



Universität des Saarlandes
Max-Planck-Institut für Informatik



Effects of cross-talk on perceived depth in stereoscopic and automultiscopic displays

Masterarbeit im Fach Informatik
Master's Thesis in Computer Science
von / by

Waqar Khan

angefertigt unter der Leitung von / supervised by
Karol Myszkowski

betreut von / advised by
Belen Masia

begutachtet von / reviewers

Saarbrücken, December 2015

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement in Lieu of an Oath

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Saarbrücken, December 2015

Waqar Khan

Abstract

Stereoscopic and automultiscopic displays suffer from crosstalk. An undesired effect which greatly reduces image quality, viewer comfort and distort the perception of depth. Previously, only a limited work has been done on understanding the relation between crosstalk and the perceived depth with respect to stimuli of different nature in stereoscopic displays. To the best of our knowledge, no such work has been done for automultiscopic displays. Moreover, most of the previous work is carried using simple monochromatic scenes. Since the human visual system uses numerous cues other than disparity to estimate the depth of an object in a stereo scene, monochromatic scenes are poor choice for understanding the aforementioned effects. We, in this work, perform experiments to better understand the effects that crosstalk might have on the perceived depth of stimuli in both, stereoscopic and automultiscopic displays. In order to obtain an accurate understanding, we perform experiments on rendered stimuli of natural scenes consisting of objects of various geometries. The model for human visual system's depth estimation via disparity as provided by the current literature fails to justify why and how the perceived depth is affected by the crosstalk. Based on the result of our experiments, we propose a modified human visual system's depth estimation model that, while estimating the viewer's observed depth of a stereoscopic stimulus, takes the ghosting in consideration as well. Finally, some improved techniques for compensation of crosstalk in automultiscopic displays are proposed.

Acknowledgements

I would like to express my gratitude to my supervisor Karol Myszkowski for introducing me to this topic, and for his remarks and his engagement through the learning process of this master thesis. Furthermore, I would like to thank Belen Masia for her immense support and guidance without which the thesis would not have been completed. Also, I would like to thank the participants in my experiments, who have willingly shared their precious time. I would like to thank my loved ones, who have supported me throughout entire process, both by keeping me motivated and helping me putting pieces together.

Contents

Abstract	v
Acknowledgements	vii
Contents	xi
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 Contribution	2
1.2 Structure	2
2 Relevant Background	5
2.1 Binocular vision, stereopsis and its limits	5
2.2 Crosstalk	8
2.3 Stereoscopic and automultiscopic screens and its crosstalk	11
2.3.1 Time sequential stereo using active shutter glasses with an LCD display	11
2.3.2 Polarized stereo	14
2.3.3 Automultiscopic Screens	16
3 Related work	21
3.1 Effects of crosstalk on perceived depth	21
3.2 Depth resolution mechanisms in HVS	24
3.3 Reduction of crosstalk	25
3.3.1 Stereoscopic displays	26
3.3.2 Automultiscopic displays	28

4 Crosstalk Experiments	31
4.1 Motivation	31
4.2 Stimuli	33
4.3 Apparatus and simulation procedure	35
4.4 Experimental procedure	37
4.4.1 Environment setup	37
4.4.2 Subjects	39
4.4.3 Experiment parameters	40
4.5 Results	40
4.5.1 Stereo	41
4.5.2 Automultiscopic	44
4.6 Discussion	47
4.6.1 Stereo	47
4.6.2 Automultiscopic	50
5 Applications	53
5.1 Observed depth prediction	54
5.2 Crosstalk mitigation	60
5.2.1 Unsharp masking in epipolar domain	60
5.2.2 Iterative crosstalk reduction	62
6 Conclusion	65
6.1 Summary	65
6.2 Future work	66
Bibliography	67
Index	70

List of Figures

2.1	HVS Depth Cues.	6
2.2	Representation of theoretical (T) and empirical (E) horopter.	7
2.3	(a-c) A simulation of 14% crosstalk. An image intended for the left eye containing 14% of image intended for right eye. (d) Illustration of viewers crosstalk with the transfer functions as defined by Woods [41].	9
2.4	A flow diagram showing the process of stereo image perception starting from stereo scene capture. The crosstalk can be induced by the display (light generation) and by the image/view separation mechanism(3D glasses or autostereoscopic parallax barrier).	12
2.5	(a) Response of a conventional LCD display (in time domain). (b) Response of a high refresh rate LCD display [41].	12
2.6	(a) A layout of a micro-polarized LCD screen where odd and even numbered rows are oppositely polarized. In this example, the viewer viewing the screen using the polarized glasses will see odd numbered pixel rows through the right eye and even numbered pixel rows in the left eye [41] (b) Construction of a typical micro-polarized LCD screen.	15
2.7	A five-view automultiscopic screen using a parallax barrier. The actual pixels on the LCD panel are called sub-pixels whereas the gaps in the parallax barrier can be thought of as a single view dependent pixel. Each of the observer's eyes can only see a specific set of sub-pixels through a view dependent pixel [42].	17
2.8	(a)Visible pixels observed through the slit of a tilted parallax barrier. It can be seen that the view is covering more than one pixels for every pixel. (b)The view that covers majority portion of a sub-pixel is assigned to it which is the reason why some light from that particular view will be leaked into the adjacent view [29]	18
2.9	The experimentally observed luminance distribution of all views across different view points. It can be seen that some portion of the adjacent views is leaked even at the sweet spots. The light leakage increases as the observation point is moved away from the sweet spot [41].	18
3.1	(a) The setup of Tsirlin's experiments [26] [24]. Left: complete experiment where the viewers see the vertical lines as a single line in depth. Right: the same line with 0, 16 and 32% simulated crosstalk. (b) Depth degradation averaged over the test subjects.	22
3.2	(a) 1D cross-correlation with varying contrast. (b) 1D cross-correlation with variable level of added noise [5].	25

3.3	(a) Input image. (b) Subtractive compensation. (c) CSF weighed compensation. (d-f) Observed images [27].	27
4.1	(A) Frontal view of the scene rendered with Blender. (B) Side view of the same scene.	34
4.2	Stereo images of the cylinder of 5 cm radius and the dragon. The theoretical depth difference between the objects and the background is 3.44 cm. Viewer can cross-fuse to see in depth.	35
4.3	Our Wheatstone setup. Physical view frusta are shaded in light blue where as the virtual view frusta are depicted in dotted gray lines. The cyan lines represent the mirrors [28]	36
4.4	View-luminance intensity profiles for an 11-view automultiscopic display simulated with Gaussians of $\sigma = 2$ centered at the sweet spots. X-axis denote the viewing position. It can be seen that the simulated luminance profiles closely matches the experimentally observed view-luminance profiles in Figure 2.9	37
4.5	(a,b) Observed left and right stereo images with 14% crosstalk and crossed depth of 3.45 cm. (c,d) Same configuration for observed images on automultiscopic display). Viewers can cross-fuse to see in depth.	38
4.6	(A) Schematics of the complete display. Upper half of the screen shows the reference stimulus, in this case, with 14% crosstalk. The bottom half shows the user-controlled test stimulus which is always free of crosstalk. (B) Layout of the experiment environment in terms of depth w.r.t. observer.	39
4.7	The median data for all observers is shown. The abscissa shows the observed depth when no crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	41
4.8	The median data for all observers is shown. The abscissa shows the observed depth when 3% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	42
4.9	The median data for all observers is shown. The abscissa shows the observed depth when 7% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	43
4.10	The median data for all observers is shown. The abscissa shows the observed depth when 14% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	44
4.11	The median data for all observers is shown. The abscissa shows the observed depth when no crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	45
4.12	The median data for all observers is shown. The abscissa shows the observed depth when 3% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	46
4.13	The median data for all observers is shown. The abscissa shows the observed depth when 7% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	46

4.14	The median data for all observers is shown. The abscissa shows the observed depth when 14% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.	47
5.1	The figure represents estimating the human observed disparity in a thin cylinder stereo pair where the disparity in terms of pixels is 9 and the crosstalk is 14%. The cross-correlation of a patch containing only the cylinder from the right eye image is computed across the entire left eye image containing both the cylinder and the ghost. (A) The blue line represents the 1D scan line of the result of local cross-correlation. In this case, the maximum cross-correlation occurs at the disparity of 9 pixels, which is the actual disparity. Orange line represents a Gaussian function with $\sigma = 9$ pixels centered at the location of the ghost that is to be used for weighing the correlation profile. (B) The resulting Gaussian weighed local cross-correlation. It can be seen that, the estimated disparity as a result of the peak of cross-correlation has shifted back from 9 to 3 pixels.	56
5.2	Comparison of experimental observed depth for cylinder of 18.9 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 15.4$ arc min.	57
5.3	Comparison of experimental observed depth for cylinder of 37.8 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 30.87$ arc min.	58
5.4	Comparison of experimental observed depth for cylinder of 56.7 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 36.54$ arc min.	58
5.5	Comparison of experimental observed depth for cylinder of 75.6 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 46.2$ arc min.	59
5.6	Representation of a 3D lightfield [6]. (A)(x,y) denotes the view dimensions while w denotes different perspective views for a light field. (B) Epipolar view of the squared region in (a)representing how every pixels shifts with respect to shift in perspective.	60
5.7	(A) 1D representation of luminance profile along an image row. Blue graph represents a luminance step and orange graph represents the luminance profile of the blurred image. Yellow graph represents corn-sweet profile centered at the luminance edge and obtained as the difference between the original and the blurred image. (B) Same procedure performed in view domain. Blue graph represents the luminance profile of a light field image and orange graph represents the non-energy preserving additive blurring due to neighboring views in an automultiscopic environment. The Yellow graph represents the difference between the view image and its view-domain blurred version that resembles closely to subtractive crosstalk cancellation rather than a corn-sweet profile.	62
5.8	Naive subtractive crosstalk compensation compared to our proposed iterative crosstalk compensation. (A) An automultiscopic display's view image containing crosstalk given by the curves in Figure 4.4. (B) Compensated view image using naive crosstalk subtraction. (C) Compensated view image using iterative crosstalk subtraction.	64

List of Tables

4.1	Description of the objects used as stimuli.	34
4.2	Number of subjects that participated in each experiment.	40
4.3	Ghosts and objects seen in an automultiscopic environment according to Equation 4.3	50

Chapter 1

Introduction

Ever since its commercial introduction to cinema in 1922, the stereoscopic approach to 3D content presentation has waxed and waned in popularity over time. Stereoscopy consists of presenting a different perspective image separately to each eye tricking the Human Visual System (HVS) to see in 3D on a 2D planer flat screen. Current commercial stereoscopic 3D displays, like any other practical system, are not devoid of imperfections. One of these imperfections known as crosstalk, is the inability to separate the different views completely for each eye. This means that some portion of the light from the image intended for one eye leaks to the other eye (and vice versa) that results in dim copies (ghosts) of the unintended images seen along with the intended images. Crosstalk is present in all commercially available 3D displays (LCD TV's and cinemas) and contributes heavily to the viewer's discomfort in the forms of reduced overall image quality and reduced image contrast. More importantly crosstalk also affects the stereopsis of HVS. Unintended depth edges conflict with the intended depth edges, thereby hindering the proper fusion of the stereo images [26]. These conflicts can result in reduced observed depth. Thanks to today's technologically impressive and immensely lucrative 3D movie industry, TV and games, the 3D industry has been seeing a sharp rise in popularity. This increase in popularity has once again motivated researchers to give attention to imperfections in commercial stereoscopy and their effects on viewers. Current methods for crosstalk compensation or ‘deghosting’ typically involve subtracting the unintended ghosts from the intended image before being displayed. This technique, however, either reduces the overall image contrast or does not remove the ghosting completely in high contrast regions of the images.

1.1 Contribution

Although the effects of crosstalk on viewer's observed quality and visual discomfort have been thoroughly studied in the past [40], its effects on the depth perception have received little attention. Since the basic purpose of stereoscopy is to display a 3D scene consisting of different objects located at different depths, any undesired variation in the perceived depths can severely hamper the aesthetics of the observed scene. Thus the main focus of this thesis is to better understand the effects that crosstalk can have on the depth perception on various types of images and 3D display technologies. Previously, some work has been done concluding that crosstalk in general reduces the perceived depth[26], i.e. objects that are distant from the plane of fixation (POF) will tend to fall back to the POF. The user studies that resulted in that conclusion were performed on monochromatic stimuli where disparity was the only cue for depth. The natural scenes usually have a lot more depth cues and hence we hypothesized that the effects of crosstalk might be different for complex stimuli. Hence we, in order to assess the effects of crosstalk of depth perception, firstly performed robust experiments on stimuli of various dimensions and shapes that were rendered images of a 3D scene containing all possible depth cues. Further, there have been no studies analyzing the effects of crosstalk on depth perception in an automultiscopic displays (glass free 3D display). We performed another set of experiments in order to analyze if the effects of crosstalk on depth perception in automultiscopic displays is any different.

We also propose a modified HVS disparity estimation model inspired by the model of Banks et al[7], that helps understand why the crosstalk results in reduced depth of objects that are located away from the plane of fixation. Finally, we propose and test some new techniques for reducing crosstalk in automultiscopic displays.

1.2 Structure

The thesis is assembled as follows. Chapter 2 discuss some background knowledge required in order to fully understand subsequent chapters. This included information about how the HVS works according to the literature, a formal definition of crosstalk and a general idea of how different 3D displays work along with the nature of their crosstalk. Chapter 3 covers some of the previous research that has been performed in order to understand the effects of crosstalk on perceived depth, the possible explanation for this effect, and different techniques that are used in order to mitigate the effect of crosstalk via image preprocessing. Chapter 4 discusses in detail the experiments we performed to quantify how the depth is degraded in both stereoscopic and automultiscopics screens,

along with their results. In Chapter 5, we discuss our proposed modified HVS depth via disparity resolution model that takes the effects of crosstalk into account as well. We also discuss our proposed crosstalk reduction techniques. Finally, in Chapter 6 we will look into some phenomenon that still needs explanation and where the future research could be headed in order to get a better understanding of them.

Chapter 2

Relevant Background

Depth perception is the ability of the Human Visual System to visualize the three dimensional world as well as measuring the distance of an object based on two dimensional images obtained from the eyes. Depth perception is imperative for performing basic everyday tasks such as avoiding obstacles without bumping into them or interacting with the world with relative ease. In animals (specially predators), it is critical to estimate the distance of a prey for an efficient attack. Depth sensation is the term used for animals as it is not known whether they sense the depth in the same way as humans do [33].

The HVS uses several cues to determine the depth of objects in sight. These cues can be categorized into two categories, i.e., cues extracted from a single image (monocular cues), and cues extracted from two images (binocular cues) [20] [33]. Figure 2.1 gives an outlook of the depth cues used by the HVS. These cues are then dynamically weighted by the HVS in order to estimate the depth of each object in the field of view [16].

2.1 Binocular vision, stereopsis and its limits

Generally speaking, all the animals with two eyes have binocular vision and they can integrate the information from two eyes based on the binocular overlap. But the term binocular vision is usually used for the animals that have a large area of binocular overlap (i.e. human and most other predators) and use it to get the depth information of the world around them. In addition to calculating depth, binocular vision also has advantages in performing other tasks such as detection, discrimination, detecting camouflaged objects or eye-hand coordination. Even the resolution of the observed world is increased with binocular vision [9]. Among all the depth cues discussed in the section above, stereopsis is the most influential of them all. Since the human eyes are located at different lateral

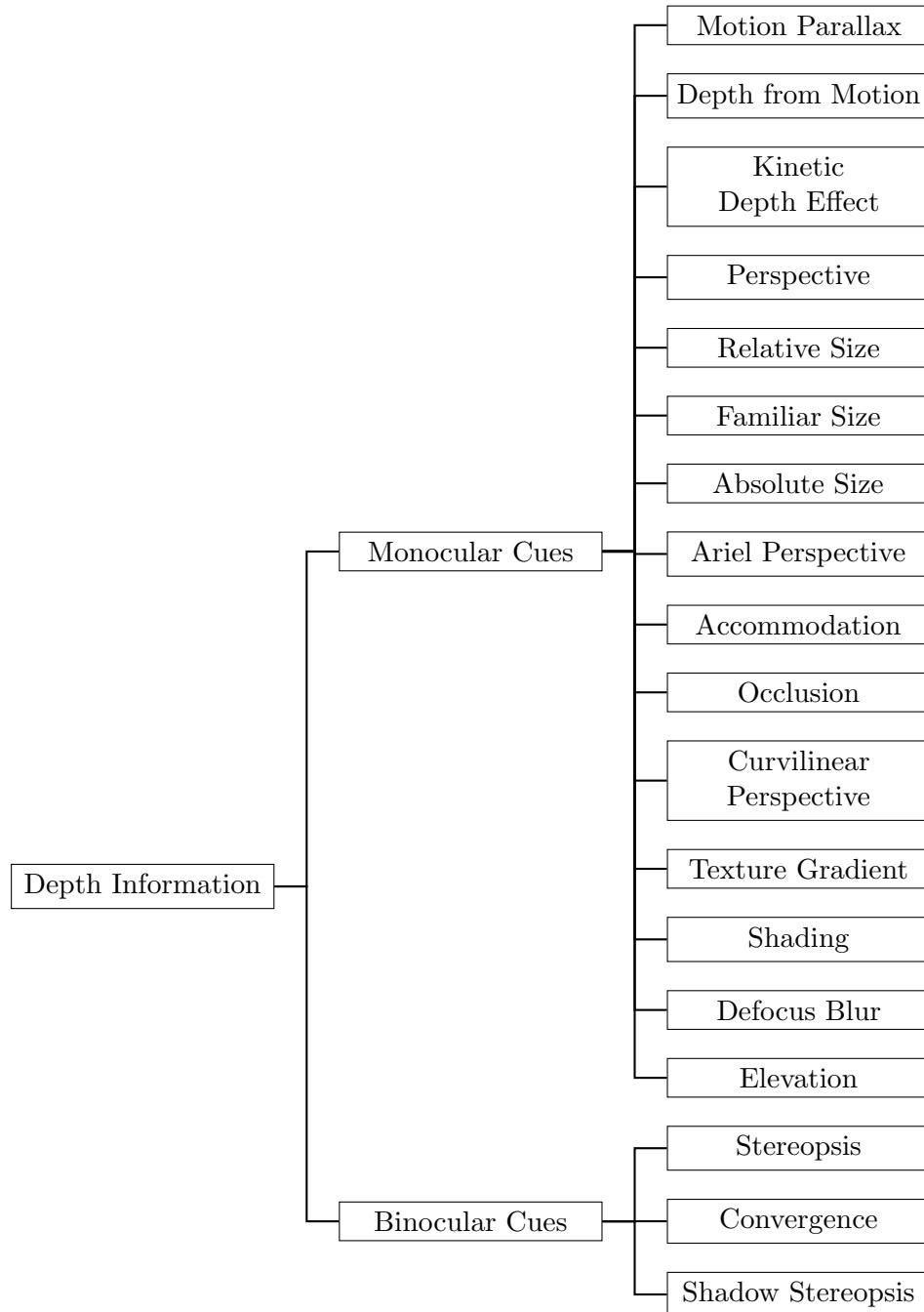


FIGURE 2.1: HVS Depth Cues.

