A. Boev, R. Bregović, and A. Gotchev, "Visual-quality evaluation methodology for multiview displays," *Displays*, vol. 33, 2012, pp. 103-112, doi:10.1016/j.displa.2012.01.002.

Visual quality evaluation methodology for multiview displays

Atanas Boev, Robert Bregovic, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

firstname.lastname@tut.fi

ABSTRACT

Multiview displays are characterized by a multitude of parameters such as spatial resolution, brightness, 3D-crosstalk, etc., which individually and in their combination influence the visual quality of the displayed 3D scene. These parameters are specified by values, precisely measured by optical means. However, it is difficult for an average consumer or content producer to compare the visual quality of two displays or judge if a given 3D content is suitable for a certain display only by this set of parameter values. In this paper, we propose a quality measurement methodology, which aims at measuring the visibility of structural distortions, introduced by a multiview display, to a number of test signals with different frequency content and apparent depth. We use these measurements to derive what we call *display passband* for signals at different disparity levels. The passband determines the frequency components for which the intended signal is predominantly visible, with respect to the distortion introduced by the display. Additionally, we propose a method to determine the approximate effective resolution of the display for signals with a given apparent depth. The result of the measurements can be used to compare the perceived visual quality of different multiview displays.

1. INTRODUCTION

Multiview displays can create visual illusion of objects floating in 3D space without requiring the observer to wear 3D glasses. Typically, multiview displays combine a pixel-addressable matrix, such as in plasma, LED, or LCD panels, with additional *optical layer* mounted on top [1]. The optical layer redirects the light generated by the pixel matrix, making the visibility of each pixel element a function of the observation angle. The set of elements visible from certain angle forms an image, also called a *view* [1][2]. A multiview display can simultaneously show a number of different views, each one visible from different direction. The process of combining multiple images in one compound bitmap is referred to as *interleaving*, and the map that links the position of TFT elements with the view number they belong to is referred to as *interleaving map*. If the views are properly selected observations of the same scene, the display recreates the scene in 3D. Even though all objects of the scene are projected on the display, they might appear as they are at different distances to the observer. The apparent distance to an object is referred to as its *apparent depth*. If the object appears at the display level, all its observations appear on the same display coordinates. If the object has different apparent depth, it appears on different horizontal coordinates in each view. The distance between the positions of an object in different views is referred to as *disparity*. The objects with positive disparity appear behind the display level, and those with negative disparity appear in front of the display.

The downside of the optical layer is that it introduces a number of multiview display specific artifacts [1]. In addition to monoscopic display parameters, such as 2D resolution or refresh rate, the visual quality of a 3D monitor is influenced by variables such as 3D-contrast and 3D-crosstalk [3]. The multitude of parameters hinders the comparison of the visual quality of different multiview displays. A number of previous works have addressed the estimation of display optical quality, ranging from theoretical considerations about the interleaving map [4][5][6] trough measuring of the optical

parameters of the display [3][7][8] to subjective tests with different multiview displays [9][10][11]. However, evaluating the quality of a multiview display based on its optical parameters only, has two main disadvantages – first, some parameters, e.g. luminance uniformity across different observation angles, are not directly related to the perceived quality; and second, visibility of 3D artifacts depends also on scene content, observation conditions and properties of the human visual system. Having human observers to rate the visibility of artifacts in all scenes would be an optimal quality assessment approach; however it is expensive and time-consuming.

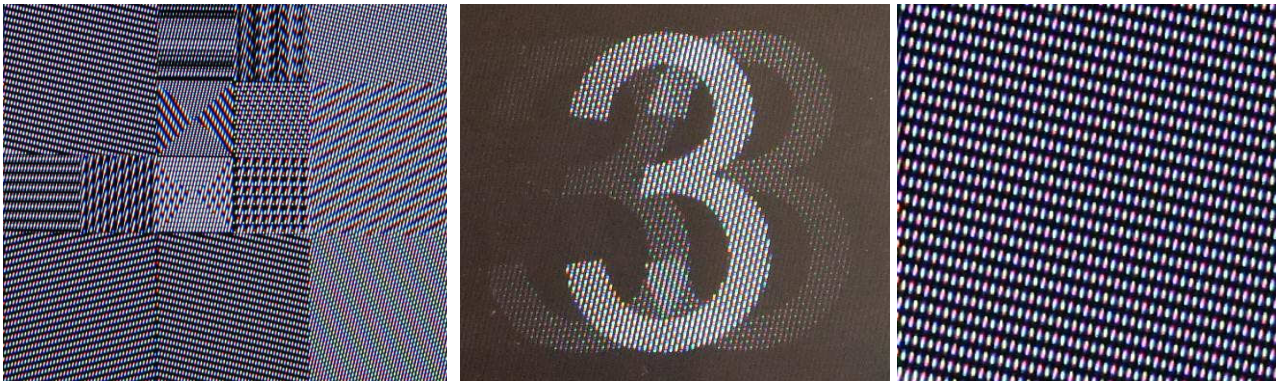
In this article we propose a methodology aimed at evaluating the level of signal distortion introduced by a given multiview display. We use multiple test signals with various frequency components and disparity levels to derive the display passband regions for planes with different apparent depth. The passband regions are estimated for two levels of distortion visibility. The shapes of the passband regions can be used to estimate the expected visual quality for different types of 3D content. Additionally, we devise a method aimed at approximating the equivalent resolution of the display for given disparity range and distortion level. The equivalent resolution of a multiview display can be directly used as a perceptually-relevant indicator of its visual quality. In a previous work, we have proposed passband evaluation methodology, which did not consider disparity [12]. Here, we extend our previous approach for a range of disparities and multiple distortion levels. In another work [13], we have estimated the distortion levels based on knowledge of interleaving pattern and angular visibility function. However, we noticed that non-linear optical effects are introducing significant differences in passband regions in measured versus simulated data. The results in the current article are based solely on measurements, and do not require knowledge of the angular visibility.

The paper is structured as follows: in Section 2, we introduce the concept of distortion visibility to be used as indicator of perceptual quality. We introduce a model of multiview display, use it to explain the most common artifacts, and give a general methodology for measuring and evaluating visual distortions. In Section 3, we give details about the measurement procedure, like test image preparation and experimental setup. We give a practical example with measurements of a 24-view display. In Section 4, we explain how the measured data is evaluated in order to obtain the passband regions of the display for a given distortion level. In Section 5 we show how using these regions one can approximate the equivalent resolution of the display for different depth planes and different distortion levels. In the last section we give concluding remarks.

2. VISIBILITY OF DISTORTIONS AS INDICATOR OF VISUAL QUALITY

2.1 Artifacts in 3D displays

The most pronounced artifacts in a multiview display are moiré and ghosting artifacts [16]. Typically, the visible pixels of a view appear on a non-orthogonal grid [1][4]. Mapping the input images, which are usually sampled on rectangular grid, to the visible pixels of a view requires special anti-aliasing filters [5][6][17]. Direct mapping of multiple images to the views of a multiview display produces moiré and color aliasing artifacts similar to the ones shown in Figure 1a. The design of a multiview display involves a trade-off between number of views, spatial resolution of a view, and visibility artifacts such as *image flipping* and *banding* [4][11]. Often, the visibility zones of different views are interspersed and from a given angle multiple views are simultaneously visible, albeit with different brightness [1][2][3]. When visualizing 3D objects with pronounced depth the combination of disparity and simultaneous visibility is perceived as ghosting artifacts [5][10]. An example for ghosting artifacts is given in Figure 1b. Often, the process responsible for ghosting is modeled as crosstalk, and the term crosstalk is used as a synonym for ghosting artifacts [3][5][10][11][16]. For displays with parallax barrier, the barrier creates visible gaps between the pixels, as seen in Figure 1c. These gaps are seen as masking artifacts [12], similar to the fixed-pattern noise exhibited by some digital projectors [18].



a) **b)** **c)**

Figure 1 Typical artifacts exhibited by multiview displays: a) moiré, b) ghosting and c) masking

