this manner different pictures can be transmitted into different directions. Usually a multitude of directions is chosen, e.g. 9 different views. Two of these views can be seen by the left and right eye respectively, and as such create a stereoscopic (3D) image (see Fig. 1a).

In Fig. 1b a schematic drawing is depicted showing a top-down view of the LCD display including the sub-pixels with the lens on top of them. The light of an individual sub-pixel is directed into a certain direction depending on the relative position of the sub-pixel and the lens. Since the lenses are arranged in columns, the periodicity of the lens structure is controlled by the width, or pitch, of one lens column. The pitch of the lens compared to the sub-pixel pitch (width) controls the number of views of the 3D display [1]. Much of the characteristics of the lenticular display are a result of the physical design of the lens. The multiple views enable a considerable freedom of movement by the viewer and the viewing by multiple viewers simultaneously. Both are very important requirements for the 3D TV application.



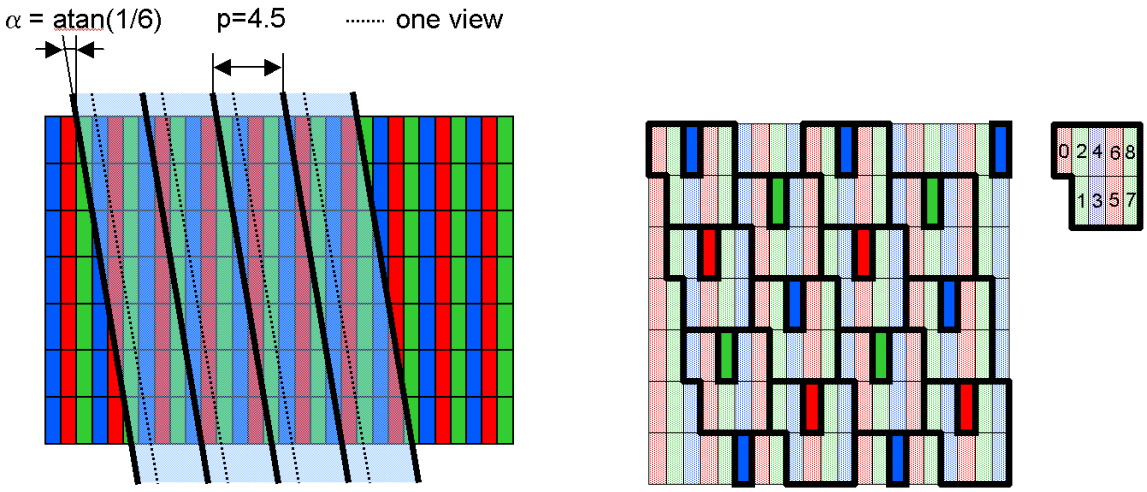
(a) (b)

Figure 1: (a) Overview of the lenticular based 3D display, showing schematically the LCD display, the lenses and a viewer receiving different views in his left and right eye. (b) The top of the LCD display showing the sub-pixels at the bottom, the glass plate on top of that, and, finally the plate with lenses. The light of the sub-pixels is directed into different directions by the lens.

2.1 Pixel Structure

Using vertical columns of lenses, as depicted in Fig. 1a, with a certain pitch, e.g. p sub-pixels, results in a $N = p$ view 3D display. When observing this 3D display at a certain angle only the sub-pixels the light of which is directed into that direction are visible. Hence, the resulting horizontal resolution for a single view would then be N times smaller than the underlying LCD panel resolution, while the vertical resolution would be the same. This gives an undesirable unbalance in the resolution of a view. Therefore, van Berkel [1] proposed to use a slanted lens array instead, with an angle relative to the vertical of , with for instance tan( ) $= 1/6$.

In Fig. 2 such a configuration is depicted. The slanted lenses are shown on top of the LCD display (Fig. 2a). All light coming from the same relative position under the lens



(a) (b)

Figure 2: (a) Front view of a lenticular 3D display with slanted lenses with an angle relative to the vertical of = arctan(1/6). The pitch of the lenses is 4.5 sub-pixels. All light coming from the same relative position under the lens is directed into one particular direction, forming one view, as indicated with the dashed lines. (b) All sub-pixels whose centers are precisely beneath the dotted line in (a) are assigned to that view and highlighted. In total this configuration results in 9 views. Groups of sub-pixels form super-pixels into which all views (directions) are represented. The actual view configuration is depicted on the right.

is directed into one particular direction, forming one view, as indicated with the dashed lines. Sub-pixels that have the largest light contribution can be assigned to this view, i.e. all sub-pixels whose centers are precisely beneath the dotted line in Fig. 2a. These sub-pixels are highlighted in Fig. 2b. In practice, also some light from neighboring sub-pixels will be visible in this view, but we will neglect this here. The pixel structure of the other views are the same but displaced spatially with respect to each other, as indicated at the right of Fig. 2b.