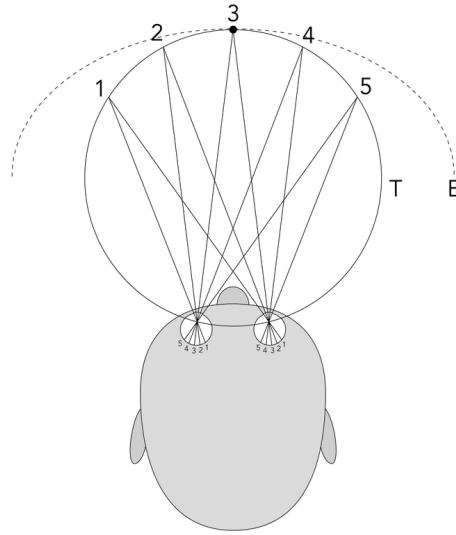


FIGURE 2.2: Representation of theoretical (T) and empirical (E) horopter.

positions on the head, the images formed on the retinas of these two eyes are slightly different. The difference is mainly the horizontal positions of the objects known as disparity [38]. The process of obtaining a fused image (Cyclopean image) and obtaining a depth map based on the horizontal disparities of the objects in these two images is known as stereopsis.

When the eyes verge in order to focus on some object (or point) in space, that object is projected at identical corresponding points in the retinas. It means that the difference between their horizontal positions is zero. This process is called fixation of the eyes and the distance of the object (point) at which the eyes are fixated is called the fixation distance. The locus of all the points in space that is projected on identical retinal points is called the horopter [34]. Theoretically, via geometrical principles, the horopter is a circular segment that passes through the fixation point. However, Wheatstone in 1838 observed that the actual/emperical horopter is much larger than that. Figure 2.2 shows both the theoretical and empirical horopter. Any object that is farther away from the horopter has uncrossed disparity in the retinas, i.e., the eyes need to be diverged (uncrossed) in order to fixate on that object. Similarly, any object that is closer than the horopter has crossed disparity in the retinas, i.e., the eyes need to be converged (further crossed) in order to fixate at it.

Stereopsis is believed to be processed in the binocular neurons of the visual cortex of mammals. The binocular neurons have receptive cells in different horizontal positions in each eye. These cells are active only when the object of interest is in certain range of disparity in one eye relative to the other, i.e., there is a maximum disparity limit. As the

objects in the images formed at the retinas of both eyes are slightly shifted horizontally, presenting two different images with shifted object to both eyes can fool the HVS into perceiving depth. This process is called stereoscopy. The first stereoscope was invented by Sir Charles Wheatstone in 1838 [39]. It used two mirrors both tilted at 45 degrees with respect to the eyes that reflected two different images from the sides. Currently all the stereoscopic screen present a different perspective image to each eye with different technologies that will be discussed in later sections.

2.2 Crosstalk

As discussed in the previous section, stereoscopy includes the process of displaying different perspective images to each eye in order to mimic the effect of depth in a scene. However, it is critical that each of these two images be segregated completely from each other. Currently, all the commercially available 3D displays (with the exception of head mounted displays or Wheatstone setups) fail to isolate the two images completely. This means that a percentage of the image intended for any eye leaks to the other eye as well. This unintended leakage is responsible for crosstalk in stereoscopic screens [41]. Figure 2.3 shows one such example where the screen has a simulated crosstalk of 14% which essentially means that 14% image intensity of the right eye image is leaked into the left eye image and vice versa.

System crosstalk is the term used to define the amount of light leakage that occurs between two views and is independent of the image contents. In simplest form, it can be mathematically defined as:

$$\text{Crosstalk}(\%) = \frac{\text{leakage}}{\text{signal}} * 100 \quad (2.1)$$

Here, *signal* is the luminance of the original image intended for an eye and *leakage* is the luminance of the light that leaks from the unintended image. Usually, on a stereoscopic screen, the amount of crosstalk is measured by observing the luminance of the left channel while a black (minimum luminance) image is displayed in the left channel and a white (maximum luminance) image is displayed on the right channel and vice versa. This definition is accurate for the screens that can manage to display a true black, i.e., zero luminance. Almost all of the LCDs (with exception of AMOLED displays) fail produce zero luminance as their minimum luminance, hence resulting in a non-zero black level at its minimum. Another mathematical representation of crosstalk takes into account this non-zero black level and is defined as:

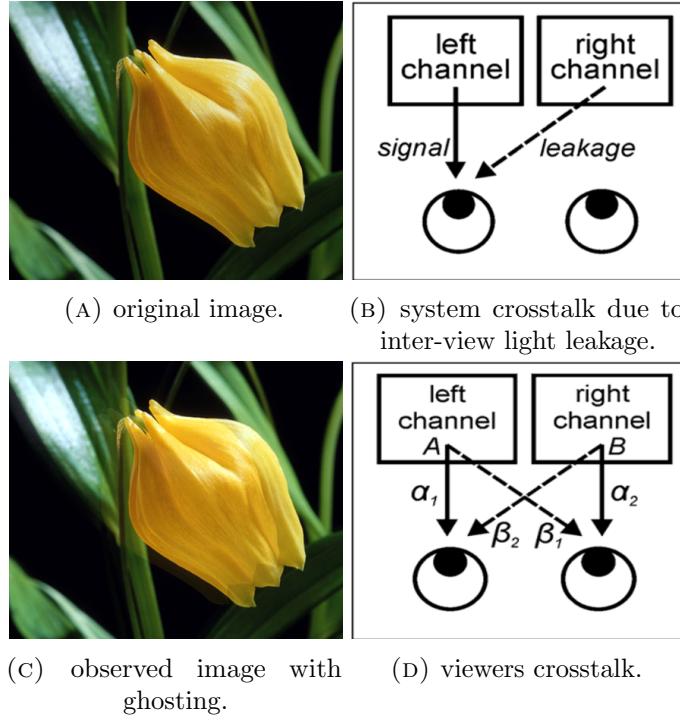


FIGURE 2.3: (a-c) A simulation of 14% crosstalk. An image intended for the left eye containing 14% of image intended for right eye. (d) Illustration of viewers crosstalk with the transfer functions as defined by Woods [41].

$$\text{Crosstalk}(\%) = \frac{\text{leakage} - \text{blacklevel}}{\text{signal} - \text{blacklevel}} * 100 \quad (2.2)$$

Throughout the rest of the thesis, we will be using crosstalk as defined by eq 2.2.

Viewers crosstalk, on the other hand, is the amount of crosstalk that can be perceived by the viewer as ghosts. It is dependent on the image contents, i.e, the contrast at the ghosting point and the parallax of objects in the scene. If the system crosstalk is defined as

$$\text{System Crosstalk (left eye)} = \frac{\beta_2}{\alpha_1} \quad (2.3)$$

Where α_1 denotes the percentage part of the left-eye image at position (x,y) as observed by the left eye and β_2 denotes the percentage amount of right eye image at the same location (x,y) leaked into the left eye. Then the viewers crosstalk is defined as

$$\text{Viewers Crosstalk (left eye)} = \frac{B\beta_2}{A\alpha_1}. \quad (2.4)$$

Where A is the luminance of that particular point in left-eye image and B is the luminance of that particular point in right-eye image (as described in Fig 2.3d). The variables α and β are characterizing the transfer functions from the displayed image to the observed image, i.e., the amount of light reaching the eyes after being displayed on the screen and

going through the glasses or any other medium that resides between the screen and the eyes. In most displays, crosstalk is an additive process and is roughly linear. This means that, to simulate crosstalk, adding a desired amount of unintended image to an intended image should be sufficient. The simulation of crosstalk in our experiments was carried out in the same manner.

The perception of crosstalk obeys the Webber's law¹ which means that leaked light will be greatly perceivable by the viewer on dark image areas rather than on bright areas. Also, the perceived crosstalk is dependent upon the contrast of the image and the binocular disparity of the stimuli i.e. crosstalk perception will increase with an increase in contrast or an increase in binocular parallax. It is commonly believed that ghosting plays the most critical role in determining image quality. Wilcox and Stewart [40] observed that over 75% of the observers in their experiments reported ghosting to be the key feature that deteriorated the image quality in stereoscopic displays. Apart from reduction of image quality, perceivable ghosting is also responsible for loss of depth, viewer's discomfort, reduction of sharpness and contrast, decreased fusion limits, and difficulty in fusion. The main reason for viewer's visual discomfort seems to be the fact that the ghost images in stereoscopic displays are always in diplopia², i.e., unfused state³. This is because the disparity gradient⁴ between the ghost and the stimulus object is far greater than HVS's threshold. Two objects in HVS can not be fused together if the angular separation between them is smaller than the disparity [4]. In fact since the ghost and the stimulus lie in the same (x, y) position but at different depth, the angular separation becomes zero and the disparity gradient of the ghost and the stimulus becomes ∞ . Hence both of them can not be fused together. In this case the more luminant object, i.e., the stimulus will be fused while the ghosts will remain in diplopia state. Crosstalk in still images is perceived to a larger extent than crosstalk in dynamic scenes (e.g. a movie). This means that motion of objects in a scene can mask the perception of crosstalk.

The stereoscopic literature provides many advices on the acceptable and unacceptable viewer's crosstalk. Woods [41] summarizes some of these advices as:

- Crosstalk between 2% to 6% significantly affects the visual quality and increase the viewers discomfort.
- In order to produce accurate depth range between 40 arcmin, the crosstalk should be as low as 0.3%.

¹Webber's law states that the just noticeable difference in a stimulus is a constant ratio of the original stimulus.

²Diplopia in vision occurs when HVS is unable to fuse two sub-images together. In this case the viewer sees two copies rather than one of the objects.

³This was reported by most of the test subjects in our experiments.

⁴Disparity gradient is defined as the ratio between disparities of two objects and the angular separation between them.

- Crosstalk is visible even at 1% to 2%.
- 5% of crosstalk is enough to induce discomfort.
- JND for crosstalk is 1%.
- 2-4% of crosstalk can significantly decrease the amount of perceived depth.

As one can observe, there is a variability in these guidelines. The reason for this variability might be because of the different setups and types of stimuli that were used by the researchers in their experiments. We observed that currently the literature is not quite thorough on how the depth perception is affected when it comes to different kinds of stimuli and relatively complex scenes. Which is why we performed some experiments of our own to verify and expand the current knowledge in this area, and is also one of the main contributions of this thesis.

2.3 Stereoscopic and automultiscopic screens and its crosstalk

In this section we will review some of the stereo technologies and their associated crosstalk. The basic setup for a stereoscopic display involves a display screen that has typically higher than average (above 120 Hz) refresh rate and a view separation mechanism. A high refresh rate is required so that images from two different view perspective can be displayed alternatively without the user realizing any glitches (after the views have been separately delivered to the eyes). Various stereo display technologies are available in the market such as CRT, DLP, plasma, PDP and LCD screens. Active/Passive 3D glasses or anaglyph glasses are generally used as view separation mechanisms. Figure 2.4 illustrates that the crosstalk is induced by the display as well as the view separation mechanisms.

2.3.1 Time sequential stereo using active shutter glasses with an LCD display

LCD screen coupled with active shutter glasses is the most common 3D technology that is commercially available. It produces an image by back-lighting a two dimensional individually addressable liquid crystal (LC) matrix. The back light is typically a cold cathode fluorescent lamp (CCFL) or light emitting diodes (LEDs). The LC matrix consists of crystals that rotate according to the voltage applied to them. This rotation limits the flow of the light that passes through each cell, hence producing different gray levels. Each pixel of the screen consists of three LC cells coupled with red, green and a blue filter. Hence the color and luminance of each pixel is created by controlling the light that

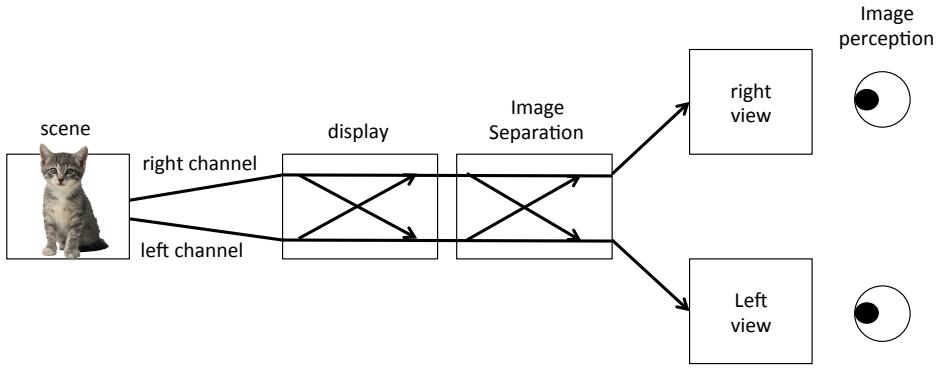


FIGURE 2.4: A flow diagram showing the process of stereo image perception starting from stereo scene capture. The crosstalk can be induced by the display (light generation) and by the image/view separation mechanism(3D glasses or autostereoscopic parallax barrier).

flows through these individual filters which is regulated by the LCs. It should be noted that even at its best, when crystals are perpendicular to the light source, they fail to block the light completely and hence an LCD screen can not display true black color.

The time required for the LCs to rotate from one position to another desired position is known as the pixel response rate or time. This response time is higher if the difference between old and the new rotation value is smaller and vice versa [41]. This means that the refresh rate of an LCD display is dependent upon this response time. LCD displays usually utilize top to bottom image update method, i.e., in order to update the image being displayed, the rows of pixels are addressed individually from top to bottom.

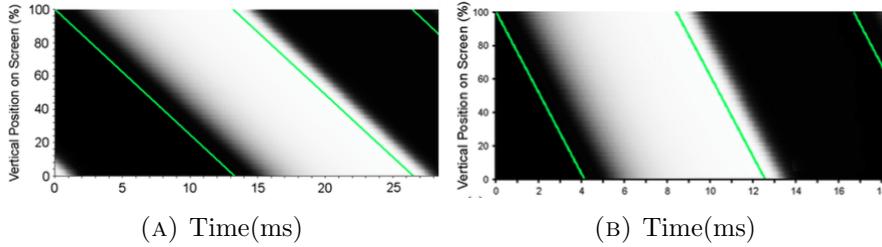


FIGURE 2.5: (a) Response of a conventional LCD display (in time domain). (b) Response of a high refresh rate LCD display [41].

Figure 2.5 shows the time domain responses of LCD displays. The green line represents the row of the pixels that are being updated while the image is being changed from complete black to complete white. It can be seen from Figure 2.5a that there is no time (a vertical line) where a complete black or white image is observed. However if the refresh rate is increased (Figure 2.5b), one can see that there is an interval in time where a stable image can be displayed.

Active shutter glasses consists of lenses that can turn opaque or transparent in order to gate the left or right image being displayed on the screen to the respective left or right eye (at any given time, one lens is in opaque state while the other is in transparent state). Each lens of these glasses has a liquid crystal called LC shutters(just like an LCD display) that blocks the light when a voltage is applied to it. The time required for these LCs to go from completely opaque to completely transparent and vice versa is called the rise and fall time [41]. It is observed that

- LC shutters do not perform identically for each wavelength,
- The LC shutters have a non zero transmission even when it is in opaque state,
- The rise and fall time are not instantaneous.

In addition to that, the light transmission in active shutter glasses also varies with the viewing angle. The highest blocking of the light occurs at an angle perpendicular to the shutters. This means that while viewing a scene through these shutter glasses, one would observe greater leakage of light in the border areas of the shutters as compared to the center (provided the eye is positioned at the center of the shutter). It is important that the shutters are opened and closed at the right timings with respect to the image being displayed on the screen. Incorrect image will be observed if the shutters are opened too early or too late thus adding to crosstalk.

Hence, the methods according to which crosstalk can occur when using active shutter glasses combined with an LCD display are:

- Optical performance of the liquid crystal cells, i.e., crosstalk is proportional to the amount of light leaked while in opaque state.
- Incorrect synchronization of shutter glasses with respect to the display.
- Viewing angle through the shutter glasses.
- Pixel response rate of the LCD display. Higher response time will result in higher crosstalk.
- Image update method. The ideal time for opening of a shutter would be when an image has been completely displayed on the screen. However most LCDs update the image using vertical scanning and hence usually there is no time at which the image is completely displayed stably.
- The (x, y) location of the image on the screen. This is related to both the image update method and the viewing angle through the shutter glasses.

- The gray level being displayed. If the change in gray level is small then the response time will be large hence causing larger crosstalk.

2.3.2 Polarized stereo

Polarization is a property dealing with the controlled oscillation of light (and other waves). Typically light waves oscillate in a way that their phase shifts over time are unpredictable. Passing the light through special polarizing filters forces them to oscillate in a controlled manner [37]. Light can be polarized in a linear manner where the light waves are forced to oscillate in a particular plane or in a circular manner if the orientation of the oscillations vary circularly with respect to time. In stereoscopy, light intended for the left and right eye can be encoded using linear polarization, i.e, the polarization of the left eye light is orthogonal to the polarization of the right eye light. Alternatively, the polarization of one view is clockwise and anticlockwise for the other view in case circular polarization is used.

In a typical 3D cinema setup, images for the left and right eye are simultaneously projected on a silver screen by two projectors. Each of these projectors has a polarizing filter (opposite with respect to each other) attached to it. Viewers view the screen through inexpensive polarized glasses that have the exact same polarizing filters in its left and right lenses as the left and right projector. This way each of the polarized lens block the light from the other view, hence giving an impression of 3D. However, imperfections are also present in this setup. Firstly, the polarizing filters mounted on the projectors are not perfect and fail to properly polarize all light wavelengths, and can not be made to be perfectly orthogonal (opposite to each other).[41]. Also, the silver screen is unable to reflect the polarized light from the projectors without distorting the polarization. Different materials have different polarization properties and until now a material that perfectly preserves the polarization for all visible wavelengths has not been discovered. Lastly, it is hard to match the polarization orientation of the projector filters to the filters present in the glasses. For example, in case of linear polarization, the orientation mismatch can easily occur if the viewers head is slightly tilted hence causing crosstalk [8]. Circular polarization however, is less prone to this mismatch as compared to linear polarization and is therefore more commonly used in 3D cinemas. In any case, failing to deliver properly polarized left and right image light to the glasses and any mismatch in the orientation of the glasses with respect to the projector filters will hinder the ability of the polarized 3D glasses to block the light from unintended view completely causing crosstalk.