2.2 Multiview display as an image processing channel

In order to assess perceptual differences between intended (input) and visualized (output) signal, we propose a model, which considers the multiview display as an image processing channel. The model follows the steps of content preparation and visualization, and has five stages as shown in Figure 2. For simplicity, the examples in the figure are given for single row of a dual-view display, however, the general signal transformations hold true for any multiview display. The channel input is an image, which is meant to be seen at a particular depth. It can be regarded as a continuous signal, as it is shown in Figure 2a. The second stage models the preparation of the views. From the interleaving map one can derive a binary mask defining the position of the samples in one view. The input signal is sampled using the binary mask of the first view. For example, let these samples appear of positions with odd number as horizontal coordinate, as shown in Figure 2b. The third stage models the presence of disparity. If the input image is meant to be seen on the level of the display, it appears on the same place in each view. In that case, the same input signal is sampled using the binary mask of the second view. For example, positions with even number as horizontal coordinate, as seen in Figure 2c. If the input image should appear at different depth, it has disparity between the views, and an offset version of the input signal is sampled, as exemplified in Figure 2f. The fourth stage models the process of interleaving. Following the interleaving map, observations of the same object with different disparity are combined together. In our model, this corresponds to a combination of multiple versions of the same input signal, sampled with different offset. For example, an interleaved version of the input signal with zero disparity is shown in Figure 2d. It is made by alternating the samples in Figure 2b (odd positions) and the ones in Figure 2c (even positions). Alternatively, an example for the same input signal interleaved with disparity 20 is given in Figure 2g, which is a combination of the samples in Figure 2b and 1f. The last stage models the influence of the optical layer. The layer changes the visibility of the individual display elements depending on the observation angle. In our example, from certain observation position the odd samples are seen with full brightness, while the even samples are seen with one quarter of the brightness (Figure 2e and 1h).

It should be noted, that for objects with zero disparity the interleaved image (Figure 2d) is a good representation of the input signal (Figure 2a). In that case, a multiview display can be modeled as a 2D display where parts of the image have partial visibility, as suggested by Jain and Konrad in [14]. The less the impact of the optical layer is, the closer the visual output to the input signal is (Figure 2e), and – as Jain and Konrad have proven – the bigger the frequency throughput of the display is. Lower visibility of the masked pixels results in alternating bright and dark pixels (Figure 2e), which can be modeled as fixed-pattern noise. However, if disparity is introduced, the interleaved signal (Figure 2g)

is quite different than the input signal. In that case, the masking effect of the optical layer makes the output (Figure 2g) better representation of the input signal. The shifted version of the input signal is meant to be fully visible from another observation angle. If it is partially present in the current observation angle, as shown in Figure 2h, it is modeled either as crosstalk between the views [7], or as interperspective aliasing [15].

The sampling stage in Figure 2 imposes an anti-aliasing filter before it. We deliberately do not include it into the model. In a multiview display, beside aliasing one has to simultaneously deal with other sources of distortions, such as imaging and crosstalk. In order to simplify the evaluation methodology we do not apply any anti-aliasing filter to the images displayed for measurements. Thus, we would see clearly the aliasing artifacts along with other artifacts as well as their interaction. From a pre-processing filter implementation point of view, this would allow addressing most of distortions by a single filter. In other words, we deal with distortions according to their visibility regardless of their origin. As seen from the model, the only place one can influence the signal is before the sampling stage as this is the only place where aliasing can be eliminated. Filters designed by the proposed methodology would generally act as anti-aliasing filters but also cancel some other frequency being source of imaging and cross-talk artifacts. Note that by measuring the artifacts in this way one can design more-restrictive or less-restrictive filters depending on the interaction between artifacts. Also the measurements can quantify in a better way the limits in changing the filter parameters for providing subjectively more pleasant visualization.

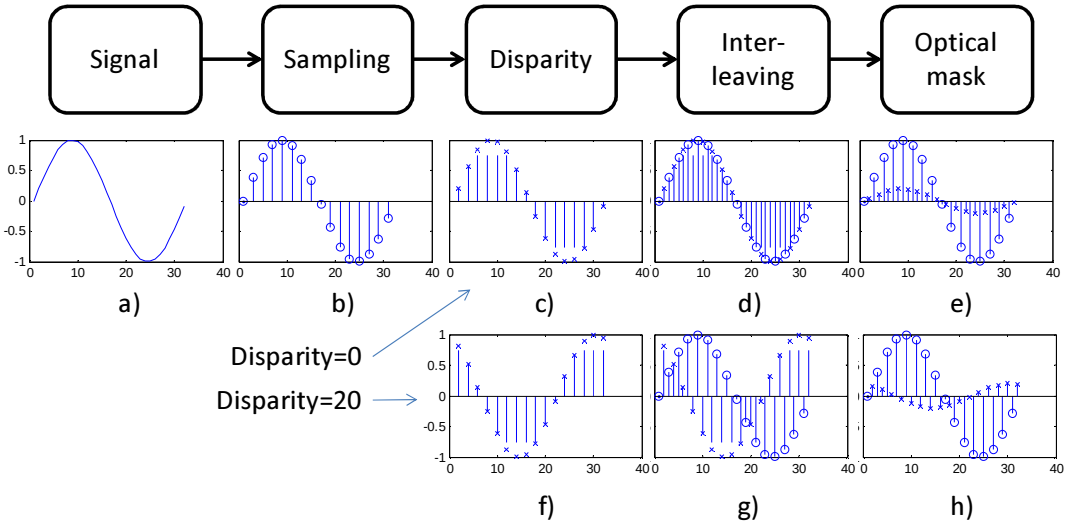


Figure 2 Generalized model of a multiview display as an image processing channel: a) input signal, b) input signal sampled in the positions which belong to view "1", c) input signal sampled in positions, which belong to view "2", d) interleaved signal, containing samples from views "1" and "2" and no disparity between the views, e) interleaved signal from "d)", after being masked by the optical layer, f) input signal, sampled in positions which belong to view "2", with an offset of 20 samples, g) interleaved signal, containing samples from views "1" and "2" and 20 samples disparity between the views, and f) interleaved signal from "g)", after being masked by the optical layer.

2.3 Properties of the human visual system

Most visual quality metrics work by assessing the perceptual difference between two images – one is the reference image and the other is the processed one. The reference is assumed to be of highest quality and the bigger the perceptual difference between the images is, the lower the quality of the processed image is deemed to be [21]. In the general case, however, the observer does not have the reference image available for comparison, and predicting the visibility of an artifact becomes a more complex task.

The human visual system (HVS) is optimized for extracting the structure from an image, and is largely insensitive to global contrast or brightness variance [19][20]. The act of seeing an object is a product of two phenomena – *visual perception*, which is the ability of the eye to collect visual data, and *visual cognition*, which is the ability of the brain to interpret visual information. The limitations of each stage are modeled as various HVS properties. For example, an observer, looking at the rectangular grating shown in Figure 3a might have a mental image that is perfect replica of the original. The visual perception is limited by physiological factors, like optical properties of the eye, and density of photoreceptors on the retina. Certain spatial frequencies are easier to perceive then others, and this limitation is modeled as *contrast sensitivity function* (CSF) [22]. For example, a dense grating such as the one shown in Figure 3b might be seen indistinguishable from an uniform shade of gray, due to the physical inability of the eye to "sample" signal with such density. On the other hand, HVS is able to reconstruct the underlying structure even if it is partially obscured. For example, the group of black polygons, shown in Figure 3c is seen as continuous black lines with white lines overlaid on top. This effect is due to visual information being processed by the V1 brain center, which behaves as a series of filters with different spatial frequency and orientation [19]. The ability of the brain to reconstruct shapes and repetitive patterns is known as the *visual Gestalt principle* [19] , and the interdependent visibility of patterns with different properties is modeled as *pattern masking* [22]. According to the Gestalt principle, closely positioned shapes are grouped according to their proximity.