The combination of a lens pitch (p) and a slant angle ( ) yields a certain number of views and a, potentially non-orthogonal, pixel structure per view. These pixel structures can be analyzed mathematically for their spectral reproduction properties, which gives a figure of merit to judge certain configurations. Each sub-pixel position in the grid ($\vec{x}$) can be described as a linear combination of two basis vectors ($\vec{v}_1$ and $\vec{v}_2$)

$$\vec{x} = \mathbf{V}\,\vec{n}, \quad \mathbf{V} = [\vec{v}_1\ \vec{v}_2], \quad \vec{n} \in Z^2 \tag{1}$$

Sampling a continuous signal s_c at the grid results in the sampled signal $s[\vec{n}]$, i.e. $s[\vec{n}] = s_c(\mathbf{V}\,\vec{n})$. The spectrum of the sampled signal, $S(\vec{F})$, can then be expressed in terms of the original spectrum S_c [2, 3], i.e.

$$S(\vec{F}) = \frac{1}{|\det \mathbf{V}|} \sum_{\vec{k}} S_c(\vec{F} - \mathbf{U}\,\vec{k}), \quad \vec{k} \in Z^2 \tag{2}$$

where $\mathbf{U} = (\mathbf{V}^{\mathrm{T}})^{-1}$. Hence the matrix $\mathbf{U}$ determines the position of the repeat spectra in the frequency domain. To avoid aliasing caused by the sampling, the image should be prefiltered such that the repeat spectra do not overlap with the baseband spectrum. This special filtering also has been investigated by Konrad [4, 5]. The maximum frequencies possible are determined by the positions of the repeat spectra, $\mathbf{U}$. This matrix is determined by the matrix $\mathbf{V}$ which in its turn is determined by the pitch of the lenses (p) and the slant angle ( ).

The maximum frequencies allowed without alias by a certain pixel structure is related to the resolution loss and, hence, can be used to define a figure of merit. It turns out that certain combinations of p and are more beneficial than others [1]. One such combination is the configuration presented in Fig. 2. This configuration will be used throughout this paper. Note that the pixel structure resembles a delta-nabla configuration, which is beneficial for image reproduction [6].

The sampling matrix $\mathbf{V}$ for the grid in Fig. 2 can be described on a "pixel" (group of 3 different sub-pixels) basis or on a sub-pixel basis. The pixel based grid is the same as the sampling grid of one color only. Assuming a sub-pixel aspect ratio of 1:3, the sampling matrices in sub-pixel width units become

$$\mathbf{V}_{\mathrm{pix}} = \begin{bmatrix} 9 & 6 \\ 0 & 9 \end{bmatrix} \Leftrightarrow \mathbf{U}_{\mathrm{pix}} = \begin{bmatrix} \frac{1}{9} & 0 \\ -\frac{2}{27} & \frac{1}{9} \end{bmatrix} \tag{3}$$

$$\mathbf{V}_{\mathrm{sub}} = \begin{bmatrix} 5 & 1 \\ 3 & 6 \end{bmatrix} \Leftrightarrow \mathbf{U}_{\mathrm{sub}} = \begin{bmatrix} \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{27} & \frac{5}{27} \end{bmatrix} \tag{4}$$

Voronoi cells are associated with the lattice in the frequency domain formed by the $\mathbf{U}$ matrix [2]. The Voronoi cell around the baseband determines the 2D area to which the signal should be confined in order to avoid aliasing, Hence, the edges of the Voronoi cell determine the maximum frequency. For the pixel basis, the maximum frequency ranges from 0.333 to 0.4 times $f_s/2$ depending on the orientation (This is in accordance with the more balanced pixel structure of Fig 2b). For the sub-pixel basis this is 0.555 to 0.676 times $f_s/2$, which is much higher. However in [7] it is shown that sub-sampling images on a sub-pixel basis results in colour artifacts due to the sampling phase differences in the colours.

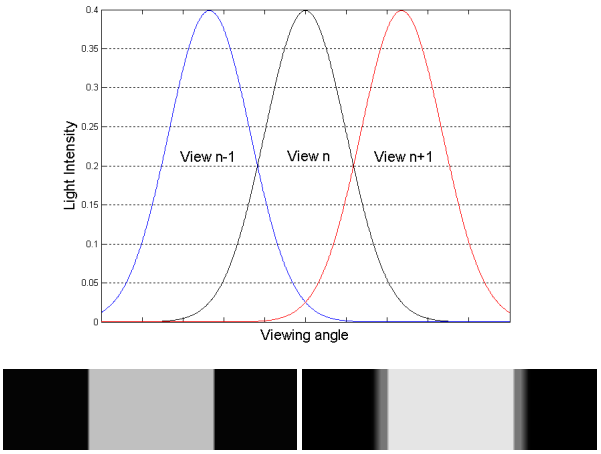


Figure 3: Schematic overview of the crosstalk between different views. The light intensity of a pixel in a view roughly has the shape of a Gaussian. The image at the bottom left shows an example image. The right image shows a simulation of the crosstalk effect when rendering the image at a certain depth.