In summary, the important factors to consider when dealing with the crosstalk in a typical 3D cinema setup are as follows:

- The optical properties of the polarizing filters used in the projectors and the glasses.
- The polarization preserving properties of the screen.
- The mismatch between the orientation of the projection filters with the orientation of the filters present in 3D glasses.

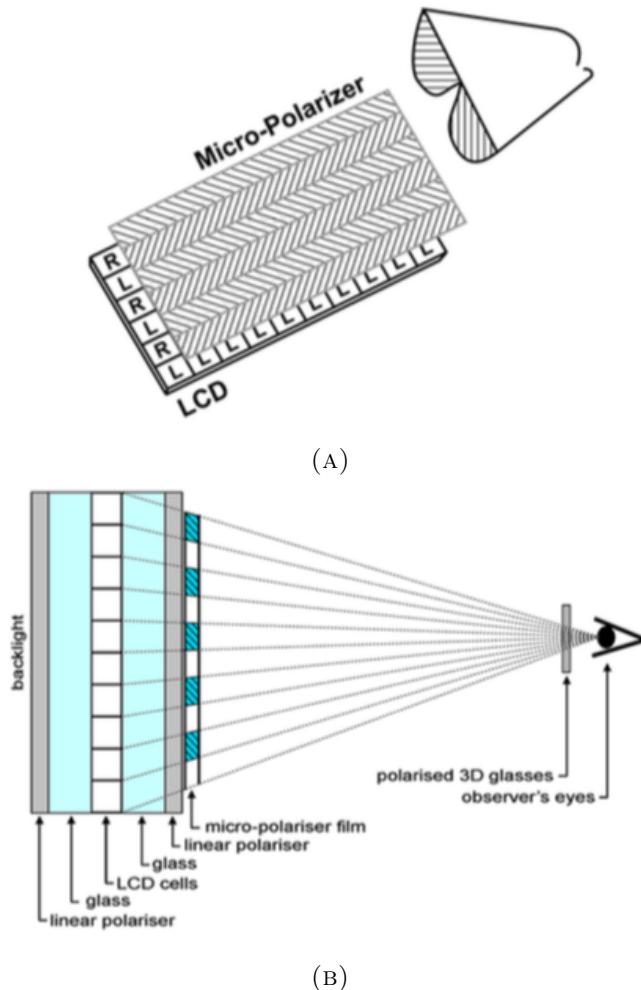


FIGURE 2.6: (a) A layout of a micro-polarized LCD screen where odd and even numbered rows are oppositely polarized. In this example, the viewer viewing the screen using the polarized glasses will see odd numbered pixel rows through the right eye and even numbered pixel rows in the left eye [41] (b) Construction of a typical micro-polarized LCD screen.

A stereo 3D setup can also be obtained by using a micro-polarized LCD screen used in conjunction with polarized 3D glasses that are similar to the cinema setup discussed above. A micro-polarized LCD screen consists of polarization filters (linear or circular) mounted on top of a conventional LCD screen. However, unlike the projective method

where two images can be projected on the screen simultaneously, alternative rows (odd and even rows) of the LCD screen pixels are polarized oppositely to each other. Micro-polarized LCD screens have the advantage over the 3D screens used with active shutter glasses that crosstalk induced due to low refresh rate of the screen is not a problem. But this also comes at a cost of reduced vertical spatial resolution of the displays as the light from half of the row pixels will be blocked by each of the lenses of the polarized glasses. Figure 2.6a illustrates a typical polarized LCD setup where the user will see different pixel rows in each eye. Figure 2.6b shows the construction of a typical polarized LCD screen. It can be seen that the LCD cells are separated from the micro-polarizer film by a glass sheet that is typically 0.5 mm thick [41]. Hence a sensitivity to the viewing distance and position due to parallax is induced. This means that if the viewer is not located at the correct position or distance to the screen, he/she will be able to see the unintended rows of pixels along with the intended rows hence inducing crosstalk. In summary, the factors that contribute to the crosstalk in a micro-polarized LCD screen stereo setup are as follows:

- Matching of the orientation of the polarization filters in the glasses to the orientation of the micro-polarizing film present on the screen.
- The quality and ability of the micro-polarizing film to properly polarize the light.
- The accuracy of the alignment of micro-polarizing film to the LCD cells.
- (x, y) screen position of the screen. Due to the fact that viewer will typically be in a position where the parallax error mentioned above will vary with the screen position.
- Viewing angle of the viewer, as this will affect the matching between the orientation of the glasses and the screen polarization filters.

2.3.3 Automultiscopic Screens

All the stereo display technologies mentioned earlier need some kind of glasses to be worn by the viewer in order to separate the different perspective views for each eye. One of many disadvantages of using glasses is that none of them are perfectly transparent and hence absorbs a portion of incoming light. This makes images darker than they actually are. Automultiscopic displays addresses this problem by placing a (vertical or tilted) parallax barrier or a lenticular sheet on top of a conventional LCD panel that results in only a specific column of pixels to be seen by an eye at some position while the adjacent columns remains occluded as seen in Figure 2.7 [42]. The number of adjacent

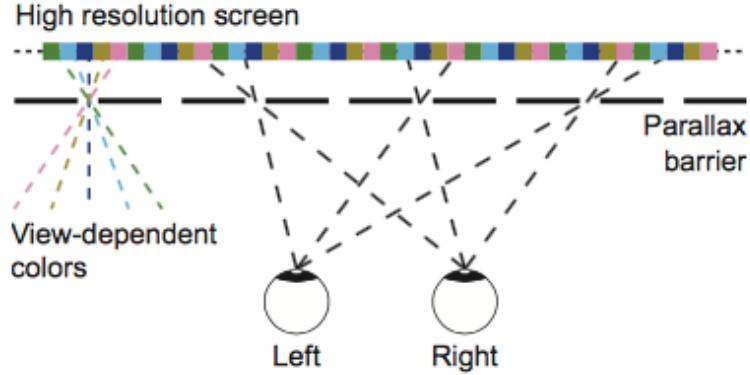


FIGURE 2.7: A five-view automultiscopic screen using a parallax barrier. The actual pixels on the LCD panel are called sub-pixels whereas the gaps in the parallax barrier can be thought of as a single view dependent pixel. Each of the observer's eyes can only see a specific set of sub-pixels through a view dependent pixel [42].

pixels columns occluded from one view determines how many different views can the screen display. The viewer can sense proper motion parallax by changing the position of his/her head in case the screen displays more than 2 views simultaneously.

Light fields is a vector function that represents the radiance of light flowing through every direction at every point in space [35]. In computer graphics, light field of a scene can be approximated as a 4D function by capturing the scene while moving the camera in space in steps. Automultiscopic screens displays this 4D light field where the fourth dimension represents the view number (more information about this can be found in chapter 6). Since the light field is usually captured at a resolution that is equal to higher than the automultiscopic screen's resolution, the view images are re sampled differently for each view. This further reduces the sampling frequency and hence cause smear aliasing if displayed directly. To resolve this problem, the re sampled view images have to be blurred in order to remove the undesired higher frequencies [42]. As seen in Figure 2.7, placing the parallax barrier vertically reduces the width dimension of the view displayed by a factor proportional to the number of views. In order to preserve the aspect ratio of the images, the parallax barrier can be tilted by some angle. This further complicates the anti aliasing and hence special filters need to be applied to the images before displaying [2].

As with every 3D display, automultiscopic screens also exhibit light leakage (crosstalk) between adjacent views. One reason for this is that the lenticules of a lenticular sheet are not perfect and do not separate the incoming light from different views completely. In case of a parallax barrier, light from the neighboring views can be occluded completely if the gaps in the barrier are extremely small (just as the case with a pinhole camera). However, there is a limit on how small the gaps can be made to be before distortion via diffraction kicks in. Hence, the gaps are made large enough so that distortion due to

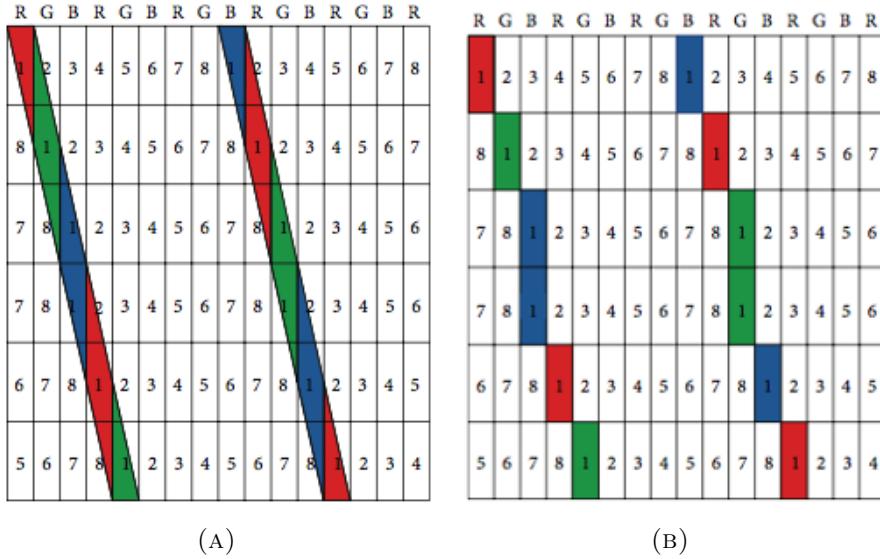


FIGURE 2.8: (a)Visible pixels observed through the slit of a tilted parallax barrier. It can be seen that the view is covering more than one pixels for every pixel. (b)The view that covers majority portion of a sub-pixel is assigned to it which is the reason why some light from that particular view will be leaked into the adjacent view [29]

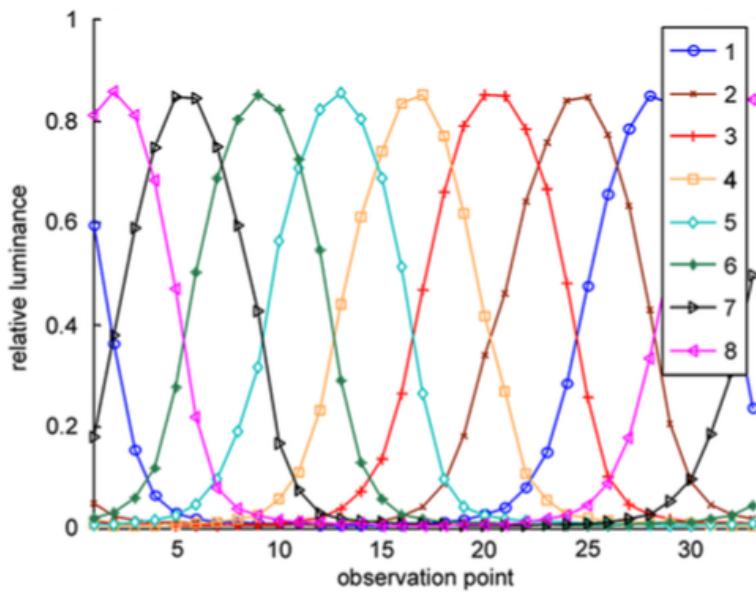


FIGURE 2.9: The experimentally observed luminance distribution of all views across different view points. It can be seen that some portion of the adjacent views is leaked even at the sweet spots. The light leakage increases as the observation point is moved away from the sweet spot [41].

diffraction can be avoided with the side effect of light leakage from neighboring views. Moreover, the lenticular sheet or parallax barrier are extremely hard to align perfectly to the pixels on such a small scale and is impossible in case the sheets are tilted. The latter will always cover a portion of pixels (Figure 2.8) which further increases the crosstalk. Typical light leakage pattern can be seen in Figure 2.9.

Crosstalk in automultiscopic screens have one advantage though. It helps smoothen the view transition while switching the position of the viewer. Hence, ideally it would be desired that there was no crosstalk present at sweet spots (the points in space where the viewers line of sight is perfectly aligned to the view image being displayed) along with some crosstalk present at points that are at the boundary of view transitioning.

Chapter 3

Related work

It is widely accepted that crosstalk in stereoscopic or automultiscopic screens results in reduced visual comfort, reduced disparity range, reduced contrast and most importantly, reduction of perceived depth [41][40] [24]. However, the literature is not thorough about how the crosstalk affects perceived depth. Wilcox et al [40] performed a large scale experiment in which 77 observers were shown a small stereo footage in two commercial large-format 3D theaters. The two major stereo degradations in theaters considered were crosstalk (ghosting) and brightness. The test subjects were asked to evaluate which one those degradations affected their experience the most. 75% of the subjects reported that ghosting was the most prominent in degrading their viewing experience. Below is a review of some of the work that has previously been conducted in order to understand the effects of crosstalk on perceived depth.

3.1 Effects of crosstalk on perceived depth

Huang et al [10] performed intensive experiments in order to obtain a threshold for the system crosstalk that on average would not mitigate the perceived depth from disparity. Since the viewer crosstalk is dependent upon the system crosstalk and the local contrast of the disparate objects in the image, a uniform level of crosstalk (same crosstalk all over the screen which does not change with time) should not have the same effect on all kinds of stereo images. This is because the HVS sensitivity to detect the change in luminance follows Webber's law [3]. This means that, compared to areas of the image with high local contrast, the HVS is more tolerant towards crosstalk where the local contrast is low. Based on this idea and the fact that the HVS also uses monocular cues to estimate depth, they performed experiments using a Wheatstone setup (Figure 4.3b) and a set of images with various contrasts and disparities to finally propose that 10% system crosstalk is the

maximum crosstalk that will mostly not nullify the depth estimation via disparity. We observed in our experiments that even though above 10% of crosstalk heavily degrades the perceived depth, it still does not result in total depth loss due to disparity. On our high contrast images, we determined that crosstalk level of greater than 16% usually resulted in the total loss of depth.

Ghosted images due to crosstalk as seen in figure 2.3c can be seen as locally decreasing the contrast of the disparate object. Rohaly et al [21] determined the effects that contrast exhibits on the stereoscopically perceived depth. They concluded that a decrease in contrast made the objects (crossed and uncrossed) appear to be farther from the viewers. This means that the crossed objects with decreased contrast appeared to be moving towards the horopter whereas the uncrossed objects appeared to be moving away from the horopter. She also found that monocular contrast reduction amplified this effect more compared to the contrast reduction in both eyes.

The work that is most relevant to our work was performed by Tsirlin et al. They quantified the amount of depth loss due to crosstalk on objects of various dimensions and disparities. In 2011 [25], they performed experiments where the test subjects were asked to specify the observed depth at various depths of a stimulus (a rectangular structure) the width of which was chosen to be such that at any disparity, the ghost would not be completely separated (Figure 3.1a). The luminance of the structure was set to be the maximum (white) on a completely black background. The subjects reported the observed depth via a slider bar located under the stimulus without any reference. They

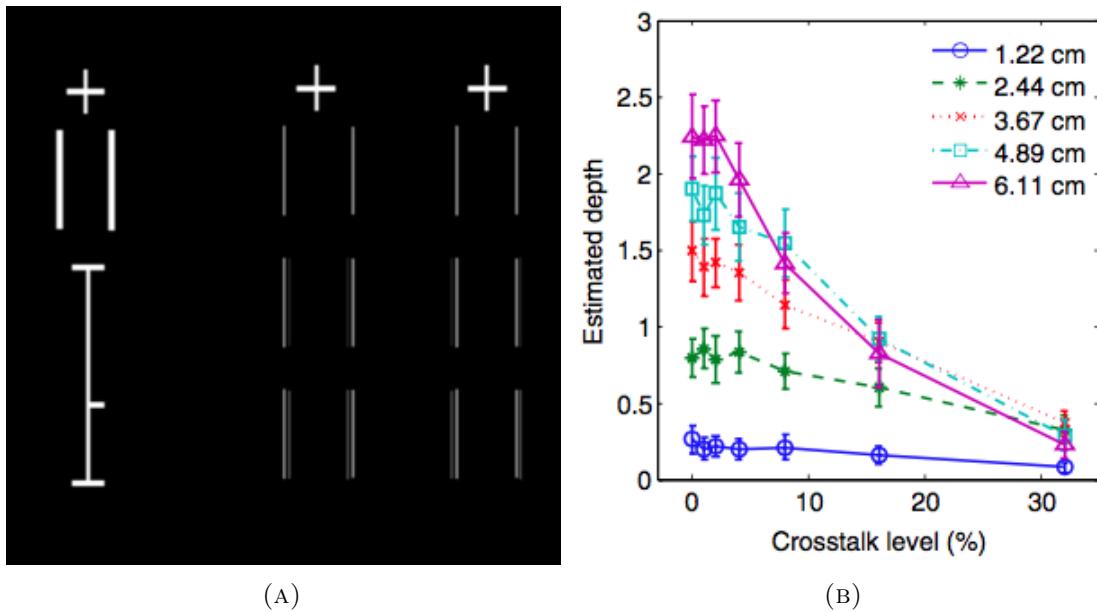


FIGURE 3.1: (a) The setup of Tsirlin’s experiments [26] [24]. Left: complete experiment where the viewers see the vertical lines as a single line in depth. Right: the same line with 0, 16 and 32% simulated crosstalk. (b) Depth degradation averaged over the test subjects.

observed that the perceived depth decreased with the increase in crosstalk as well as increase in disparity. In the second part of the experiment, they observed the effect of crosstalk on binocular occlusion and found out that effect of crosstalk on the perceived depth due to occlusion was even more severe. In 2012 [26], they performed similar experiments to observe the effects of crosstalk on thin structures where the ghost is always completely separated from the stimulus (which is usually the case with vertical thin structures present at a significant distance from the plane of focus). Again they found that observed depth degrades significantly as the disparity or the crosstalk level was increased. However this time the degradation was observed to be higher than the case where the ghosts always overlapped the stimulus. One problem with these experiments is that the stimuli are presented as uniformly luminant objects on a black background where there is no other depth cue present. This is usually not the case with the complex scenes that we typically observe in 3D movies. We observed in our test experiments that in such monochromatic scenes, the depth of the stimuli was extremely hard to perceive even at moderately high disparities. Tsirlin et al also observed that the perceived depth degraded with disparity even in the base case, i.e., the case where crosstalk was set to zero. This is not usually the case in complex scenes. Secondly, there was no reference used and the subjects were simply asked to report the observed depth via a slider bar that was controlled with a mouse. This ‘rating’ setup can be problematic because it can not be guaranteed that all the subjects would rate the same depth equally. And finally, the relation of the stimulus width to the perceived depth at different disparities was not examined, which we hypothesize could be relevant.