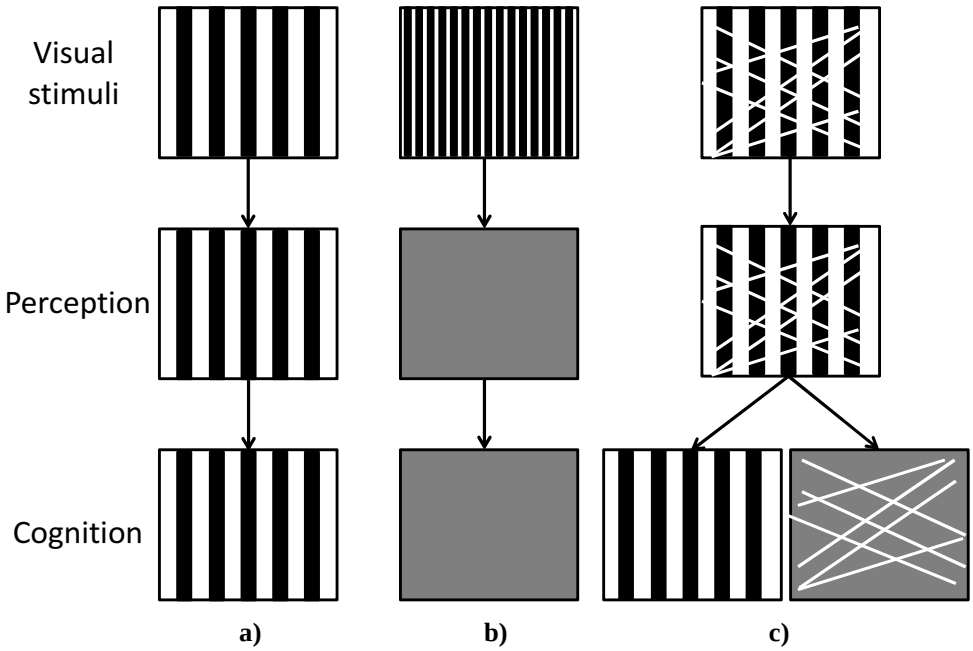


Figure 3 Examples for visibility as a combination of perceptual and cognitive processes: a) sparse grating, b) dense grating, and c) sparse grating with masking

Predicting the visibility of an image detail is a complex task, as it is influenced both by HVS parameters (contrast sensitivity function, pattern masking, etc.) and observation conditions (distance to display, ambient light) [22]. Still, there are works on the general visibility of stereoscopic crosstalk in typical observation conditions [10],[11]. According to the Weber-Fechner law the perceptibility of a change in stimuli is proportional to the amplitude of the stimuli. This fact also holds true for perception of brightness [19]. Following the Weber-Fechner law, the crosstalk is measured as percentage of the intended signal (intended signal is the input signal as perceived on the display). Crosstalk of less than 5% is considered under the visibility threshold and crosstalk of 25% or more is considered unacceptable [10]. The threshold level of barely acceptable crosstalk depends on the local contrast of the content and on the white-to-black contrast ratio of the display. This level has been reported as 10% [24], 15% [25] and 25% [10]. Typically, the level is

measured with high-contrast, black-and-white patterns, but for natural images, a higher crosstalk level might be acceptable [24], [25]. In our paper, we use 20% as the value for barely acceptable crosstalk.

2.4 Criteria for visibility of distortions in multiview displays

In order to estimate the visibility of distortions, we model three HVS properties – brightness perception, contrast sensitivity function and Gestalt principle. This is done by a three-step procedure in frequency domain. First, we model the contrast sensitivity function by applying a circular weighting window. The weights in the window change as a function of the distance to the center of the coordinate system, and the shape of the function follows the shape of the spatial CSF at photopic level as described in [23]. Then, according to the Gestalt principle, we identify the visually dominant pattern by searching for the lowest spatial frequency regardless of the orientation. In frequency domain this is expressed as proximity of the peak of the signal to the center of the coordinate system (DC). Finally, according to the Weber-Fechner law, the eye senses brightness approximately logarithmically for typical observation conditions. Thus, we measure the visibility as the ratio between the amplitude of the distortion introduced by the display and the amplitude of the intended signal.

$$\delta(f_x, f_y) = \frac{\text{amplitude of the distortion}}{\text{amplitude of the signal}} \cdot 100 \ \%$$

In the rest of paper this is referred to as the *distortion to signal ratio*. Since the display behaves differently for different frequencies, the display distortion will depend on the horizontal, f_x, and vertical, f_y, signal frequency.

In this work we analyze the distortion by applying threshold at two different levels – 5% distortion level, which represents unnoticeable levels of distortion, and 20% which represents visible, but still acceptable artifact levels.

2.5 Evaluation methodology

In order to assess the visual quality of a multiview display, we prepared test images with varying spatial frequency, orientation and depth, and for each test image we measured the relative distortion introduced by the display. Our measurement methodology has six steps, which are shown in Figure 4.

The first step is to prepare number of *test signals*, which contain a 2D sinusoidal pattern with varying horizontal and vertical frequency components. Then, each test signal is extended to a number of *test images* each one with different apparent depth. This is done by mapping the same signal to each view of the display, adding different amount of disparity to each view and interleaving all views in a test image. The third step involves automated visualization of all test images on the display and making a snapshot of each one with a high resolution camera. The output of that step is a collection of *test shots* of all test images. In the next step the *spectra* of each test shots are analyzed, in order to determine the distortion to signal ratio, that is, the ratio between the magnitude of the distortion frequency component introduced by the display and the magnitude of the intended frequency component in the test signal. The distortion frequency component is selected as the largest peak in the spectra, which is positioned closer to the center than the intended frequency component of the input signal. Based on the selected threshold (distortion level), the intended frequency of the test image is marked as being inside of the display passband (if the distortion to signal ratio is smaller than the threshold) or being outside of the display passband (otherwise). At step five, all frequencies with the distortion to signal ratio smaller than the threshold are combined into the display passband area. In the final step, each passband is approximated by a rectangle, with the same area and the same horizontal to vertical ratio of the evaluated display passband. The horizontal and vertical sizes of that rectangle are used for estimating the equivalent resolution of the

display for the corresponding disparity. The output of the last step is a list of equivalent display resolutions for multiple disparity levels and different thresholds. In this paper the analysis is done for two thresholds, 5% and 20%.

Step two of the quality evaluation methodology requires knowledge of the interleaving pattern of the measured display. Note, that it is possible that the interleaving pattern, if provided by the display manufacturer, is a simplified version of the actual one and does not accurately identify the groups of display elements with similar angular visibility functions. Since the interleaving pattern is an important part in the described quality evaluation method, it has to be known (or evaluated) as correctly as possible. In a previous work, we have described an approach for deriving number of views and interleaving pattern of a given display [12].

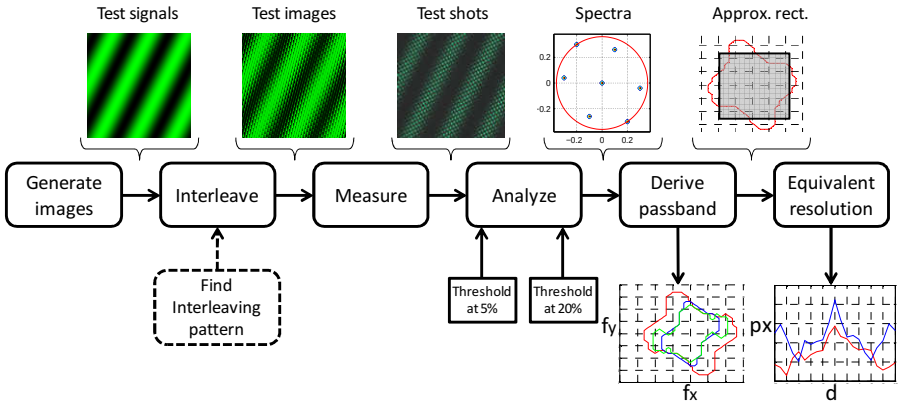


Figure 4 Block diagram of the proposed quality evaluation methodology.