Therefore, there is a trade-off between sharpness and colour artifacts. In Section 3, we describe an experiment to visually measure the maximum frequency, to see which description is most suitable to describe the measurements.

2.2 Crosstalk

The spatial extent of the views is broadened to improve the uniformity of the 3D display [8]. This, however, increases the crosstalk. At the top of Fig. 3 a schematic overview of the crosstalk between different views is shown. The light intensity of a pixel in a view roughly has the shape of a Gaussian. This can be verified through simulation and measurement. The actual shape depends on the lens design. The light intensity diagram shows three pixels from collocated views having equal luminance. As shown in the diagram, the tails of the intensity profiles overlap.

Moving the eye from one viewing angle to the other results in a smooth transition between the different views. However this also means that given a certain view n that part of the light of neighboring views is visible in that view. This portion of the light is defined as the crosstalk. As one can see from the diagram, the amount of crosstalk changes with the viewing angle.

Rendering a part of an image at a depth other than zero implies that that image part shifts horizontally from view to view. The effect of the crosstalk is illustrated at the bottom of Fig. 3. At the left an example image containing two sharp vertical edges is shown. At the right a simulation of the crosstalk effect is shown when this image is rendered at a certain depth. The crosstalk causes blurring of the edges because shifted versions of the image are partially visible in the rendered view. Hence, the crosstalk reduces the visibility of high frequencies because of interference between frequency patterns.

The crosstalk can be reduced by compensating for the additive effect of the 3D display. However, as mentioned before the additional amount of light from a neighboring view depends on the viewing angle and the lens design. This makes exact modeling difficult and, therefore, we started with a simple model, namely a fixed amount of crosstalk per view. Hence, given the light intensity I_n of view n, the compensated intensity $\hat{I}_n$ becomes

$$\hat{I}_n = I_n - \quad I_{n-1} - \quad I_{n+1} \tag{5}$$

where is the level of crosstalk modeled. Because the crosstalk reduction has to be done in the linear light domain, the same operation can be done using the luminance of the views L_n by applying gamma and inverse gamma correction, i.e.

$$\hat{L}_n = \left(L_n - \quad L_{n-1} - \quad L_{n+1}\right)^{\frac{1}{}} \tag{6}$$

where $= 2.5$ for a typical display.

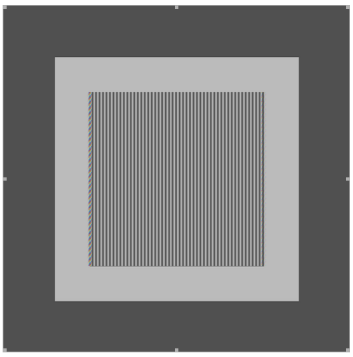


Figure 4: An example test image used in the experiment. It contains 2 DC windows and a middle part with a sine wave having a chosen frequency. The DC windows are added to more easily see the image at a certain depth.

3. EXPERIMENT

To assess the effect of the pixel structure and the crosstalk we performed an experiment. Both effects limit the maximum visible frequency on the display as explained in Section 2. Therefore, we setup an experiment to visually measure the maximum visible frequency. We started with an artificial image containing two DC level windows and a sine wave with a controllable frequency in the middle, see Fig. 4 for an example.

If we choose for the phase of the sine $= f_s x$, then the horizontal frequency becomes

$$f_x = \frac{1}{2}\frac{}{x} = \frac{1}{2}\ f_s, \quad 0 \leq \quad \leq 1 \tag{7}$$

However x is continuous while the pixel grid is a discrete sampling. Substituting k/f_s for x in the phase gives $= k$, where k is the discrete horizontal pixel position. Rendering this sine wave at a certain depth requires shifting the views with a disparity d with respect to each other. Hence, for view n the sine wave function becomes: $\sin(\quad (k - nd))$. To generate the 3D image we sampled this function according to the pixel structure of each view (see Fig. 2b).