Tsirlin et al [24] again performed similar experiments but this time with complex natural scenes. The subjects were shown a complex crosstalk-free scene with two objects of interest marked by arrows. Once the observers had memorized the perceived depth difference between the two objects, they were shown the exact same images with crosstalk induced and were asked to rate the depth difference they observed between those two objects. The results again concluded that the perceived depth difference decreased as the theoretical depth difference between the objects of interest and the crosstalk increased. One problem with this experiment is that it reports the observed depth in a complex cluttered scene where the ghosts from the neighboring objects will play a significant role in determining the observed depth difference. Hence, the effect of how the HVS will respond to the geometry of the stimulus can not be isolated. Moreover, just as in her previous experiments, the users rating can easily deviate from each other.

In contrast, in our experiments, we present the test subjects with a depth adjustable crosstalk-free stimulus and ask them to match the depth according to a reference crosstalk added test stimulus. This way, it can be ensured that all the subjects would register the

same depths equally. Moreover, the stimuli are rendered images of a 3D scene containing all possible depth cues. The experiments are explained in detail in Chapter 4.

3.2 Depth resolution mechanisms in HVS

Once we accept the hypothesis that crosstalk in stereoscopic displays reduces the perceived depth of an object, and that it does so proportionally to the level of crosstalk and the disparity of the object, then, in order to understand why the perceived depth is degraded, we need to refer to literature where the depth information extraction by the HVS from disparity is investigated.

It is widely believed that the HVS uses some kind of cross-correlation of retinal patches between the two eyes in order to determine the location (or in other words the disparity) at which some object is located in a retinal image with respect to the other. Cormack [5] investigated the binocular fusion process with respect to contrast of the stimuli and concluded that the binocular fusion was dependent upon the contrast and the strength of the binocular cross-correlation profile of the stimuli. The cross-correlation was computed as a function of image luminance. The stimuli used in his experiments were random dot stereograms in which the number of white dots matching between the left and the right eye images at some disparity represented the interocular correlation. He concluded that at low contrast, the binocular fusion increased when the contrast of the stimuli (white dots in the stereograms) increased. The reason behind this can be seen in Figure 3.2a where the cross-correlation (interocular correlation) profiles at different contrasts has been plotted. The peaks in these profiles represent the disparity at which the two images matched the most. It is clear that, as the contrast decreases, the distinctness of the these peaks also degrades which might make it difficult for the HVS to identify. The same is also true if noise is added to the images as seen in Figure 3.2b. This can be one of the reasons for the degradation of the perceived depth in ghosted images as the ghost (due to crosstalk) in an image is just a low contrast version of the stimulus added into the image at some disparity effectively reducing the local contrast. We used these finding in deriving our own model for the HVS that will be explained in Chapter 5.

Filippini and Banks [7] explained the limits of stereopsis and modeled them by using windowed interocular cross-correlation. The neighborhood of each pixel in one stereo images (e.g., left one) was weighed by an isotropic Gaussian window followed by horizontally displacing a similar window in the right eye image. For each displacement, the cross-correlation of the windowed patch was computed. The disparity was obtained at the displacement value where the maximum correlation for the patches was obtained.

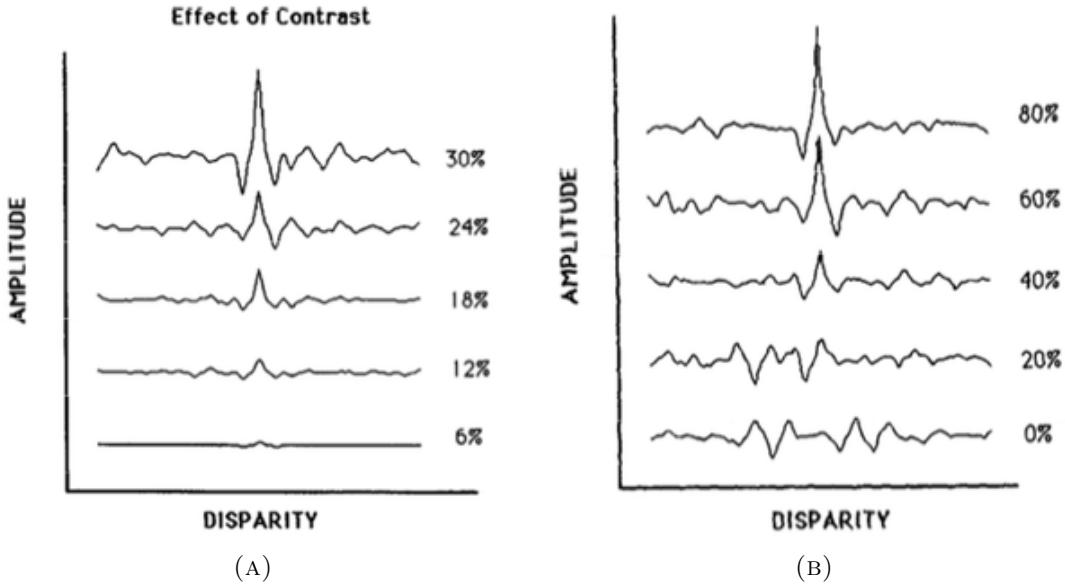


FIGURE 3.2: (a) 1D cross-correlation with varying contrast. (b) 1D cross-correlation with variable level of added noise [5].

The 2D cross-correlation formula used is shown in equation 3.1.

$$c(\delta_x) = \frac{\sum_{(x,y) \in W_L} [(L(x,y) - \mu_L)(R(x - \delta_x, y) - \mu_R)]}{\sqrt{\sum_{(x,y) \in W_L} (L(x,y) - \mu_L)^2} \sqrt{\sum_{(x,y) \in W_R} (R(x - \delta_x, y) - \mu_R)^2}} \quad (3.1)$$

where $L(x,y)$ and $R(x,y)$ are the image intensities for the left and right eye, W_L and W_R are the Gaussian windows applied to the images, μ_L and μ_R are the mean intensities within the two windows, and δ_x is the displacement (disparity) of window W_R relative to W_L . This way, the correct disparity for each object in stereo images can be computed. The problem is that this method also computes the correct non-degraded disparity even if the crosstalk is induced to the images. This does not correspond well to the HVS behavior. Hence, we need a model that can simulate not only the correct disparity estimation but also the degraded disparity when crosstalk is added to the images.

3.3 Reduction of crosstalk

As mentioned in the previous sections, crosstalk severely hinders the perceived depth and reduces the overall visual quality and the viewer comfort. Hence, it is imperative for the crosstalk levels to be mitigated or at least lowered to an unperceivable level. Almost all of the 3D displays using present hardware technologies exhibit some amount of crosstalk. Fortunately, it is possible to calculate the amount of leakage between views at any position for any display, which can help us compensate for it. Usually, a luminance

meter is used to detect the amount of light leaked between views. Once the leakage levels for a display system are computed, image processing techniques can be used to preprocess the stereo images in such a way that results in mitigation or minimization of the perceived ghosting once they are displayed on any traditional 3D screen. Early attempts relied on subtracting from one view image, the leaked image intensity from the other perspective view before displaying (Figure 3.3). This technique might result in negative light intensities in low contrast areas of the images. Since negative light can not be produced by any display, the areas of the preprocessed image that has negative intensities are clamped to zero. This means that the crosstalk in these areas is not fully subtracted and hence the ghosting still persists. One way to avoid such problem is to increase the overall image intensity by some value such that no negative values are obtained after crosstalk subtraction. This, however, comes at a cost of decreased average contrast of an image and degrading the overall image quality. In the following sections we will review some of the contemporary techniques that are used in order to remove the crosstalk while maintaining the overall image quality.

3.3.1 Stereoscopic displays

Konard et al [13] proposed a crosstalk reduction technique that took the ghosting perception thresholds into consideration. For every possible combination of crosstalk level and the contrast of the scene, they experimentally computed the minimum level of light intensity that, after subtraction, would bring the ghosting to an undetectable level by a human observer. Ghosting was later minimized efficiently by storing and using these values from a lookup table. In order to avoid the negative light intensity values after crosstalk subtraction, they proposed to increase all the pixel values by the maximum negative intensity obtained, hence, reducing the overall image contrast. Limpcomb and Wooten [18] took the spatially non-uniform nature of crosstalk in the display into consideration and proposed a technique that consisted of dividing the screen into 16 horizontal bands. The crosstalk was evaluated for each of these bands followed by the crosstalk subtraction in the images accordingly. The problem with this technique is that the non uniformity of the crosstalk level between and across the bands is continuous in nature which is why it resulted in over and under subtraction of crosstalk between bands that was still perceivable by the viewers. Another problem is that all of the aforementioned techniques assumes static scenes and hence ignore the temporal aspects.

In time sequential displays, crosstalk is induced not only by the view image from the opposite view but also by the views in the previous frame. Smit et al [22] took this phenomenon into consideration as well. However, as with every subtractive approach, this technique also suffered from over subtraction. Sohn and Jung [23] proposed a technique

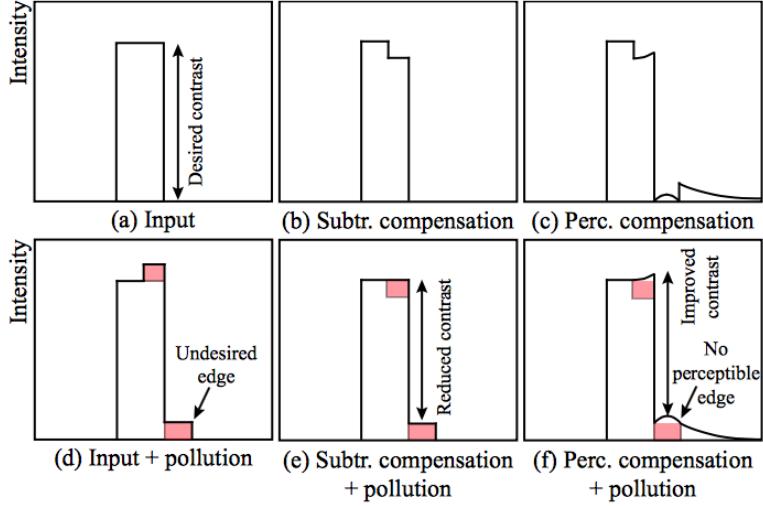


FIGURE 3.3: (a) Input image. (b) Subtractive compensation. (c) CSF weighed compensation. (d-f) Observed images [27].

that utilized disparity adjustments for a scene such that the least amount of uncorrectable crosstalk occurs in perceptually important regions. A crosstalk visibility estimator that simulated human sensitivity to luminance changes according to Webber’s law was used in order to determine the areas in a scene where the spatio-temporal crosstalk was visible. In the second step, the global disparities in the scene were changed (shifting the zero plane) and the amount of perceivable uncorrectable crosstalk was computed using the crosstalk visibility estimator for every shifted zero plane. The global disparity shift that yielded the lowest amount of uncorrectable visually important crosstalk was chosen for displaying. In order to remove the already minimized uncorrectable crosstalk, the intensities in the scenes was locally increased in order to preserve as much dynamic range as possible.

Most recently, Poulakos et al [27] proposed a crosstalk reduction minimization technique that utilize the contrast sensitivity function¹ (CSF) to steer the minimization in such a way that only perceptually important areas in a scene are corrected for crosstalk. This is attained by minimizing 3.2 the residual r 3.3:

$$\operatorname{argmin}_x \|\lambda_l \cdot r\|^2, \quad s.t. \quad 0 \leq x \leq 1. \quad (3.2)$$

$$r = x + \phi(x) - \bar{x} \quad (3.3)$$

where λ_l is a diagonal matrix of the CSF spectral coefficients, \bar{x} and x denote the original and the compensated images, respectively. $\phi()$ denotes the non-linear function representing the crosstalk. The constraints on the optimization are necessary so that the

¹Contrast sensitivity is the measure of human ability to discern between luminance of different levels at different spatial frequencies in a still image [31].

minimized image values do not exceed the displayable range. Using the CSF helps spread the intensities of the uncorrectable crosstalk to the areas which are perceptually less important (as seen in Figure 3.3). Also in this figure, it can be seen that the optimization introduces a Cornsweet profile at the edges effectively improving the local contrast. One other factor that is important for the perception is visual masking, i.e., the phenomenon where a signal is undetectable in the presence of another signal with similar a pattern. Poulakos et al further improved the optimization by incorporating into Equation 3.2 an additional weighing with a visual masking operator:

$$\underset{x}{\operatorname{argmin}} \|\lambda_n \cdot \lambda_l \cdot r\|^2, \quad s.t. \quad 0 \leq x \leq 1. \quad (3.4)$$

Here, λ_n is a visual masking binary operator that will help not strive to minimize x for non-perceivable areas of the ghosted image due to visual masking. Although they did not mention it, this idea can also be used for optimization of light fields for crosstalk reduction in automultiscopic displays.

3.3.2 Automultiscopic displays

As discussed in Chapter 2, crosstalk in an automultiscopic displays is introduced by the neighboring views that is the horizontal neighbors of a pixel in case the lenticular sheet is not tilted, and both horizontal and vertical neighbors of the view in case the lenticular sheet is tilted. This is because a tilted sheet can not cover a whole pixel of the underlying LCD panel completely (Figure 2.8a). Due to this reason, eliminating crosstalk from automultiscopic displays is more complex than stereoscopic displays.

Campisi et al [1] proposed that, for each view of an automultiscopic display, the ghosting from within and across the views can be observed and composed as coefficients of a circulant Toeplitz matrix. Hence, for an automultiscopic screen with n views, the perceived views can be modeled as:

$$L_p = (E + A) \cdot L_d \quad (3.5)$$

where L_p and L_d are the vectors of displayed and perceived images in luminance domain, A is a circulant Toeplitz matrix approximating the crosstalk coefficients and E is $n \times n$ identity matrix respectively. A generic reduction of crosstalk was derived from Equation 3.5 for each channel by inverting $(E + A)$:

$$L_{rd} = (E + A)^{-1} \cdot L_{ri} \quad (3.6)$$

where L_{ri} is the gamma corrected intended luminance of the red channel. The implementation of this will generate luminance values that are beyond the range of the display (negative values). To avoid this, equation 3.6 can be modified in order to increase the average contrast of the image:

$$L_{rd} = (E + A)^{-1} \cdot \left(\frac{r_i \cdot (255 - \beta) + 255 \cdot \beta}{255^2} \right)^\gamma \quad (3.7)$$

where the term β denotes the level to which the black level must be shifted up. As expected, this will avoid getting any negative values for luminance but decrease the image contrast. The authors claimed that a certain level of negative values can be tolerated. Therefore, for each view, the value of β was computed iteratively so that it resulted in no more than a desired number of negative values.

Since the light leakage from neighboring left and right views can be considered as non-energy preserving blurring, inverse filtering might be applied to the images in order to minimize the perceivable ghosting. Jain and Konard [11] derived a blurring filter that simulated the light leakage from the neighboring views, combined it with the anti-aliasing filter typically used in automultiscopic displays and used Wiener inverse filtering to preprocess the images in order to mitigate the crosstalk. Li et al [17] argued that light leakage from the same view due to incorrect sub-pixel approximation is more severe than light leakage from the neighboring views and hence proposed to use an appropriate filter that mimics the sub-pixel light leakage in order to be used for inverse filtering.

Generally, the simulation of crosstalk can be written as a system of equations $AX=B$ where A is the crosstalk coefficient matrix as described above, X is the set of intended images and B is the set of crosstalk added view images. Naively, one can compute crosstalk free X by solving $X = A^{-1}B$. However, for this to work, A should be a full rank matrix and even then the result will be unacceptable since the resulting preprocessed images will contain negative values. Wang et al [29] first proposed a mathematical model for computing the crosstalk coefficient matrix A only by using the parameters of the screen such as screen resolution, pixel dimensions, and the angle of tilt of the lenticular sheet. They argued that, usually, the crosstalk is measured experimentally which is always prone to errors. The mathematical model should provide a more reliable crosstalk estimation. Secondly, they proposed solving this linear system of equations in a constrained manner where the values of X can not exceed the range $[0, 255]$. They named it as Box-Constrained Integer Least Square (BILS) solution.

$$\begin{aligned} & \min_{X \in BOX_x} \|B - AX\|_2^2 \\ & BOX_x = X \in \mathbb{Z}^{w*h} : L \leq X \leq U \end{aligned} \quad (3.8)$$

where $\|.\|_2$ denotes the Euclidean norm, w and h are the width and height of the screen resolution, and L and U are vectors of 0 and 1 respectively.

Chapter 4

Crosstalk Experiments

In this section we discuss the details of thorough experiments that we conducted in order to quantify how the crosstalk affects the perceived depth. For the sake of completeness, we performed various experiments on considerably complex (realistic looking) rendered images that consisted of stimuli of various widths and geometric complexities. The effect of crosstalk were observed for both stereoscopic and automultiscopic displays.

4.1 Motivation

Some preliminary work has been done by Tsirlin et al[25][24][26] in order to understand these effects. However, several aspects of their experiments were somehow limited for the aim of applying the findings in a practical scenario. Firstly Tsirlin et al's experiments were performed on monochromatic images where no depth cues other than disparity were present. As mentioned in Chapter 3, we conducted a pilot experiment using such stimuli (a thin rectangular white bar on a black background) using a similar Wheatstone setup and found that it was very hard to sense any depth of the stimulus even at high disparities. Even after some time, guessing the apparent distance of the stimulus from the screen in length units was quite unreliable. This would be an indicative that the naive test subjects in their experiments may also have encountered the same problems. This observation is backed by the fact that the test subjects in [26] and [25] reported reduced depth of the stimuli even when no crosstalk was present. More reliable experiments would be ones where the test subjects report the actual theoretical depth in the base case, i.e., when no crosstalk is present. Additionally, the test subjects were simply asked to report the perceived depth via a sliding bar representing zero or some maximum depth at extremes of the bar; this methodology, relying on the subjects choosing within a scale is susceptible to varying answers both intra-subject, and especially, inter-subjects.

Further, the scenes of the natural world usually have at least some other monocular cues e.g. proper light shading, texture gradients, defocus blur etc. Finally, since one factor contributing to visibility of the ghosts is also how much it is separated from the object, wide objects at some disparity will exhibit less visible ghosting than thin stimuli. For these reasons, we decided to use a more reliable, reference matching experimental setup to analyze the effects of crosstalk on the observed depth on rendered stimuli of different widths in a natural scene .