3. MEASUREMENTS

3.1 Generation of test images

The aim of the test image generation procedure is to create a collection of images containing various frequency components, with different apparent depth. The first step is to generate a number of 2D test signals where the brightness of each pixel is calculated by using

$$I = \frac{1}{2}\sin(\pi \cdot x \cdot f_x + \pi \cdot y \cdot f_y) + \frac{1}{2},$$

where I is the brightness, x, y are the horizontal and vertical coordinates of the pixel and f_x, f_y are the horizontal and vertical frequency components, correspondingly. Furthermore, $x = 1,2,...,x_{\max}$, where $x_{\max}$ is the width of the test signal, $y = 1,2,...,y_{\max}$, where $y_{\max}$ is the height of the test signal, $f_x \in [0,1]$ where 1 is normalized to be half the horizontal sampling rate and $f_y \in [-1,1]$ where 1 is normalized to be half the vertical sampling rate. Although it is beneficial to have as many as possible frequency pairs (f_x, f_y), in order to keep the number of test images reasonably small, in our experiments we increase f_x and f_y with step of 0.025. It should be pointed out that in order to generate signals with all possible frequencies that can be displayed we have to use $f_x, f_y \in [-1,1]$. However, due to the symmetrical properties of the 2D discrete Fourier transform (DFT) when applied on real-valued signals it is enough to use only half of the frequency space as selected above – the magnitude response for a signal with frequency (f_x, f_y) is identical to the one for signal with frequency $(-f_x,-f_y)$. Magnitude response for frequencies $f_x \in [-1,0]$ and $f_y \in [-1,1]$ can be generated by mapping the ones for $f_x \in [0,1]$ and $f_y \in [-1,1]$. This is illustrated in Figure 5. Every

point in the 'green' part of the frequency domain has a symmetrical point in the 'red' part of the frequency domain. For illustrative purposes, few of those point-pairs are highlighted.

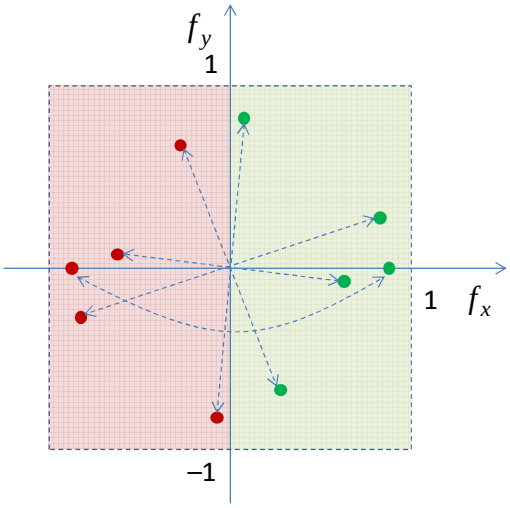


Figure 5 Symmetry of 2D real-valued signals in frequency domain (2D DFT)

The next step is to render a number of test images from each test signal, by adding different disparity to each view. First, one should take v copies of the test image, where v equals the total number of views for the measured display and assign them to the views of the display. Then, the contents of each view are shifted horizontally with an offset $s_n = d \cdot n$, where s_n is the offset for the n-th view, n is the view number and d is the targeted disparity. In our experiments d varies between -10 and 10. Finally, the n views are interleaved into a test image, according to the interleaving map of the display. Test images with negative d have apparent depth in front of the screen and test images with positive d have apparent depth further away than the screen plane. Note that in this case *disparity* refers to the disparity between the views, and not the perceived disparity. The former is the offset in pixels between images in neighboring views and the latter is the offset between the images seen by each eye. The artifacts caused by the optical layer are visible by a single eye, and can be measured by a single camera. Such artifacts do not affect the perceived disparity; therefore second camera is not necessary.

In our experiments, we used 23" 3D-display manufactured by X3D-Technnologies, hereafter referred to as *X3D-display*. The display is marketed as an 8-view display, has TFT-LCD matrix with resolution of 1920x1200 and wavelength-selective optical layer which acts as parallax barrier [22]. Since we had a rough estimate of the passband of the display from previous measurements, our test signals did not include all frequency combinations. We prepared 441 tests signals with 21 disparity steps, which resulted in 9261 test images. Three of our test images are shown in Figure 6. Each of them is generated using $f_x = 0.2$ and $f_y = 0.1$, the test image in Figure 6a has disparity $d = 0$ and the test images in Figure 6b and Figure 6c have $d = 1$ and $d = 5$, respectively.

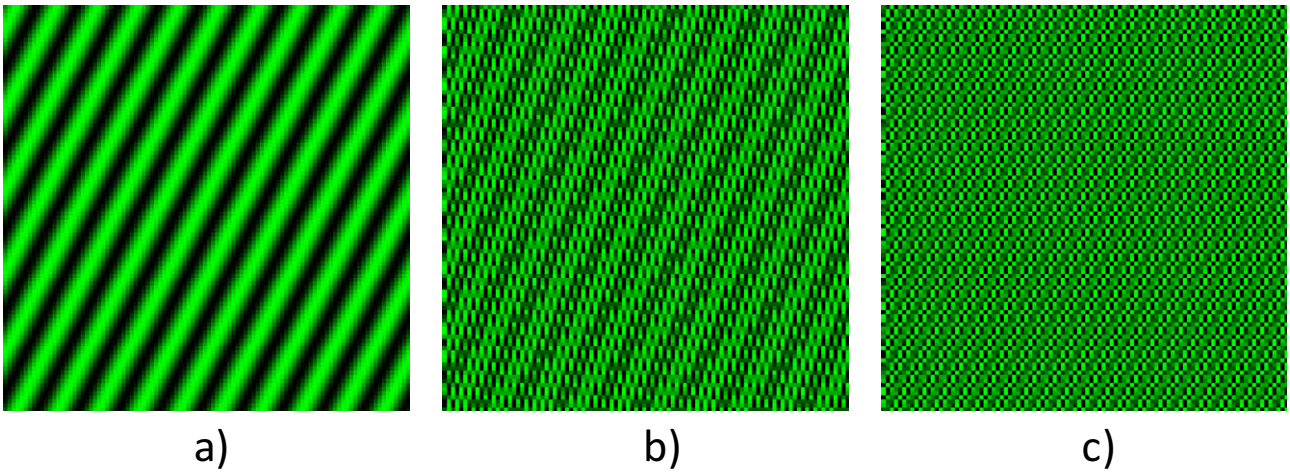


Figure 6 Example of test images with $f_x = 0.2$, $f_y = 0.1$ and different disparity (enlarged details): a) $d = 0$, b) $d = 1$, c) $d = 5$

3.2 Experimental setup

The next step is to visualize each test image on the display and photograph it. In our experiments, we used an HD resolution camera with GigE interface and custom software which automates the visualization-capture-store cycle. The camera was positioned at a distance of 70cm from the display, which is within the nominal observation distance of the X3D-display [27]. In order to avoid aliasing and minimize the influence of the camera, one should use zoom factor that gives the highest possible ratio between the number of photographed display pixels and the number of pixels in the test shot. In our measurement the ratio was 2.24 pixels in the test shot for each pixel in the test image.

The brightness and contrast settings of the display can affect the visibility of image details and thus – the perceptibility of artifacts. Naturally, the visual quality of any display is influenced by its calibration. Our suggestion is to measure the display in typical observation conditions, with values for contrast, brightness and gamma perceptually calibrated to ensure the largest amount of distinguishable levels of gray. In our measurements, the contrast of the display was set to 50%, brightness to 100% and the gamma was set using the visual gamma calibration procedure provided by the drivers of the video card. In order to avoid measurement noise, one should select the lowest ISO sensitivity of the camera and choose exposure time that gives sufficient dynamic range without saturation. We used ISO 50 and exposure time of 1.5sec. The images were captured in gray scale with intensity range between 0 and 217. The three test shots shown in Figure 7 are photographs of the corresponding test images in Figure 6. By comparing images in Figure 6b and Figure 7b, one can see that for the selected f_x and f_y the optical layer of the display works well for $d = 1$, leaving the intended signal predominantly visible. The comparison between images in Figure 6c and Figure 7c, shows that for $d = 5$, the optical layer has undesirable effect on the same combination of f_x and f_y.