For disparities ranging from 0 to 5 pixels in steps of 0.25 we visually determined the maximum frequency by increasing with steps of 1/24 with a maximum of 1 until the frequency became invisible. The result is the blue line shown in

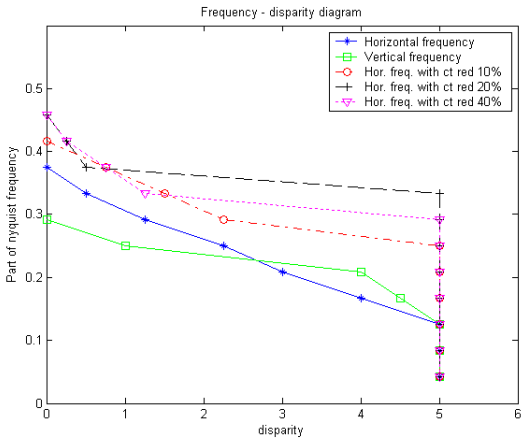


Figure 5: Diagram showing the maximum visible vertical frequency as a function of the disparity, and the maximum horizontal frequency with and without crosstalk reduction applied, using 3 different parameter values.

Fig. 5. As can be seen, the maximum frequency falls off approximately linearly with the disparity. The same experiment was performed using a vertically varying frequency. This is shown with the green curve in Fig. 5. One can see that the behaviour is similar to the horizontal frequency, although it starts somewhat lower.

Next we applied the crosstalk reduction algorithm from Section 2 to judge the influence of the crosstalk. For the parameter we used values 0.1, 0.2 and 0.4. The results are shown in Fig. 5. For $= 0.1$ and $= 0.2$ a clear increasing improvement is visible, but for $= 0.4$ the improvement decreases again compared to $= 0.2$. Furthermore, the maximum frequency stays almost constant for the larger disparities as opposed to the case without crosstalk reduction.

4. DISCUSSION

The maximum visible horizontal frequency for zero disparity was measured to be around 0.375 times $f_s/2$, and the maximum vertical frequency measured was around 0.30 times $f_s/2$ (see Fig. 5). This is in best accordance with the pixel based description of the pixel structure as derived in Section 2. When crosstalk reduction is applied the maximum visible frequency increases above the theoretical limits set by the pixel based description hinting that one profits from sub-pixel sampling. The limits of the sub-pixel based description are not reached, but this was to be expected looking at the work presented in [7].

The crosstalk reduction using a simplified model proved to increase the maximum visible frequency. Although, the crosstalk reduction with $= 0.4$ proved to be worse than with $= 0.2$, while in practice the overlap between the views is significant (see Fig. 3). This is most probably due to the simplified model used in this paper. The next step would be to investigate more elaborate models and to see if the results can be further improved.

The knowledge of the presented curve gives the opportunity to optimally render an image for the 3D display. Two approaches are possible. First of all given a local frequency in the image its depth can be clipped to the maximal depth at which it could be rendered without alias. Secondly, given a local depth value, the image can be pre-filtered such that no artifacts will occur at that depth.

Furthermore, sharpness enhancement methods [9] boost the higher part of the image spectrum. Again the presented curve can help control the amount of enhancement, in order to obtain the optimal picture quality.

5. CONCLUSIONS

Currently the picture quality of a lenticular based 3D TV still lags behind that of regular 2D TVs. This is partly due to the special characteristics of a lenticular based 3D displays. Knowledge of these characteristics can help to improve the picture quality by optimizing image rendering, e.g. the presented crosstalk reduction increases the maximal visible frequency for disparities over 4 pixels by a factor 2. Besides that, image enhancement methods can be controlled using the knowledge.

REFERENCES

[1] C. van Berkel, "Image preparation for 3D-LCD," in *Proceedings of the SPIE*, 1999, vol. 3639, pp. 84–91.

[2] A. Tekalp, *Digital Video Processing*, Prentice-Hall, 1995, ISBN: 0-13-190075-7.

[3] E. Dubois, "The sampling and reconstruction of time-varying imagery with appliction in video systems," *Proceedings of the IEEE*, vol. 73, no. 4, pp. 502–522, 1985.

[4] J. Konrad and P. Agniel, "Artifact reduction in lenticular multiscopic 3-D displays by means of anti-alias filtering," in *Proceedings of the SPIE*, Jan. 2003, vol. 5006A, pp. 336–347.

[5] J. Konrad and P. Agniel, "Non-orthogonal sub-sampling and anti-alias filtering for multiscopic 3-D displays," in *Proceedings of the SPIE*, Jan. 2004, vol. 5291, pp. 105–116.

[6] L. Silverstein et al., "Effects of spatial sampling and luminance quantization on the image quality of color matrix displays," *Journal of the Opt. Soc. Am. A*, vol. 7, no. 10, pp. 1955–1968, 1990.

[7] M. Klompenhouwer and G. de Haan, "Subpixel image scaling for colour matrix displays," *Journal of the SID*, vol. 11, no. 1, pp. 99–108, 2003.

[8] W. IJzerman et al., "Design of 2d/3d switchable displays," 2005, Accepted for publication in *Proceedings of the SID*.

[9] E. Bellers and J. Caussyn, "A high definition experience from standard definition video," in *Proceedings of the SPIE*, Jan. 2003, vol. 5022, pp. 594–603.