It is commonly believed that the reason for degraded observed depth in the presence of crosstalk is due to the fact that the human visual system is confused in choosing the exact location (or disparity) of the correct match for an object in the corresponding binocular retinal image. To the best of our knowledge, it was unclear whether this HVS confusion is elevated or reduced when the geometric complexity of the object (stimulus) is increased. For this reason, we set out to analyze the effect of crosstalk on depth perception of geometrically complex stimuli. Another motivation for our proposed thorough analysis of effects of crosstalk on perceived depth of thin stimuli was, that current studies suggest that for a certain crosstalk level and considerably smaller width of a disparate object, the extent of observed depth degradation increases as the disparity increases. This sounds counter-intuitive because, for substantially high disparity of a thin object, the ghost will be completely separated and far away from the actual object itself. If, the reason for degraded depth is actually the confusion of HVS in finding the proper match between the actual stimulus and its ghost, then in case of thin stimulus and large enough disparity, the HVS should be able to find the proper match relatively easily. The reason being that it is easier to distinguish between the ghost and the stimulus when they are completely separated as compared to the case of overlapping ghost. We hypothesized that the observed depth of any object with respect to the actual theoretical depth should degrade as long as some part of the ghost overlaps the object itself, and the observed depth should improve when the ghost separates completely. Hence the graphs of Figure 3.1b, intuitively, should be Parabolic shaped.

In stereoscopic displays, the arrangement of object-ghost pair is antisymmetric between the eyes. This, however, is not the case with automultiscopic displays. As seen in Figure 2.9, at least two of the neighboring views are responsible for adding crosstalk at any view point. Also, at the sweet spots, the number of views involved in light leakage from the left is always equal to the number of light leaking views from the right. Because the automultiscopic display displays light fields, the arrangement of object-ghost tuple is symmetric in both eyes (at sweet spots). The HVS might react differently to this major difference. To the best of our knowledge, until now there have been no studies that observed the effect of crosstalk on perceived depth in automultiscopic displays. For this reason we decided to analyze the effects of an automultiscopic display's sweet

spot crosstalk on the perceived depth of all the stimuli we used in experiments for the stereoscopic case.

4.2 Stimuli

Our stimuli consisted of an object (a cylinder or a dragon) placed in front of a plane that had a wooden texture. The reason for choosing a textured background was the HVS's efficiency in distinguishing between textures. This would make it easier for the observer to correctly guess the distance between the background and the foreground object. In order to make sure that our stimuli contained most of the basic monocular cues, we rendered them using blender 2.73. The scene (coordinates represented in meters) as shown in figure 4.1a, consisted of a an object of interest placed at origin. The background plane with wooden texture was placed 3 meters behind the object parallel to the camera plane. In order to capture the shading effects on the object, the scene was lit with a light source (sun) that was placed 10 meters in front and 3 meters at height of the object. Finally the camera was 10 meters in front of the object. The stereo scene was generated by capturing the two images at some distance on each side parallel to the image plane. This distance represented the baseline of the cameras, and was varied.

Our experiments were based on test subjects judging the observed distance between the cylinder and the textured plane. This could have been achieved by either moving the cylinder or by moving the background plane in space relative to the camera. However, this would have changed the object or the texture size which could have been used as an undesired (not affected by crosstalk) cue by the observers. In order to avoid that, we simply generated a dense series of images with varying camera baselines and keeping the position of the background plane in lock with respect to the camera. This way, the background was always at zero disparity (i.e. the plane of focus) and the cylinder always had a crossed disparity (proportional to the baseline distance) meaning the cylinder always appeared to be in front of the plane at different distances. In order to avoid resizing while being displayed on our display, images were rendered at the resolution of 768x432. With the smallest baseline, the observed distance between the cylinder and the plane was negligible whereas this distance appeared to be increasing as the baseline increased all while keeping the size of object, size of texture and the proximity of the object to the texture constant. For the objects we chose four cylinders of different radii and a dragon from Stanford's 3D scanning repository¹. Figure 4.2 shows a sample. The details of each of the objects can be found in table 4.1.

¹<http://graphics.stanford.edu/data/3Dscanrep/>

TABLE 4.1: Description of the objects used as stimuli.

Object	Scene Dimensions (xyz)	On-screen Dimensions (arc min)
Cylinder(thin)	5cm x 2m x 5cm	18.9 x 357
Cylinder(medium)	10cm x 2m x 10cm	37.8 x 357
Cylinder(thin)	45cm x 2m x 45cm	56.7 x 357
Cylinder(thickest)	80cm x 2m x 80cm	75.6 x 357
Dragon	0.31m x 0.22m x 14.17cm	380 x 260.4 (max)

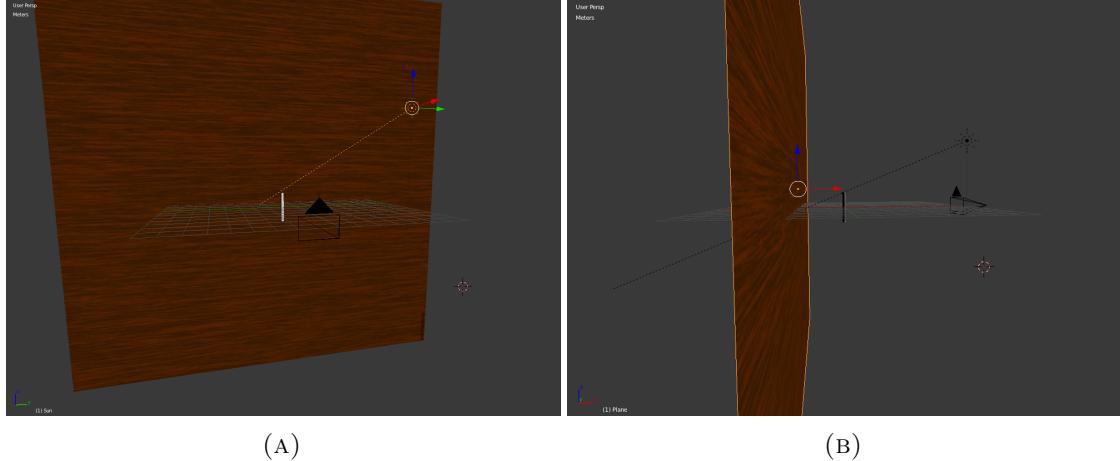


FIGURE 4.1: (A) Frontal view of the scene rendered with Blender. (B) Side view of the same scene.

In order to generate a set of stereo images with the object of interest at different disparities, the scene was captured with the camera baseline varying (in scene coordinates) between 0.002 meters till 0.12 meters with the step of 0.002 meters. This gave us a set of 60 stereo images where the theoretical depth of the object of interest ranged from 0.04 cm to 3.45 cm (0-10.5 arc minutes in terms of crossed disparity) in front of the plane of focus (wooden plane). Figure 4.2 shows stereo images of one of the cylinders and of the dragon. In a reference matching experiment, test subjects might use the discreet nature of the test stimulus as a matching cue. In order to avoid that, the step size was chosen small enough so that the transition from the minimum to the maximum depth for an object would seem continuous to the observer.

In order to compensate for symmetrical ghosts on both sides while the theoretical depth of object of interest remain unchanged, the images for automultiscopic experiments were rendered similarly with difference being that the range of baseline started from 0 until 0.24 meters. The reason for this is explained in the next section.

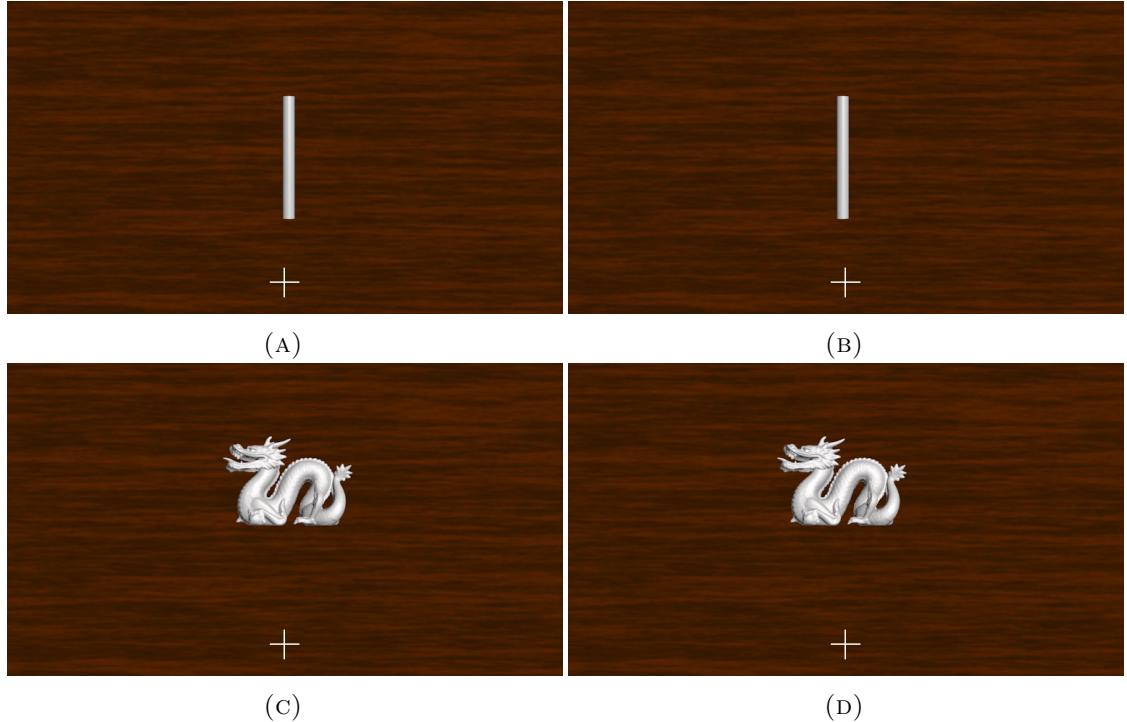


FIGURE 4.2: Stereo images of the cylinder of 5 cm radius and the dragon. The theoretical depth difference between the objects and the background is 3.44 cm. Viewer can cross-fuse to see in depth.

4.3 Apparatus and simulation procedure

The experiments were conducted on a commercial 47" SIMS2 HDR47E LCD display with a spatial resolution of 1920x1080. The left and right half of the screen each with a resolution of 960x1080 were gated to the left and right eyes using a custom built Wheatstone setup [39] that consisted of two mirrors M1 and M2 set at a 45° angle with respect to the screen. The optical length from the eyes to the screen via mirrors M1 and M2 was measured at 87.3 cm. The whole setup was enclosed in a black box with an aperture through which the viewer could view the screen via mirrors M1 and M2. The display, even though built for HDR viewing had a *DVI plus* mode for which the luminance measured with luminance meter via the aperture ranged from ?? to ?? cdm^{-2} . The *DVI plus* mode should (according to the documentation) mimic a traditional LCD display, hence we chose to use this mode.

Since on an LCD panel, all the RGB channels equally and additively contribute to the crosstalk, we simulated the ghosted images simply by adding a percentage image of unintended view to the intended view image. For stereo, simple addition of a percentage of the right image into the left and vice versa sufficed. However, for the automultiscopic case, we simulated the view-luminance profiles (Figure 2.9) by fitting a Gaussian at each view. We found that an upscaled Gaussian such that it attains the value 0.8 at the mean

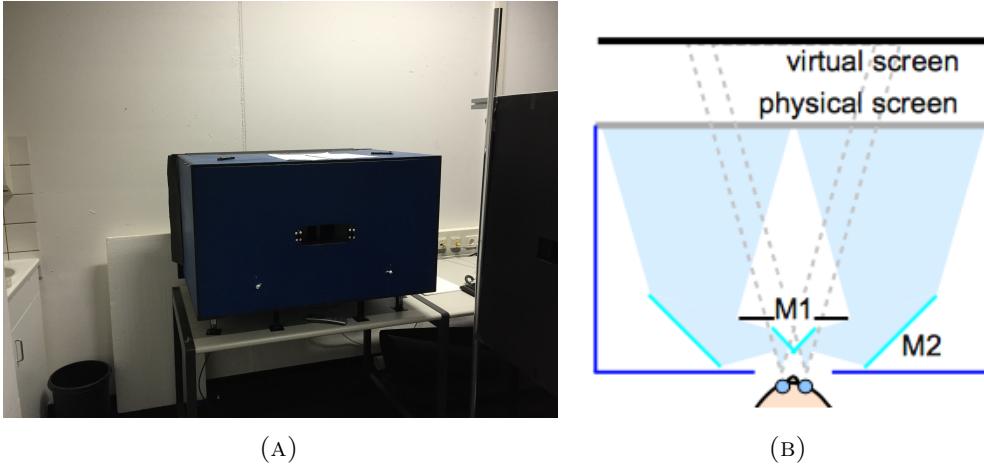


FIGURE 4.3: Our Wheatstone setup. Physical view frusta are shaded in light blue where as the virtual view frusta are depicted in dotted gray lines. The cyan lines represent the mirrors [28]

(μ) and with standard deviation $(\sigma) = 2$ matched the required view-luminance profile appropriately. Figure 4.4 shows 11 such Gaussians, the means (μ) of whom are shifted according to the sweet spots of an 11-view automultiscopic display. Using our simulated profiles, we found that at any sweet spot viewing position, the amount of light leaked from views other than the two adjacent views can be considered as negligible. Hence, in our simulation we only considered light leakage from the two adjacent views. Figure 4.5 shows the simulated crosstalk on stereo and automultiscopic display. Considering that an automultiscopic screen displays light fields sampled at frequency $1/\delta_x$ (δ_x being the step size at which the images were captured along the camera baseline), the leaked light from a position ‘y’ to position ‘x’ can be modeled as:

$$l(y, x) \simeq a_y(x) \cdot f(y) \quad (4.1)$$

where $a_y(x)$ is the value of crosstalk as given by the Gaussian(corresponding to location y) at position x, and $f(y)$ is the light field image at position ‘y’. The observed image g at any viewing position x can be mathematically modeled as:

$$g(x) \simeq \dots + l(x - 2\delta_x, x) + l(x - \delta_x, x) + l(x, x) + l(x + \delta_x, x) + l(x + 2\delta_x, x) + \dots \quad (4.2)$$

This would mean that the amount of ghost separation at any disparity for an object would be constant on either side proportional to the object’s parallax in δ_x . In order to keep the nature of the automultiscopic experiments similar to that of the stereo (where the ghost separation increases as the disparity increases), we changed the sampling rate δ_x with respect to the disparity of the object. I.e. For a disparity $2d$, the observed left

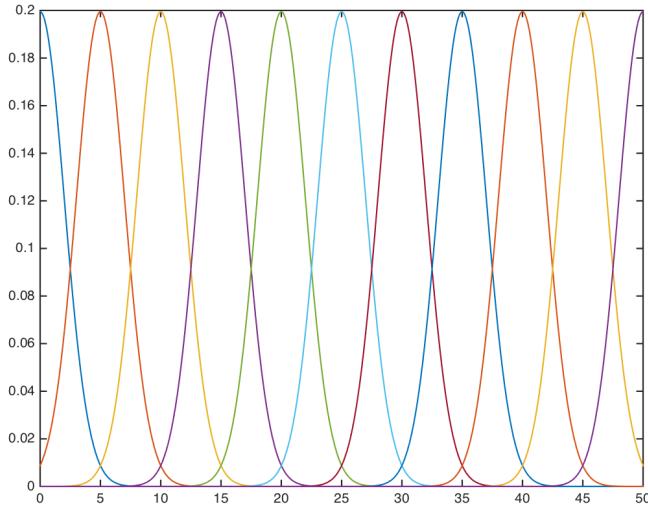


FIGURE 4.4: View-luminance intensity profiles for an 11-view automultiscopic display simulated with Gaussians of $\sigma = 2$ centered at the sweet spots. X-axis denote the viewing position. It can be seen that the simulated luminance profiles closely matches the experimentally observed view-luminance profiles in Figure 2.9

and right images can be written as:

$$\begin{aligned} g(d)_{left} &\simeq l(2.d, d) + l(d, d) + l(0, d) \\ g(d)_{right} &\simeq l(-2.d, -d) + l(-d, -d) + l(0, -d) \end{aligned} \quad (4.3)$$

4.4 Experimental procedure

4.4.1 Environment setup

For our experiments, we subdivided the screen horizontally: the upper half where the test subjects would see the crosstalk added scene (stimuli) with some fixed depth of the object; and the lower half, where the subjects were shown the same scene without any crosstalk. We will refer to the upper half of the screen as *reference* stimulus and the lower half as *test* stimulus. Also we will call the depth difference between the background (wooden plane at zero disparity) and the foreground object (at some crossed disparity) the depth of the stimulus (*DoS*). As mentioned in Section 4.2, *DoS* in the presented stimuli ranged from 0.04 to 3.45 cm. The subjects could vary *DoS* in the test stimulus in the above mentioned range via left and right mouse buttons. Due to the dense sampling of the light field (small baseline distance change between stimuli), the change in the depth appeared continuous, i.e., the viewers could not observe any jumps while changing it. Pressing the left mouse button continuously for 6 seconds would change the depth

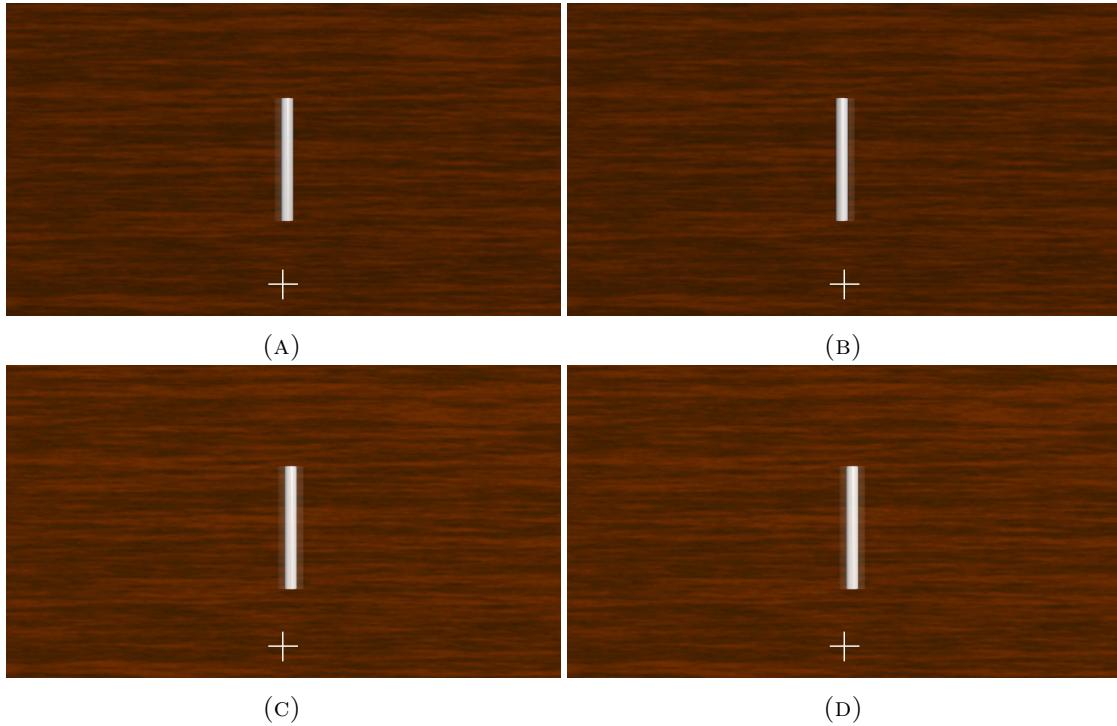


FIGURE 4.5: (a,b) Observed left and right stereo images with 14% crosstalk and crossed depth of 3.45 cm. (c,d) Same configuration for observed images on automultiscopic display). Viewers can cross-fuse to see in depth.

of the object from 0.04 cm to 3.45 cm, and vice versa for the right click. The subjects could stop anywhere in between by releasing the mouse button. During the experiment, the subjects were tasked with matching the *DoS* in the test stimulus to the *DoS* of the reference stimulus. Upon matching *DoS*, the subjects would register the answer using the center mouse click. Figure 4.6a shows the overall layout of the screen from a viewer's perspective. To avoid the possibility of the subject matching the absolute depths of the objects in stimuli rather than *DoS*, as required, we shifted the global depth of the reference stimulus by 3.45 cm away from the screen depth (towards the viewer) whereas the global depth of the test stimulus was shifted equally and similarly away from the viewer. We thus ensured that the absolute depths of the objects in stimuli would never coincide, and hence the subjects would not be able to, and would not attempt to match the absolute object's depth. In order to make this arrangement obvious to the viewer, a nonious cross-hair representing the plane of focus (located at the depth of the physical screen) was added at the bottom center of the screen. Figure 4.6b shows the top view of the visualized environment. The experiment environment was created using the Psychtoolbox-3 package for Matlab, which was also used to record the subject's answers.