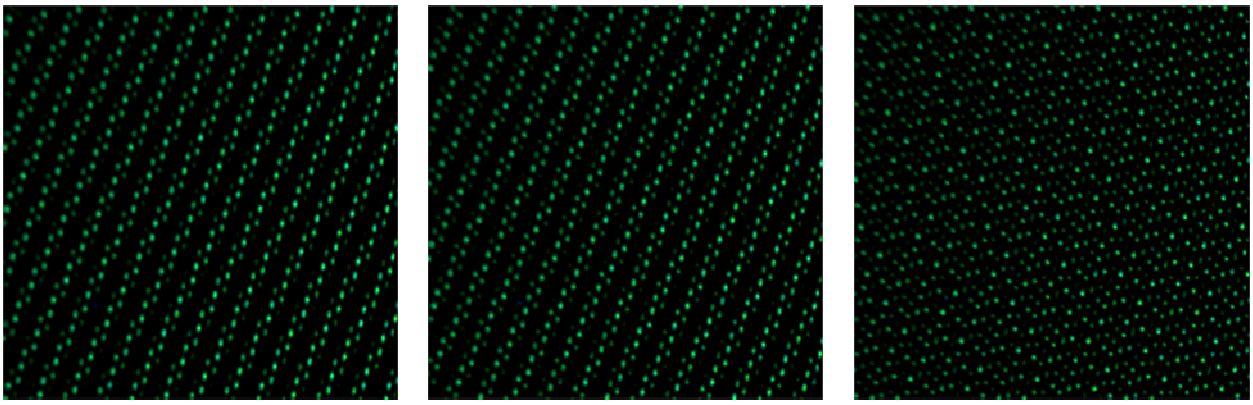


Figure 7 Example of test shots with $f_x = 0.2$, $f_y = 0.1$, and various disparity, acquired during the experiment (enlarged details): a) $d = 0$, b) $d = 1$, c) $d = 5$.

4. DISPLAY PERFORMANCE IN FREQUENCY DOMAIN

In this section we describe the procedure for evaluating the performance of the display in the frequency domain by processing the test shots obtained as described in the previous section. We do all processing in the frequency domain in order to simplify dealing with various problems that might occur during measurement, for example, camera position, difference in pixel size in the camera and pixels on the display, etc. [12].

4.1 Analysis of frequency components

In an ideal case, a test shot should be visually identical to the test image. However, as discussed in Section 2.1, this is not the case in practice due to the various distortion introduced by a multiview display. For a viewer, when looking at the display, it is obvious if an image is properly represented on the display (see Figure 7a, Figure 7b) and when it is not (see Figure 7c). Such clear identification of visible pattern, as can be done by a human, is not straightforward to obtain by using a signal processing algorithm. Moreover, some of the introduced distortions are more disturbing for a viewer than others. For example, as seen in Figure 7a and Figure 7b, there are many high frequency distortions (dark gaps in the lines). However, HVS is trained to find underlying patterns by grouping features together (*Gestalt principle*) and is sensitive to the predominant frequency components (*pattern masking*) [19]. Therefore, the distortions seen in Figures 7a and 7b, do not hinder the visibility of the original texture. In this figure we still easily see the diagonal lines that we rendered on the display. On the other hand, in the case of low frequency distortions as seen in Figure 7c our original signal is lost. In some cases we will even see signals that did not exist in the test signal but were created by the display and became dominant. Based on the above discussion, we determined criteria in the frequency domain for estimating if a signal of a particular frequency will or will not be properly represented on the display. The overall procedure implementing the criteria can be summarized in the following four steps[1]:

First, we calculate the spectrum (magnitude of the 2D DFT) of a test shot. Due to various properties of the display, the spectrum of the test shot is very different from the spectrum of the input image. As an example, the spectra for shots shown in Figure 7b and Figure 7c are shown in Figure 8c and Figure 8d, respectively. For comparison, the spectra of the corresponding input signals are shown in Figure 8a and Figure 8b. Since disparity corresponds to shifts in the spatial

[1] This procedure has been originally introduced in [12] for evaluating the frequency behavior of an autostereoscopic display at zero disparity. Here we will repeat it for completeness together with some additional clarifications and observations

domain and the magnitude of the DFT is shift invariant, the spectrum (magnitude of DFT) does not depend on the disparity. Therefore, Figure 8a and Figure 8b are identical. On the other side, in both spectra of test shots, Figure 8c and Figure 8d, there are many dominant components. They appear due to the optical effects of the display (optical layer) as it was discussed in Section 2. However, as mentioned before, most of those are high frequency distortions that we can ignore since they will be partially masked by the contrast sensitivity function of the HVS. Moreover, we are not able to do anything about them since they are always present in a multiview display.

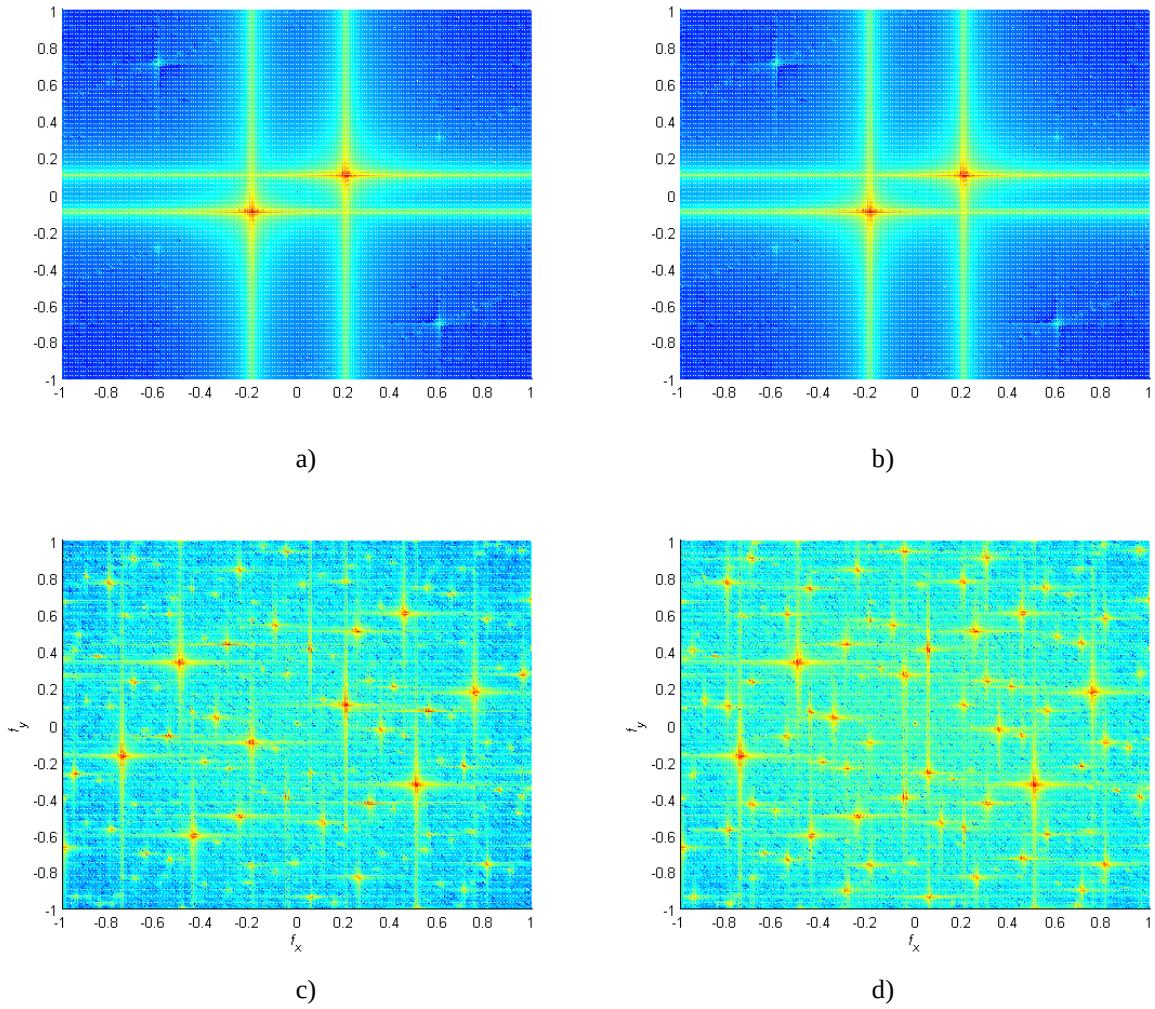
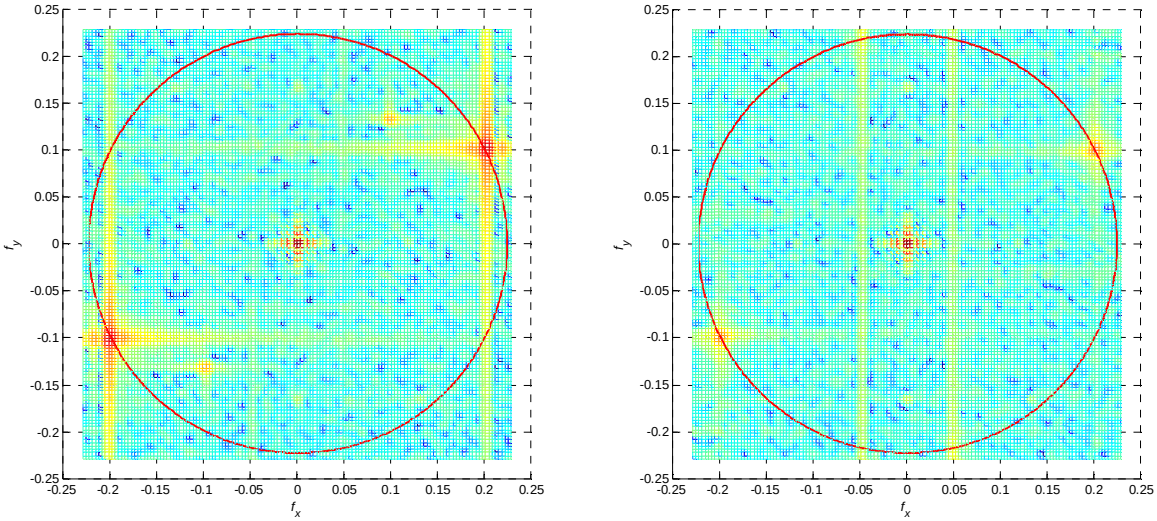