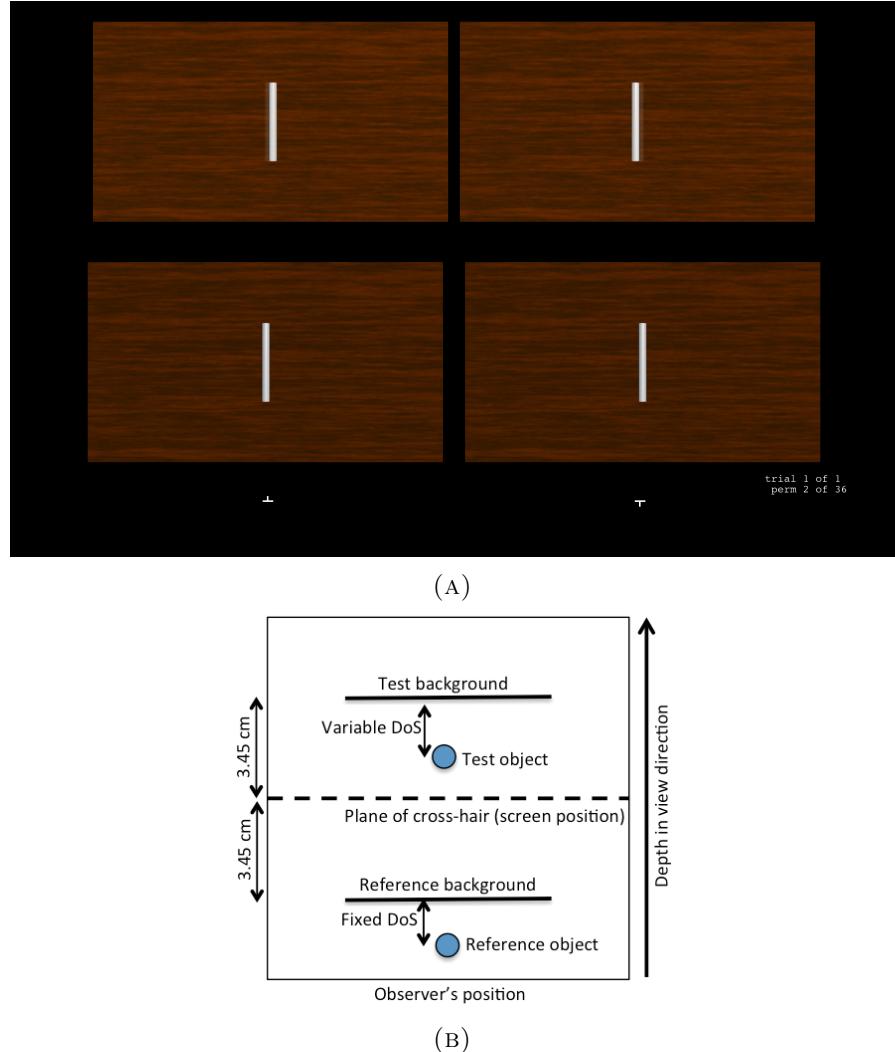


FIGURE 4.6: (A) Schematics of the complete display. Upper half of the screen shows the reference stimulus, in this case, with 14% crosstalk. The bottom half shows the user-controlled test stimulus which is always free of crosstalk. (B) Layout of the experiment environment in terms of depth w.r.t. observer.

4.4.2 Subjects

Several paid test subjects and the author participated in the experiments. In order to ensure subject's correctness of the stereo vision, they were first asked to identify the proximity of the test and reference stimuli with respect to the cross-hair. Then they were asked to identify the polarity of the change in depth of the test stimulus while the author changed it slightly in each direction randomly. Table 4.2 shows the number of participants in each kind of experiment.

TABLE 4.2: Number of subjects that participated in each experiment.

Object	Stereo	Automultiscopic
Cylinder(thin)	11	13
Cylinder(medium)	11	11
Cylinder(thick)	7	13
Cylinder(thickest)	8	13
Dragon	13	13

4.4.3 Experiment parameters

Four levels of crosstalk (0%, 3%, 7% and 14%) and seven levels of DoS (crossed disparity of object) (0.04, 0.73, 1.18, 1.53, 1.75, 2.76 and 3.44 cm) were chosen to be tested. We neglected any level of crosstalk greater than 14% because in our pilot experiments we noticed that any greater crosstalk level almost always led to matching of the stimulus with the ghost, instead of the corresponding stimulus, resulting in total loss of depth. Further we could not find any commercial stereo display where the crosstalk levels were that high: Usually they are below 10%. During the experiment, each subject was asked to match the observed depths for the 28 possible combinations (4 crosstalk levels \times 7 DoS levels) for each kind of stimulus. Trials for stereoscopic and automultiscopic cases were conducted separately.

4.5 Results

In this section, we show the results obtained for the different experiments performed. In all cases, we show the median data for all the observers along with the standard error for each crosstalk level, DoS and geometry (Figure 4.7 to Figure 4.14). Individual lines in the graphs denote the different objects used as stimuli (See Table 4.1). The x and y axis denotes the actual (theoretical) and the viewer-observed depths of the stimuli respectively. Ideally, if crosstalk had no effect on the perceived depth, all the graphs should be diagonally coinciding lines. Even though generally this is not the case, it can be observed that in the base case (no crosstalk), all the graphs, for all objects, almost approximates a diagonal line for both stereo and automultiscopic experiments. This is an improvement over [26] meaning that the reported amounts of depth degradation for different cases would be more reliable, since in our case, the base case (0% crosstalk) is consistent with what previous literature reports [find the literature](#).

4.5.1 Stereo

It can be seen in Figure 4.7 that in the absence of crosstalk, observers mostly perceived depths correctly. A slight overestimation of depth in case of the smallest depth (0.04 cm) can be seen for the thickest cylinder (width = 36 arc min). Given the fact that the user-controlled depth variation of the test stimulus was continuous, it is reasonable to assume that it would be quite unlikely for the observers to exactly match the depths. Hence the slight overestimation of depth can be considered as an observer approximation error. The same, however, could not be straightforwardly said about the underestimation for the dragon at 2.76 cm depth which is mapped to approximately 2 cm observed depth. But since the next greater depth shows no underestimation, we can safely assume it to be a human error.

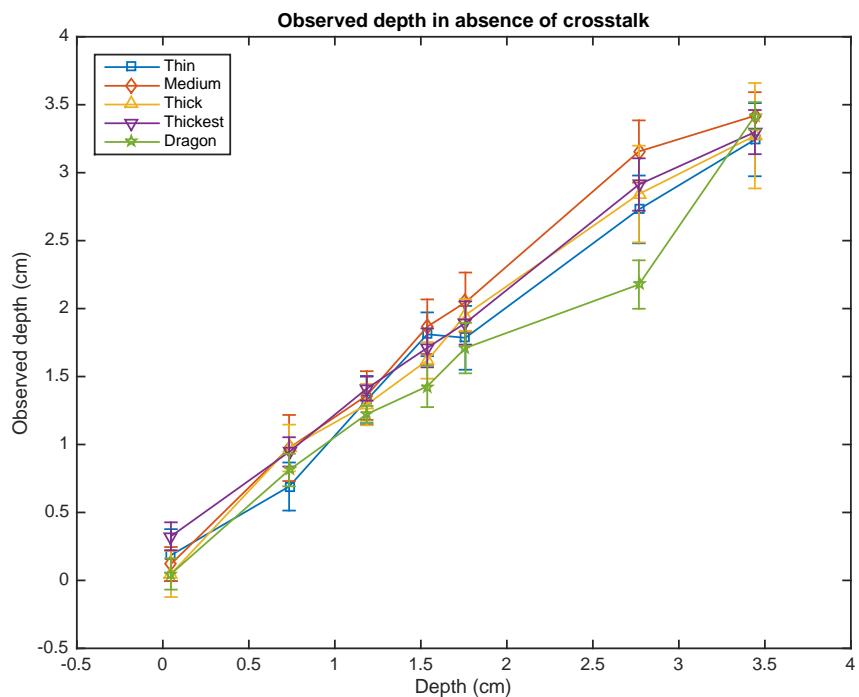


FIGURE 4.7: The median data for all observers is shown. The abscissa shows the observed depth when no crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

When the crosstalk is increased to 3% (Figure 4.8), it is clear that the observed depth of the thin (18.9 arc min) cylinder starts degrading at depths greater than 1.75 cm. The angular disparity at this depth is 5.35 arc min which would give a ghost separation² of 28%. A slight overestimation of the 2.76 cm depth for the thickest cylinder (75.6 arc min) can be considered as approximation error, since the next greater depth is observed

²Ghost separation is defined as the percentage of ghost stimulus that does not spatially overlap the stimulus.

without any degradation. We think it would be reasonable to consider this as a human approximation error because the HVS probably does not gets confused for some disparity and then behaves without any error for the following larger disparities. I.e. Once the observed depth starts degrading at some disparity, provided all the other conditions are kept constant, it should keep on degrading for all larger disparities as well.

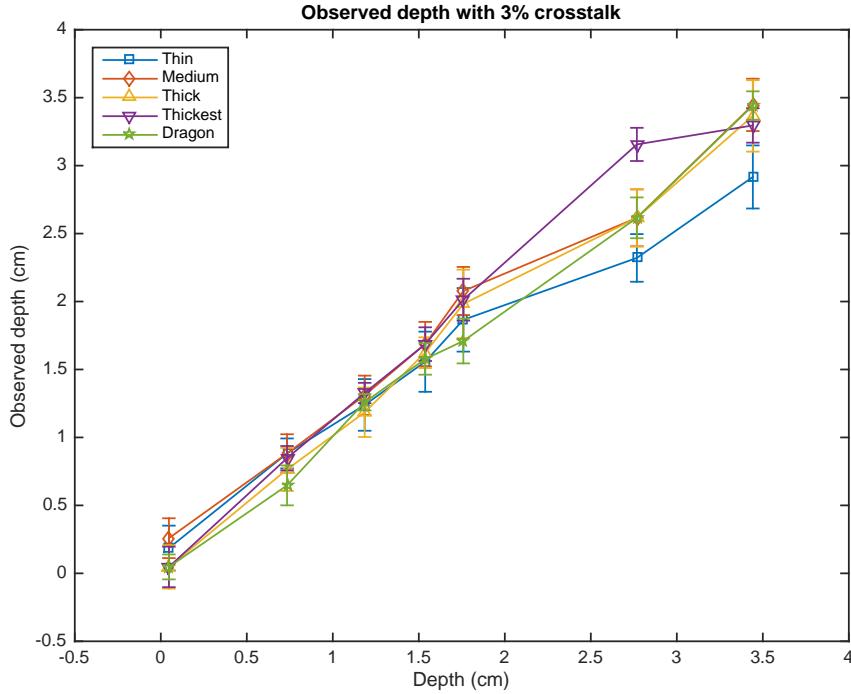


FIGURE 4.8: The median data for all observers is shown. The abscissa shows the observed depth when 3% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

Substantial effects on the observed depth start appearing once the crosstalk level is increased to 7% (see Figure 4.9). The cylinders of all widths show reduced observed depth. The reduction is most severe for the thin (18.9 arc min) cylinder. The loss of observed depth appears to be lower as the width of the cylinder increases while the observed depth of the dragon seems to be unaffected. One interesting point to observe is that for all the cylinders, the perceived depth starts degrading at the actual depths of 1.75 cm or higher. For the thin cylinder, the ghost is completely separated only at the maximum depth i.e., 3.45 cm. However the observed depth appears to be the same for the maximum and the second maximum i.e. 2.76 cm depth. Also it should be noted that the graph for the thin cylinder shows signs of degradation even at a depth of 1.13 cm.

Observed depth is severely affected for all the objects once the crosstalk reaches 14% (see Figure 4.10). The thin cylinder again shows signs of degradation starting at a depth of 1.13 cm where as the observed depths of all the other objects start deteriorating

at the theoretical depth of 1.76 cm. The pattern of wider objects showing less depth deterioration compared to thin objects is still observed. However, with such extreme crosstalk level, even the observed depth of the dragon is reported to be substantially affected.

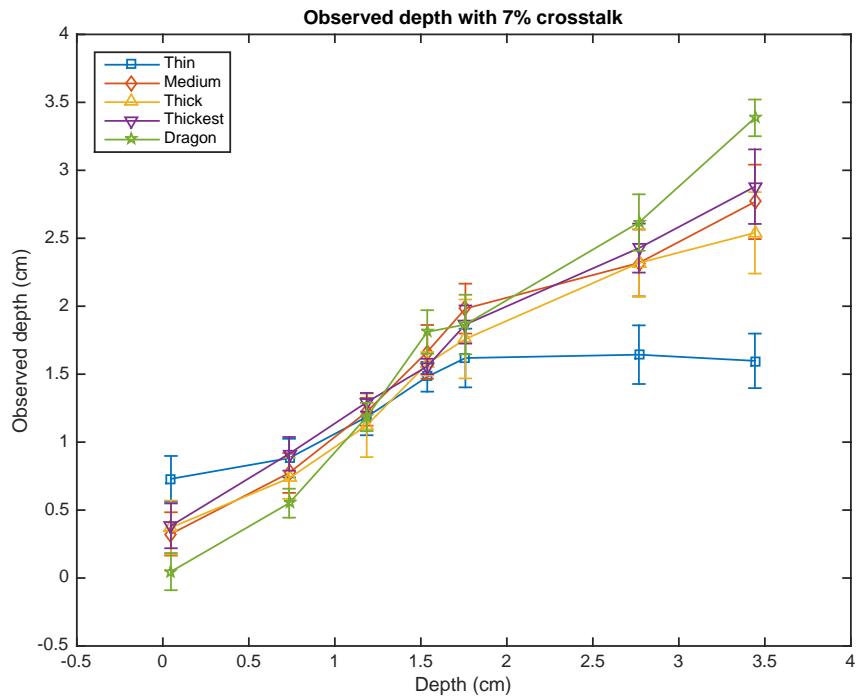


FIGURE 4.9: The median data for all observers is shown. The abscissa shows the observed depth when 7% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

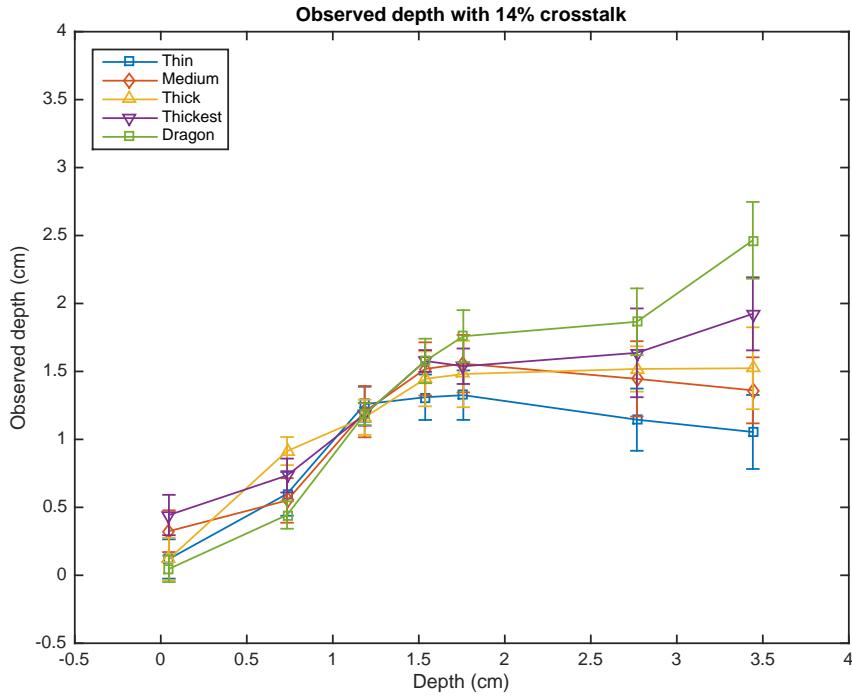


FIGURE 4.10: The median data for all observers is shown. The abscissa shows the observed depth when 14% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

4.5.2 Automultiscopic

Figure 4.11 shows the pattern of the observed depth in the base case (i.e., 0% crosstalk). It appears that the observed depth for the dragon is slightly overestimated at the actual depth of 2.76 cm. At the same instance, the observed depth of the thick cylinder is underestimated. These exceptional cases may be considered again as human error, firstly because the next greater depth is observed without any error and secondly because this case is exactly the same as the base case of the stereo experiments. However, the fact that there is mostly (both stereo and automultiscopic case) some mis-judgment of the depth may also indicate that the HVS for some reason is being confused by this particular depth.

We notice that the crosstalk has no effect on the perceived depth when the crosstalk level is raised to 3% (Figure 4.12). This is because for each of the perceived depth, the base case observed depth is in the range of the standard error. This is different than the stereoscopic scenario where we observed some minute depth reduction for the thin cylinder. Increasing the crosstalk to 7% (Figure 4.13) however, shows signs of degraded depth for the medium (37.8 arc min) cylinder depths larger than 2.76 cm. Meanwhile all

the other objects show no evidence of degraded depth even though the crosstalk level at 7% is higher than what we usually get from a conventional automultiscopic display [41].

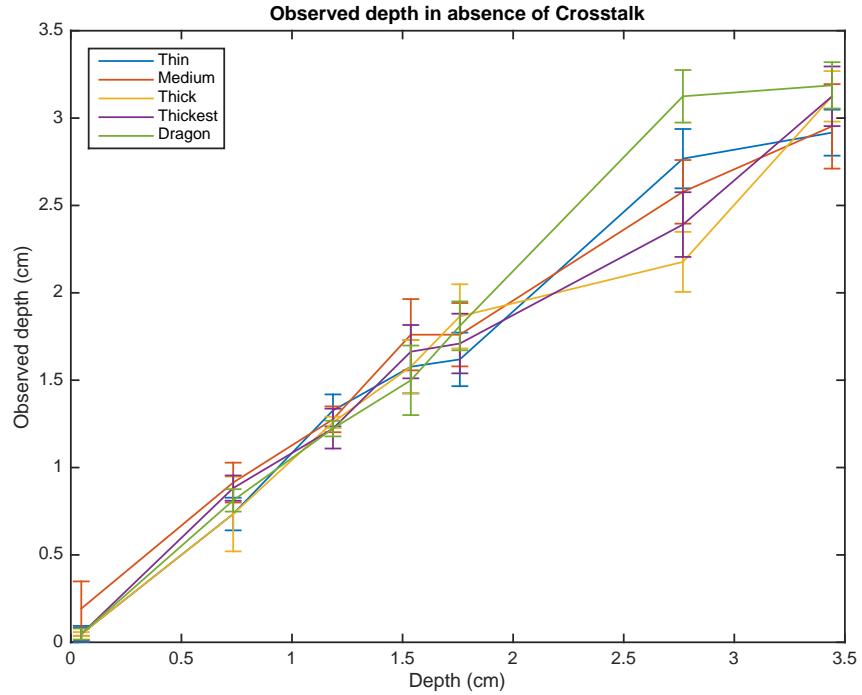


FIGURE 4.11: The median data for all observers is shown. The abscissa shows the observed depth when no crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

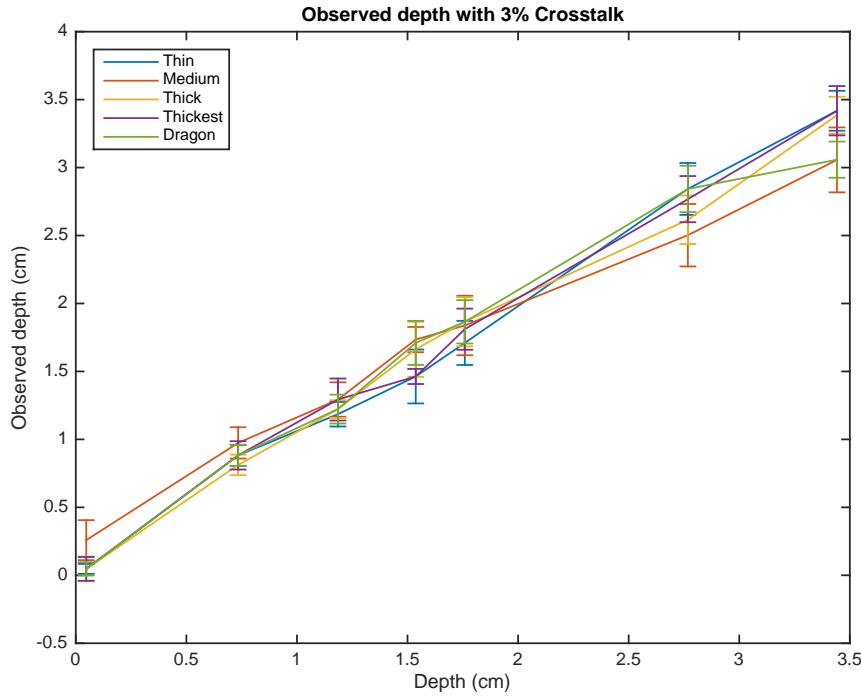


FIGURE 4.12: The median data for all observers is shown. The abscissa shows the observed depth when 3% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

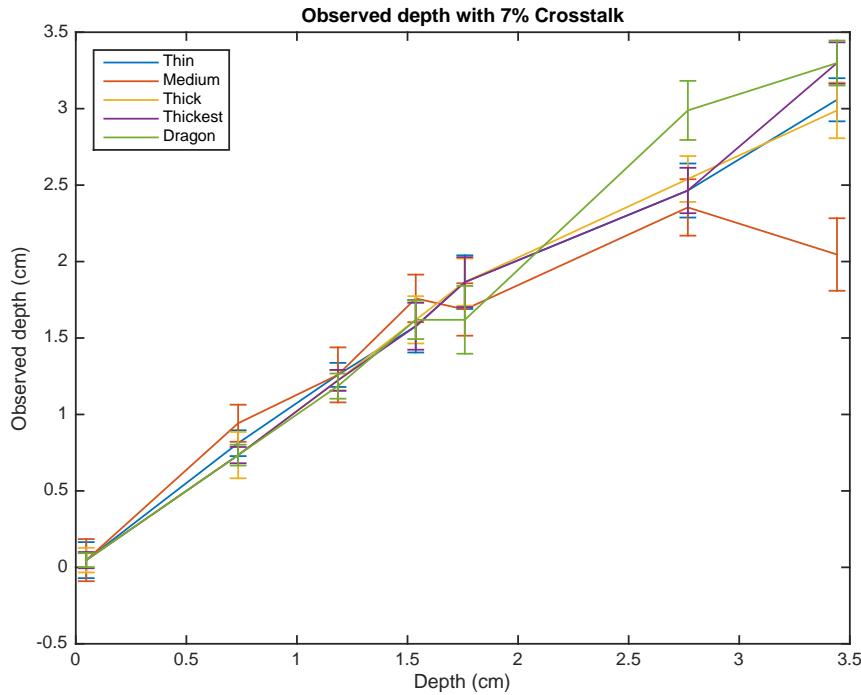


FIGURE 4.13: The median data for all observers is shown. The abscissa shows the observed depth when 7% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

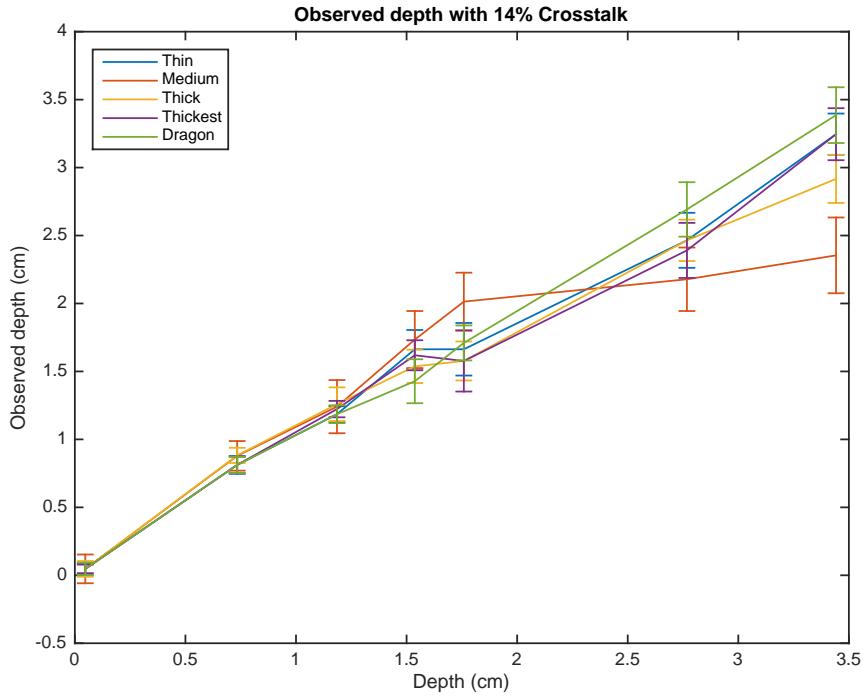


FIGURE 4.14: The median data for all observers is shown. The abscissa shows the observed depth when 14% crosstalk is present, in comparison with the theoretical depth. Different lines represent different stimuli. The error bars indicate ± 1 standard error.

Interestingly, increasing the crosstalk level to 14% also shows that for most cases, the perceived depth is not different than the depth observed in absence of crosstalk. It is not clear why only the medium cylinder shows depth degradation in this case. However, based on the results it is clear that the effect of crosstalk on the perceived depth is significantly lower (or negligible) in the automultiscopic environment. Possible explanations for all these behaviors can be found in the next section.

4.6 Discussion

4.6.1 Stereo

In these experiments, we found that for the stereoscopic case, crosstalk has a substantial degrading effect on the magnitude of perceived depth. The perceived depth magnitude relative to the 0% crosstalk base case decreases as the crosstalk level increases. Moreover, for a fixed value of crosstalk, the degradation effect seems to be more pronounced as the disparity increases. This effect is stronger on the thin objects as compared to the wider ones. One of our initial hypothesis suggesting that the perceived depth magnitude might not be degraded once the ghost and the object separate completely (i.e. no overlap

between them) was also proven to be wrong as the results clearly show degradation even in those cases. In order to explain these results, we have to delve into the theories regarding how the HVS extracts depth from disparity.

There is evidence to suggest that the HVS uses the cross-correlation between horizontally displaced patches on the two retinal images to obtain the disparity of objects.[7][12]. The disparity, once the cross-correlation is performed, is extracted by locating the peaks in the resultant correlation profiles [5]. It is easier for the HVS to identify clear sharp peaks compared to broad noisy ones as shown in Figure 3.2. Since adding crosstalk additively to an image effectively reduces the global contrast and severely reduces the local contrast around the object, it is obvious that the HVS will have more difficulty in resolving the proper disparity, a possible reason for the observed depth degradation. Also it is believed that the patches chosen for interocular cross-correlation are filtered by two dimensional isotropic Gaussians[7] (Gaussian windows)[7]. It is not clear what the size (constant or variable) of those windows is.

If the HVS could somehow match the patch containing the ghost-object pair in one retinal image to the other, then according to windowed cross-correlation theory, because there would be no ambiguity, disparity computed would either be zero or the actual disparity without any degradation. However, we think that the HVS in case of stereo only uses the object-containing patch in one retinal image to obtain the horizontal disparity based on correlation with the other image. One of the reasons for this is that HVS has a certain limitation and it cannot fuse two objects together if the disparity gradient (defined as the ratio of disparity difference between two objects and their cyclopean angular separation) is greater than 1-1.5 [9] [12]. In case of stereo pairs, where the objects in both images have a crossed disparity, their ghosts have an uncrossed and equal disparity relative to the plane of focus. This, in terms of geometry, is only possible if the ghost is located exactly behind the object (similar to the double nail illusion as illustrated in [26] [15]) making the disparity gradient ∞ . In such cases, the stimulus with higher luminance (the actual object) is fused leaving the ghost in diplopia state. The other reason is that the HVS also takes into consideration the polarity of the luminance changes between patches in order to find a correct match. This means that an object whose left side is black and right side is white will not be matched to the object in the other retinal image who's left side is white and right side is black [9]. In case of stereoscopy with added crosstalk, the polarity of the ghost-object luminance change in one retinal image is the opposite of the other. Hence it is reasonable to assume that a patch that only contains the object of interest from one retinal image is chosen to compute the horizontal interocular cross-correlation with the other. We tried simulating the disparity estimation via interocular cross-correlation using the model proposed in [7] and found that the crosstalk had no effect on the computed disparity.

For the x, y position of an object in one of the stereo images, the ghost in the other image is present at the exact same location while the actual object is located at some horizontal disparity d . Computing the cross-correlation in this case would result in two distinct peaks, one at zero disparity and the other at disparity d . Hence the HVS might apply some averaging in order to obtain a disparity. This, however, is only possible when the disparity is large enough to separate the ghost from the object completely. As seen in Figure 5.1a, this is not the case when the ghost is not fully separated from the object. In this case the correlation will result in one peak. Considering the correlation profile, we observe that there is difference between the slope around zero disparity (location of the ghost) and the actual disparity d . It is known [9] that the HVS also uses the first derivative of luminance in order to find matching portions of stereo images. In that case, the HVS can still isolate two distinct peaks in the cross-correlation profile where one peak results at zero disparity, representing maximum correlation with the ghost, and the other at disparity d , representing the maximum correlation with the object. Since the HVS gives preference to lower disparities [15], if the averaging is performed based on the heights and the distances (relative to the zero disparity) of the peaks, then this might explain the loss of perceived depth magnitude that is proportional to the crosstalk level. This would also explain why at some crosstalk level threshold (16% in our experiments), the ghosts are matched to the original objects and the viewer perceives two copies of the object side by side (double nail illusion [26]) located at the plane of focus. However, this will still not explain why wider objects result in lower loss of perceived depth magnitude as compared to thin objects.

In our experiment results, we also noticed that the perceived depth magnitude remains unaffected for some small disparities even when the crosstalk level is at 14%. For 3% crosstalk, this threshold is the angular disparity of 5.25 arc min (1.75 cm depth) whereas, for the crosstalk levels of 7% and 14%, these thresholds seem to be 4.2 arc min (1.5 cm) and 3.15 arc min (1.18 cm) respectively. This suggests that the HVS probably does not average the disparities at locations of interocular cross-correlation peaks as that would have resulted in a degraded perceived depth at every disparity. Our proposed modified windowed interocular cross-correlation model explained in chapter 5 might explain these results.

Our results are consistent with the current literature and show a clear degrading effect on the perceived depth that increases with the increase of crosstalk level and the decrease of object's width. Kooi et al [14] showed that as low as 5% crosstalk can result in reduced visual comfort. Considering that and based on our results, we recommend that the crosstalk levels in stereoscopic 3D displays should not exceed more than 3%. At this level we found that the perceived depth is not significantly degraded for stimuli of any shape or geometry.

4.6.2 Automultiscopic

At first impression, one might hypothesize that the crosstalk should have a greater effect on the observed depth in automultiscopic scenes for the simple reason that the number of ghosts per image is great (at least one on each side of the object as opposed to only one ghost in case of stereo). Our results however, indicate that there was no substantial effect of crosstalk on the perceived depth magnitude in automulticopic environment. We believe that the HVS in case of automultiscopic environment, given the existing symmetry, matches the patches containing the tuple (left-ghost, object and right-ghost) in one retinal image to the similarly appearing patch in the other one. One reason for this as mentioned in the previous section is that the HVS considers the polarity of luminance changes in the object. The tuples in both the retinal images have the same luminance profiles. The other reason is that matching the tuples containing patching in this way will not violate any disparity gradient which was the case with the ghosts in stereoscopic environment.

TABLE 4.3: Ghosts and objects seen in an automultiscopic environment according to Equation 4.3

Image	Stimulus	Location(disparity)
Left eye image	Left ghost	0
	Object	$\frac{d}{2}$
	Right ghost	d
Right eye image	Left ghost	$-d$
	Object	$-\frac{d}{2}$
	Right ghost	0

It can be seen from table 4.3 that the tuple appearing in the left eye is located at a disparity d in the right eye image. And since matching tuple containing patches will not cause any confusion (due to disparity gradient violation etc), it is clear why the viewers were able to report the proper depths (in most cases) for the objects despite of raising the crosstalk level to 14%. Based on our results we can conclude that crosstalk does not have any significant effect on the observed depth in automultiscopic displays. However, since the number of ghosts are twice the amount found in stereoscopic screens, we can predict that crosstalk in automultiscopic screen may affect the viewers visual comfort to a greater extent. Our experiments simulated only for the sweet spot of an automultiscopic screens. It should be noted that even though amount of light leakage from neighboring views will change as the viewer moves away from the sweet spot, the pattern of the luminance changes throughout the objects will remain the same. Hence, the observed depths of the objects will not be affected. The observed geometry of the

objects however might change (depending on the scene) due to crosstalk if the viewer is not located at the sweet spots.

Given the fact that the amount of crosstalk changes as the viewer moves his/her viewing position, and the fact that crosstalk might change the observed geometry of the objects in the scene while viewing from a non-sweet spot position, it is difficult to recommend a minimum level of crosstalk that ensures a pleasant and comfortable viewing experience. We know that as low as 5% of crosstalk in a stereoscopic display is enough to result in visual discomfort. We can predict that this threshold would be much lower for an automultiscopic display. However, further research must be carried out in order to provide such objective measurements. In this work, what we did establish, for the first time, is the absence of a significant effect of crosstalk on perceived depth magnitude in the case of automultiscopic displays, which is different to what happens in stereoscopic displays.

Chapter 5

Applications

In the previous chapters we learned that in addition to viewer's discomfort, reduction of contrast, reduction of stereo acuity, and reduction of overall image aesthetics, crosstalk in stereoscopic screens also has a substantial effect on the observed depth of the disparate objects (i.e., the objects not located at the plane of focus (POF)). The general rule is that the observed depth for any object not lying on the POF will tend to fall back to the POF (losing depth) as the crosstalk level increases, or as the disparity of an object increases. The effect seemed to be more pronounced for objects with narrow width compared to the objects with comparatively wider width, the reason being that the ghost separation (which causes the confusion for HVS) for thin objects is larger for any given disparity. We also learned that, contrary to our intuition, the crosstalk in an automultiscopic screen has little to no effect on the perceived depth for objects of any width for even the most extreme level of crosstalk (14%). The main reason why crosstalk did not affect the observed depth might be that in the automultiscopic case, there are two similar ghosts present on each side of the object in both retinal images. This makes the ghost-object combination in both eyes very similar and the HVS might consider them as the same object. Hence, there is nothing to confuse the HVS in estimating the correct disparity between two retinal image patches.