Figure 8 Spectra of signal $f_x = 0.2$, $f_y = 0.1$ at various stages. a) input signal for $d = 1$, a) input signal for $d = 5$, c) test shot for $d = 1$, d) test shot $d = 5$.

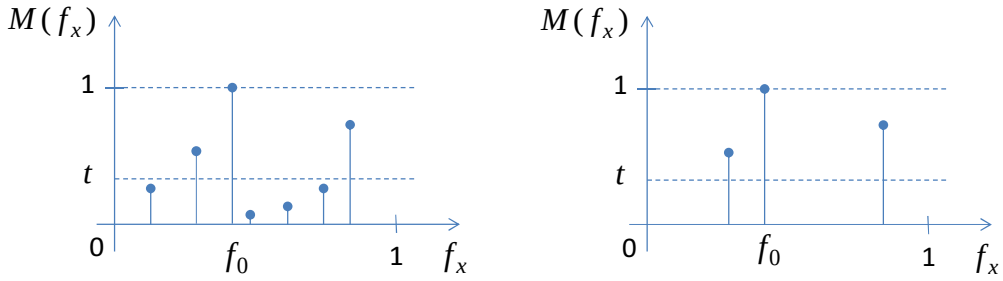
Second, based on the observation discussed at the beginning of this section, from the distortion viewpoint, we are only interested in the area containing frequencies lower than the frequency of the input signal. These frequencies lie inside a circle with the center at DC and radius $r_0 = \sqrt{f_{x_0}^2 + f_{y_0}^2}$ with f_{x_0} and f_{y_0} being the frequencies of the input signal in horizontal and vertical direction, respectively. Zoomed detail of the spectra given in Figure 8c and Figure 8d is shown in Figure 9a and Figure 9b, respectively.



a) b)

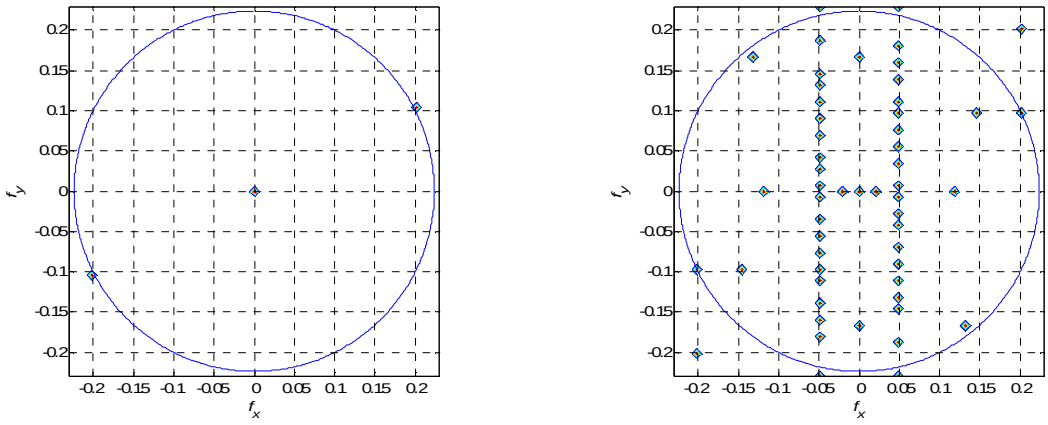
Figure 9 Zoomed spectra of test shots for $f_x = 0.2$, $f_y = 0.1$ in the area of interest. a) $d = 1$, b) $d = 5$.

Third, as seen in Figure 9, there are many different signal components present in both test shots. In practice, many of those will not be visible because the amplitude is too small. Therefore, we have to threshold the spectra, that is, determine when a component is significant and when not. This is directly related to the visibility of various distortions as discussed in Section 2.3. Although, in Section 2.3 the distortion criteria were stated in the time domain, due to the fact that the DFT is a linear transform we can directly apply the same thresholds in the spectral domain. Moreover, if the magnitude of the intended signal is scaled to one, then no additional scaling is required. In the evaluation, we assume that every component that is below the threshold does not contribute to the output signal (will not be visible based on the desired criteria) and therefore we ignore it. This is illustrated by means of a simple 1D example in Figure 10. In this figure, f_x is the sampling frequency in one direction, $M(f_x)$ is the magnitude of the 1D DFT, t is the threshold and f_0 is the frequency of the intended signal with magnitude scaled to 1. After applying the threshold, the original spectra in Figure 10a becomes as shown in Figure 10b. As seen from the figures, all frequency components with magnitude less than t are removed from future analysis. Similarly, after applying the threshold of 5% on the spectra of Figure 9 the thresholded spectra are shown in Figure 11. In this figures, for a better visualization, only the centers of the peaks are shown.



a) b)

Figure 10 A simplified 1D example of tresholding in the spectral domain. a) Spectra before tresholding. b) Spectra after tresholding (all components below the threshold level t have been removed).



a) b)

Figure 11 Spectra of test shots for $f_x = 0.2$, $f_y = 0.1$ in the area of interest represented by the circle after a 5% threshold has been applied. a) $d = 1$, b) $d = 5$.

Fourth, if after applying the threshold there are no signals left with frequencies lower than the input signal, then we assume that signal of this frequency is represented properly on the screen. Consequently, we declare that this frequency is in the passband of the display. This is illustrated in Figure 11a. Since after tresholding, there are not any components left inside the area of interest (marked by circle with radius r_0), this signal ($f_x = 0.2$, $f_y = 0.1$ and $d = 1$) will be properly represented on the display. On the other hand, if there are one or more signal components left, the image on the display will be considerably distorted. Those frequencies we declare as stopband. This is illustrated in Figure 11b. Since after tresholding, there are several components left inside the area of interest, this signal ($f_x = 0.2$, $f_y = 0.1$ and $d = 5$) will not be properly represented on the display.

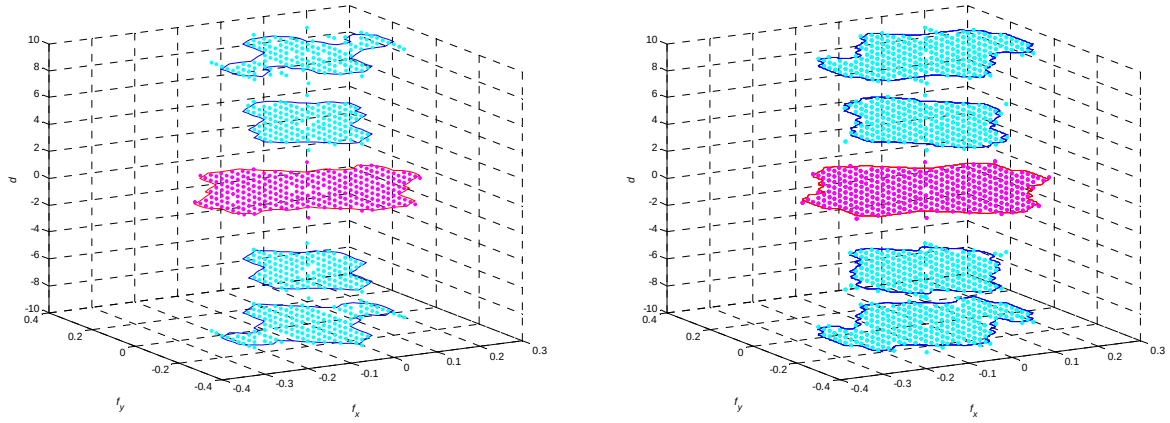
4.2 Calculation of display passband

We repeat the above procedure for all shots (for all input frequencies and all disparities). This results in data describing the passband regions at different apparent depths. Furthermore, for each disparity level, we applied a 3x3 median filter in order to smooth the passband region and remove possible gaps caused by non-ideal measurements. The effect of the median filter is rather positive in filling in gaps and the errors it might introduce are negligible with respect to the subsequent approximation of the filter passband. A filter approximating the measured passband region being of reasonable size will always have quite wide transition band, that is, it will be far away from an ideal one and as such the errors introduced by the filter around the edges of the passband will be bigger than the ones introduced by the median filter.

The passband regions for disparities (d = -10, -5, 0, 5 ,10) and thresholds 5% and 20% are given in Figure 12a) and Figure 12b), respectively. The dots show the evaluated data and the solid line around shows the passband edge after median filtering.

Three observations can be made based on the presented figures. First, the passband form is clearly disparity-dependent. Having the measured pass-bands for different disparities, one can more accurately prepare 3D content to be shown on the display. Second, the passband is dependent on the chosen distortion threshold. The level of 5% corresponds to the visibility threshold and the level of 20% corresponds to a high, but still acceptable, amount of distortions. Thus, measurements at those two levels set up the quantitative compromise between allowing more frequency content to pass versus increasing the amount of visible distortions. In other words, one can design a set of filters ranging from more-restrictive to less-restrictive ones and corresponding to different amount of visible distortions. It can be left to the user's preference to select which filter is to be applied to the watched content. Third, this figure can be used as quality profile.

By comparing the passbands of two displays one can judge which of them is better in representing 3D content within given disparity range. The bigger the area of a passband is, and the closer it is to a square, the better suited the display is for visualizing natural content.



a) b)

Figure 12 Display passband for different disparities d = -10, -5, 0, 5 ,10. a) distortion threshold t = 5%, b) distortion treshold$t$ = 20%.

5. EQUIVALENT RESOLUTION

In this section we introduce the notation of equivalent resolution. The equivalent resolution is a simplified way to interpret the measured passband for a given threshold and given disparity. Referring to the quality profile of the display given in Figure 12, we note that it might be a bit difficult to use it when comparing this display with other displays. In an attempt to find a simplified yet reasonable representation of the shapes, we approximate the passband for each disparity level with a rectangular shape. The approximating rectangle is centered at the origin, has the same area (in size) as the original passband and overlap as many as possible passband points, while keeping the aspect ratio between maximum values in horizontal and vertical direction. In order to do this, the following set of equations has to be solved:

$$\frac{y_m}{x_m} = \frac{b}{a}$$

$$a \cdot b = A\,,$$

where a and b are the horizontal and vertical width of the rectangle, respectively, x_m and y_m are the maximum width and height of the original shape, respectively, and A is the area of the original shape. These parameters are illustrated in Figure 13. After some trivial mathematical operations, a and b, can be evaluated as follows:

$$a = \sqrt{A\frac{x_m}{y_m}}$$

$$b = \sqrt{A\frac{y_m}{x_m}}\,.$$

As an example, Figure 13 shows the approximation for zero disparity and 5% threshold.

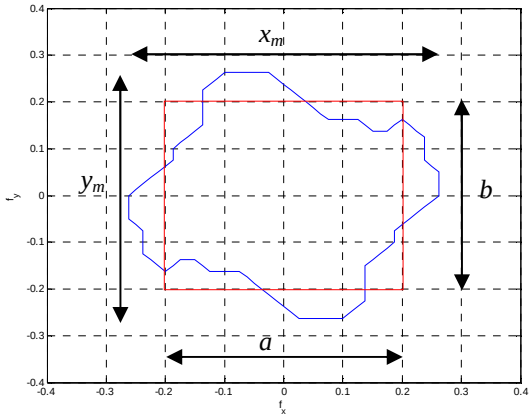
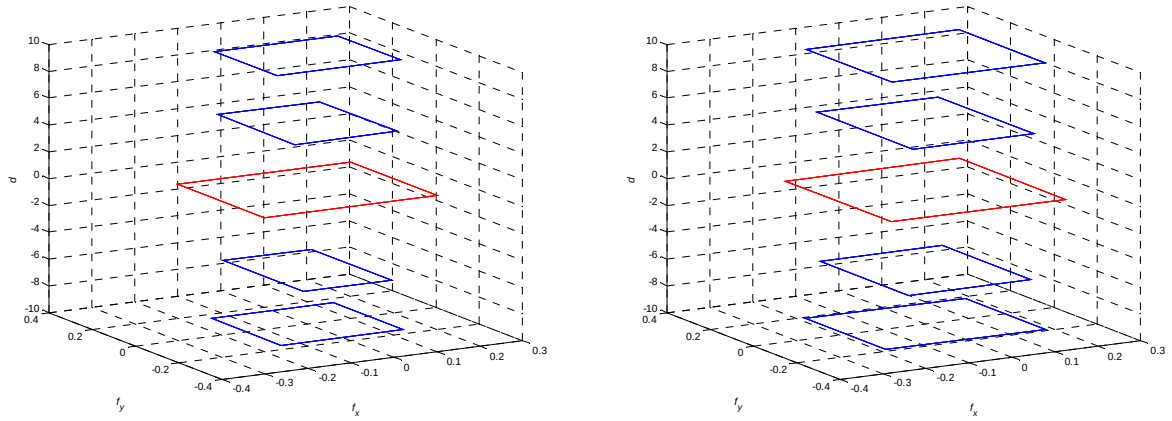


Figure 13 Fitting rectangle to the passband. Example for t = 5% and d = 0.

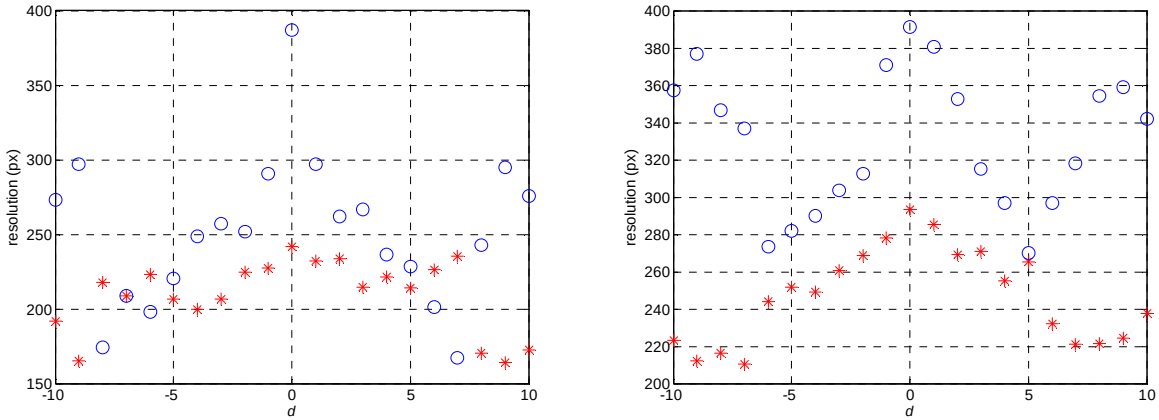
By fitting rectangles for all disparities, for the X3D-display, the equivalent passbands for t = 5% and t = 20% are shown, in Figure 14a) and Figure 14b), respectively.



a) b)

Figure 14 Display passbands approximated with rectangles for different disparities d = -10, -5, 0, 5 ,10. a) distortion t = 5%, b) distortion t = 20%.

In order to represent this figure in a more understandable way, we transfer the pasbands in the equivalent resolutions (in number of pixels) in horizontal and vertical direction and plot them with respect to disparity. The equivalent resolution is obtained by multiplying the passband width (height) with the resolution of the display’s TFT-LCD matrix in horizontal (vertical) direction. In the case of X3D-display, the TFT-LCD resolution is 1920 by 1200. The equivalent resolution for the X3D-display for the t = 5% and t = 20% is shown in Figure 15a) and Figure 15b), respectively.



a) b)

Figure 15 Equivalent resolution in horizontal (circle) and vertical (star) direction as a function of disparity. a) t = 5%, b) t = 20%.

The equivalent resolution given in Figure 15 is a simplified representation of the true bandwidth in terms of spatial resolution (number of pixels in horizontal and vertical direction for each disparity level). It is calculated to serve two main purposes. First, it enables a fast and easy comparison between different displays. A display that has a higher equivalent resolution at a given disparity will pass more data and as such will be better. Second, the equivalent resolution is useful when preparing content to be represented on the display. It suggests in an immediate way what the limits in terms of spatial and frequency resolutions are so to avoid preparing images which will be shown improperly on the display. We choose to express the equivalent resolution in pixels, because most users know what visual quality to expect for an image with a resolution given in these units.

6. CONCLUSIONS

In this work, we have drawn a generalized model of a multiview display and used it to explain the reason behind common artifacts, such as aliasing and crosstalk. We have proposed a measurement methodology, which can assess the visibility of these artifacts in patches with different spatial frequency, orientation and disparity. Using these measurements, we have shown how one can derive the display passband for images with different apparent depth. The measurements for display passband versus object disparity can be used for comparing the visual quality of different multiview displays. Additionally, we have given an example about how the display passband can be used to approximate the effective (equivalent) resolution of a multiview display for 3D content with given disparity.

Other comparative studies focus on characterizing the optical quality of a multiview display. In these studies, a large number of parameters of each display are measured and analyzed - e.g. twelve display parameters in [3], six parameters in [8] and four parameters in [7]. Such large variety of parameters allows displays to be characterized in different ways; however it also makes the comparison and choice of a display complex and rather non-intuitive task for display users. In our work, we propose that the display passband is used as (an additional) indicator of perceptual quality of a multiview display. The advantage over other approaches is twofold: first, it is easier to compare two displays - larger and more uniform passband corresponds to a 3D display capable of visualizing a wider range of spatial frequencies; second, it is easier to judge the expected quality for 3D content with given resolution and disparity range - by analyzing the frequency components of a 3D content one can judge if it is suitable for a given display. The measurement results for equivalent resolution versus disparity can be used to optimize content resolution for a given multiview display.

REFERENCES

[1] S. Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

[2] N. Dodgson, "Autostereoscopic 3D Displays," Computer , vol.38, no.8, pp. 31- 36, Aug. 2005, IEEE (2005)

[3] M. Salmimaa, T. Jarvenpaa, "Optical characterization of autostereoscopic 3-D displays", *J. Soc. Inf. Display 16*, 825 (2008)

[4] V. Berkel and J. Clarke, "Characterisation and optimisation of 3D-LCD module design", *in Proc. SPIE Vol. 2653, Stereoscopic Displays and Virtual Reality Systems IV*, (Fisher, Merritt, Bolas, edts.), p. 179-186, May 1997

[5] J. Konrad and P. Agniel**,** "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Process.*, vol. 15, pp. 128-140, Jan. 2006

[6] V. Saveljev, J.-Y. Son, B. Javidi, S.-K. Kim, and D.-S. Kim, "Moiré Minimization Condition in Three-Dimensional Image Displays," J. Display Technol. 1, 347- (2005)

[7] M. Salmimaa, T. Jarvenpaa, "3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays", *J. Soc. Inf. Display 16*, 1033 (2008)

[8] P. Boher, T. Leroux, T. Bignon, V. Collomb-Patton, "A new way to characterize auto-stereoscopic 3D displays using Fourier optics instrument", in *Proc. of SPIE, Stereoscopic displays and applications XX* , 19-21 January 2008, San Jose, California, USA

[9] J. Hakkinen, J. Takatalo, M. Kilpelainen, M. Salmimaa, and G. Nyman, "Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width", J. Soc. Inf. Display 17, 433 (2009)

[10] F.Kooi, A. Toet, "Visual comfort of binocular and 3D displays", Displays, Volume 25, Issues 2-3, August 2004, Pages 99-108, ISSN 0141-9382, DOI: 10.1016/j.displa.2004.07.004

[11] S. Pastoor, "Human factors of 3D images: Results of recent research at Heinrich-Hertz-Institut Berlin" in Proceedings of IDW'95, vol. 3D-7,pp. 69-72, 1995.

[12] A. Boev, R. Bregovic, A. Gotchev, "Measuring and modeling per-element angular visibility in multiview displays", Special issue on 3D displays, Journal of Society for Information Display, vol. 18, no. 9, pp. 686-697, September 2010.

[13] A. Boev, R. Bregovic and Atanas Gotchev, "Design of tuneable anti-aliasing filters for multiview displays", Conference "Stereoscopic Displays and Applications", a part of Electronic Imaging Symposium 2011, San Francisco, CA, USA, January 2011.

[14] A. Jain and J. Konrad, "Crosstalk in automultiscopic 3-D displays: blessing in disguise?", Stereoscopic Displays and Virtual Reality Systems XIV. Edited by Woods, Andrew J.; Dodgson, Neil A.; Merritt, John O.; Bolas, Mark T.; McDowall, Ian E.. Proceedings of the SPIE, Volume 6490, pp. 649012 (2007).

[15] Christian N. Moller and Adrian R. L. Travis. 2005. Correcting Interperspective Aliasing in Autostereoscopic Displays. IEEE Transactions on Visualization and Computer Graphics 11, 2 (March 2005), 228-236. DOI=10.1109/TVCG.2005.28 http://dx.doi.org/10.1109/TVCG.2005.28

[16] W. IJsselsteijn, P. Seuntiens and L. Meesters, "Human factors of 3D displays", in (Schreer, Kauff, Sikora, edts.) 3D Video Communication, Wiley, 2005.

[17] A. Boev, R. Bregovic, A. Gotchev, "Methodology for design of antialiasing filters for autostereoscopic displays", Special issue on Advanced Techniques on Multirate Signal Processing for Digital Information Processing, Journal of IET Signal Processing, to be published (2010)

[18] Shree K. Nayar, Vlad Branzoi, Terry E. Boult, "Programmable Imaging Using a Digital Micromirror Array," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, pp. 436-443, 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 1, 2004

[19] B. Wandell, “Foundations of Vision”, Sinauer Associates, 1995

[20] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, IEEE Trans. Image Processing, vol. 13, No. 4, 2004, pp. 600-612

[21] S. Winkler, “Perceptual Video Quality Metrics – A review”, in H. Wu and K. Rao, eds. “Digital video image quality and coding”, ch. 5, CRC press, 2006.

[22] E. Montag and M. Fairchild “Fundamentals of Human Vision and Vision Modelling”, in H. Wu and K. Rao, eds. “Digital video image quality and coding”, ch. 2, CRC press, 2006.

[23] Maria Amparo Díez-Ajenjo, Pascual Capilla, Spatio-temporal Contrast Sensitivity in the Cardinal Directions of the Colour Space. A Review, Journal of Optometry, Volume 3, Issue 1, 2010, Pages 2-19, ISSN 1888-4296, 10.3921/joptom.2010.2.

[24] L. Wang, K. Teunissen, T. Yan, C. Li, Z. Panpan, Z. Tingting, I. Heynderickx, "Crosstalk Evaluation in Stereoscopic Displays," *Display Technology, Journal of* , vol.7, no.4, pp.208-214, April 2011 doi: 10.1109/JDT.2011.2106760

[25] P.J.H. Seuntiëns, L.M.J. Meesters, W.A. IJsselsteijn, “Perceptual attributes of crosstalk in 3D images”, *Displays*, Volume 26, Issues 4-5, October 2005, Pages 177-183, ISSN 0141-9382, 10.1016/j.displa.2005.06.005. (http://www.sciencedirect.com/science/article/pii/S0141938205000429)

[26] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in Proc. SPIE Photonics West 2002: Electronic Imaging, vol. 4660, pp. 212-221, 20023D

[27] X3D-23” Users’ Manual. NewSight GmbH. Firmensitz Carl-Pulfrich-Str. 1 07745 Jena, 2006