Current literature agree on the theory that the HVS resolves the estimated disparities of objects in a binocular scene by computing the cross-correlation of several patches located at different positions between two perspective retinal images. Finally, we learned that in order to reduce the viewers crosstalk, most of the current state of the art techniques rely upon preprocessing the stereo images in such a way that when, viewed on a stereoscopic or automultiscopic screen, the images with system crosstalk added resemble the desired images. Most techniques subtract the pre-calculated light intensity leakage between views from the actual images, and in addition, to using some perceptual

measures in order to mask the ghosting that still persists due to the limited dynamic range of the display.

Observed depth reduction due to crosstalk is a serious problem which might alter or degrade the artistic viewing experience for a scene. Similar to we have image quality metrics e.g. SSIM [30], MSE[36] or VDP[19] used to calculate the effects of noise in an image on its perception by a human observer as compared to the noise-free images, it would be useful to have a quality metric that is able to predict the observed depth of objects of interest in a 3D scene based on the crosstalk level of the display along with the dimensions and disparity of the object. Once this prediction is available, the scene artists can then alter the theoretical depths of the objects in order to bring the observed depths as close to the desired depth as possible. For this to work, however, we should have a good understanding of how the HVS estimates the depth from disparity between two retinal images, and hence derive an accurate HVS model. Moreover, reducing the viewer's crosstalk for any given display is also as important if not more for a better viewing experience. The most promising techniques to achieve that use complex optimizations to be performed in order to pre-process the images[27]. It would be helpful if similar effects can be achieved without the mentioned complex and time consuming optimizations.

In this chapter, we will look into a proposed modified HVS depth from disparity resolution model that provides a good initial approximation to the observed depths obtained from our experiment results discussed in the previous chapter. We will also propose two new crosstalk reduction techniques along with there advantages and disadvantages and their results.

5.1 Observed depth prediction

Filippini et al [7] proposed a model that simulated the depth estimation of the HVS via disparity. In short, this model computes the local cross-correlation that is defined by Eq 5.1.

$$c(\delta_x) = \frac{\sum_{(x,y) \in W_L} [(L(x,y) - \mu_L)(R(x - \delta_x, y) - \mu_R)]}{\sqrt{\sum_{(x,y) \in W_L} (L(x,y) - \mu_L)^2} \sqrt{\sum_{(x,y) \in W_R} (R(x - \delta_x, y) - \mu_R)^2}}, \quad (5.1)$$

where

$$W_L = W_R = e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} \quad (5.2)$$

are anisotropic Gaussian windows (patches) in the left and the right stereo images. The window W_L is fixed in one image while W_R is displaced horizontally while keeping the vertical position constant. For each displacement δ_x , $c(\delta_x)$ is computed resulting in

a value in the range [-1,1] (1 if the patches are perfectly correlated, -1 if the patches have no correlation at all). The value of δ_x for which the maximum cross-correlation is attained is considered to be the disparity of the pixel (centered at the window) in the right image w.r.t the left image. The process is repeated for all the pixels in the left image. The depth estimates using this model, and the human observed depths of the peaks and troughs in random dot stereogram consisting of saw-toothed corrugations of different frequencies and amplitudes were compared by the authors. They found that in addition to their estimated depths matching closely with the viewer's observed depths, their model also correctly estimated the HVS's threshold (for distinguishing between disparities) for the upper and lower spatio-temporal frequencies. However, the random dot stereograms used were devoid of any crosstalk.

We tested the same model on our stimuli (Chapter 4) and found that this model always resulted in the correct disparity of the objects between a stereo image pair. That is, the crosstalk had no effect on the disparity the disparity estimated by the model. This did not match the results we obtained in our experiments with the human observers. The reason is that even though the areas on the images where the ghosts are present will show some positive correlation, the maximum correlation is always obtained at the location where the actual object is located.

Hence it is clear that the HVS does not simply estimate the depth from disparity as the maximum of local cross-correlation only. It is believed that the HVS, while matching patches between retinal images, prefers lower disparity over higher ones [9]. This means that, for any patch in e.g. the left eye retinal image, if the HVS finds two matching patches in the right image, it will choose the one that is closest to the location of the left image patch. Recall from the previous chapter that in the case of stereo displays with some level of crosstalk, for a disparate object in the left image, the same object is located at some horizontal disparity d where as the ghost is located exactly at zero disparity (same location as the location of the object in the left image). Computing a cross-correlation in this case would result in some positive correlation because of the ghost at zero disparity and a highly positive correlation at disparity d . To match the results from human observers, the disparity estimator model should have the following characteristics:

- Select a disparity d based local cross-correlation.
- In the presence of crosstalk, estimated disparity d should shift towards zero disparity gradually as the actual disparity increases over some threshold.

- Estimated disparity should be estimated as 0 if the crosstalk level increases over some threshold (e.g. more than 14%) or if the disparity is above some level (exceeding Panum's fusion area ¹).

Hence we may assume that the model should somehow weigh the resulting cross-correlation profile prior to obtaining the disparity estimate. We observe that weighing the cross-correlation profile with a Gaussian function centered at the location of ghost in the left eye image fulfills the aforementioned requirements. This can be seen in Fig 5.1.

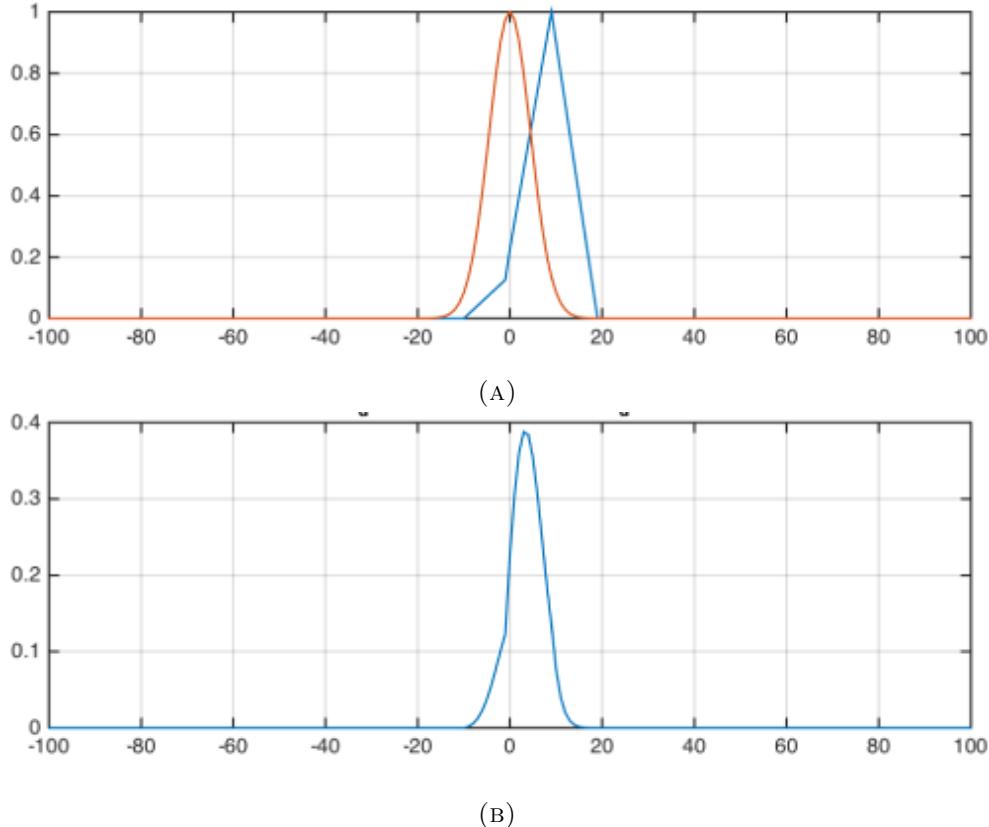


FIGURE 5.1: The figure represents estimating the human observed disparity in a thin cylinder stereo pair where the disparity in terms of pixels is 9 and the crosstalk is 14%. The cross-correlation of a patch containing only the cylinder from the right eye image is computed across the entire left eye image containing both the cylinder and the ghost. (A) The blue line represents the 1D The scan line of the result of local cross-correlation. In this case, the maximum cross-correlation occurs at the disparity of 9 pixels, which is the actual disparity. Orange line represents a Gaussian function with $\sigma = 9$ pixels centered at the location of the ghost that is to be used for weighing the correlation profile. (B) The resulting Gaussian weighed local cross-correlation. It can be seen that, the estimated disparity as a result of the peak of cross-correlation has shifted back from 9 to 3 pixels.

In our disparity estimation model, we assume that the HVS will be able to isolate the object of interest in one (source) retinal image (due to its luminance profile) and

¹Panum's fusion area is the space around the POF where single binocular vision is observed. Double vision (diplopia) is observed for objects outside the Panum's fusion area.

try to find an appropriate match for that object in the other target image. The target image has of two possible matches i.e. the ghost and the object itself. For simplicity, we show the resulting cross-correlation for a single scan line of the stereo image pair. This is sufficient to estimate the disparity of a single fixed width object. Figure 5.1a (blue line) represents the cross-correlation profile for a cylinder that has a width of 9 pixels and is located at the disparity of 9 pixel. The background for simplicity in this case is black. We can see that there is some positive correlation at zero disparity due to the 14% crosstalk induced ghost. However, the full correlation (+1) is attained at the disparity 9, indicating the actual disparity of the object. The orange line represents how the resulting correlation profile is weighted by a truncated Gaussian that has a standard deviation of 9 pixels. The Gaussian weighed correlation profile in Figure 5.1b is obtained by simple point-wise multiplication of the two functions in Figure 5.1a. It can be seen that after weighing with the Gaussian window for which the standard deviation σ is roughly equal to the width of the object, the maximum correlation occurs at a disparity of 3 pixels. This means that, the perceived depth via disparity would be approximately 3 pixels (around 1.18 cm). So, in this, case the estimated disparity is quite close to the result that we obtained from the experiments.

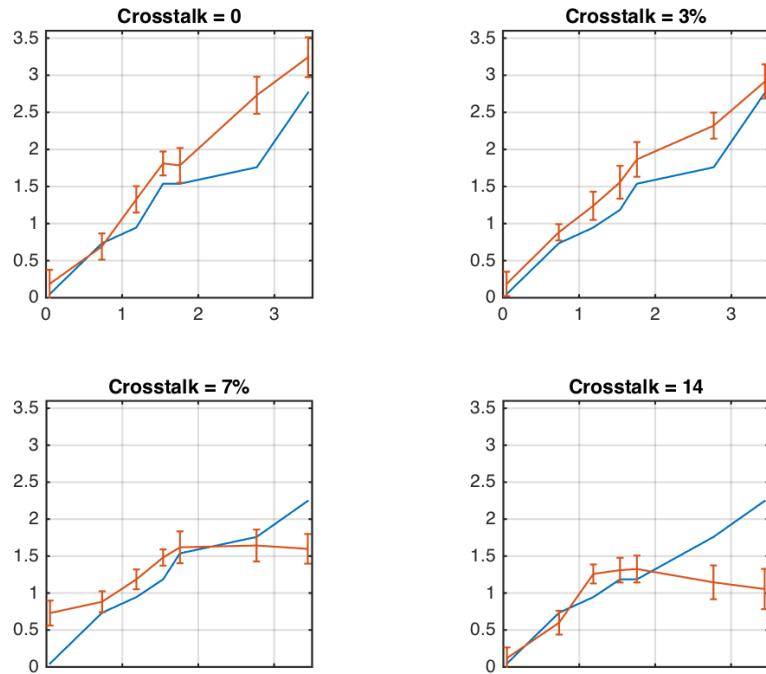


FIGURE 5.2: Comparison of experimental observed depth for cylinder of 18.9 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 15.4$ arc min.

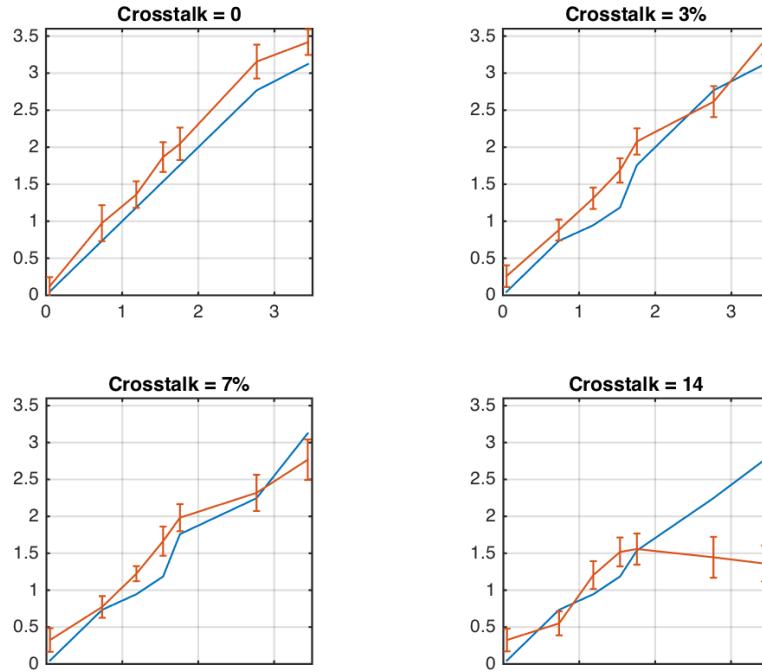


FIGURE 5.3: Comparison of experimental observed depth for cylinder of 37.8 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 30.87$ arc min.

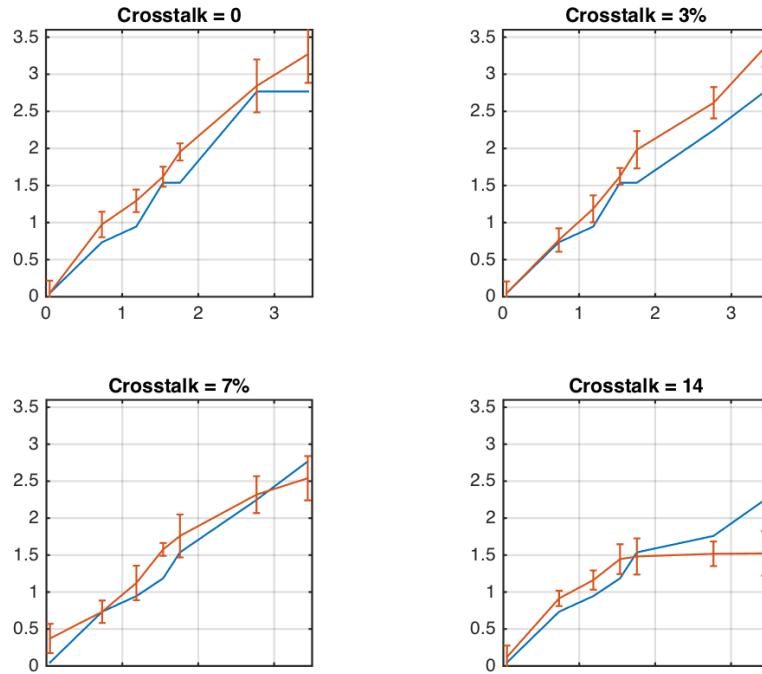


FIGURE 5.4: Comparison of experimental observed depth for cylinder of 56.7 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 36.54$ arc min.

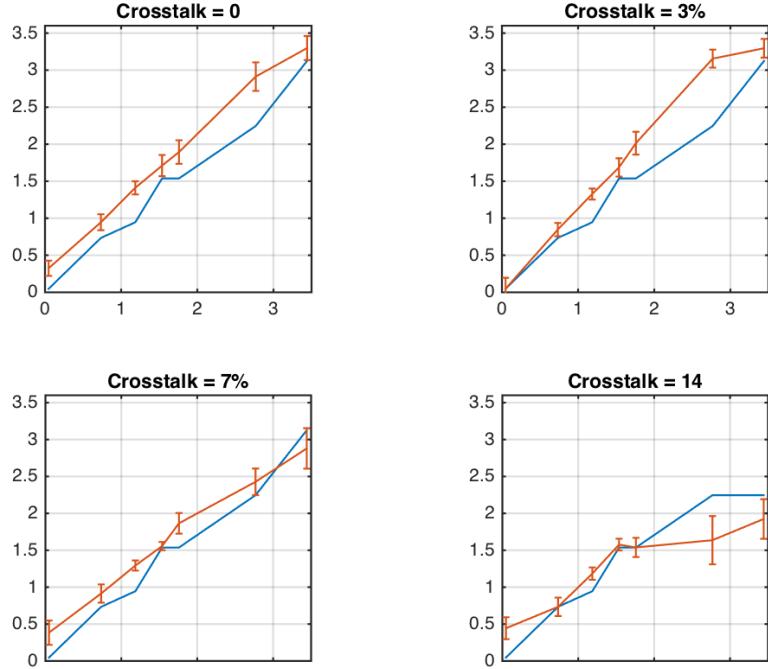


FIGURE 5.5: Comparison of experimental observed depth for cylinder of 75.6 arc min width (orange) vs estimated depth via windowed cross-correlation. The best results were obtained with $\sigma = 46.2$ arc min.

We simulated our stereo experiments for the cylinders and computed the estimated depths using our Gaussian weighted cross-correlation model. Figure 5.2 to Figure 5.5 shows the results that we obtained and there comparison with the data we got from our experiments. We observed that even though this model had some shortcomings and did not fit to all of the data properly, it was able to mimic the degradation of estimated depth as the crosstalk level or the disparity increased. The figures above represent the estimated depths using our Gaussian weighted local cross-correlation model with fixed σ for each cylinder that best matched the data. For almost all the cases (specific disparities and crosstalk levels) where the estimated depth was incorrectly estimated, we were able to get approximately matching results using some σ . However, we were not able to find one σ that matched all cases for a particular cylinder. Perhaps the HVS weighs the cross-correlation profile with Gaussians of different standard deviations for different object widths and disparities. Our Gaussian window-weighted local cross-correlation model has the advantage that it resolves to a degraded estimated depth in the case where certain level of crosstalk is present. However, the shortcoming is that it also, to some extent, penalizes the estimated depth even if there is no crosstalk present. This does not conform fully to our experimental data. However, it may give the right direction for the future research in this area.

5.2 Crosstalk mitigation

During the course of this thesis, we devised and tested some new ideas to improve the current crosstalk reduction techniques. In this section we review those ideas.

5.2.1 Unsharp masking in epipolar domain

The Corn-sweet illusion [32] is used in contrast boosting image processing techniques such as unsharp masking. The basic idea of unsharp masking is to add to an image a high frequency image of itself. The high frequencies containing image is obtained by subtracting a blurred version of the image from itself. Baar et al [27] used a similar technique to blur the uncorrectable crosstalk in combination with increasing the contrast of the objects around the edges. One can generally use unsharp masking to increase the contrast of an image suffering from contrast loss due to preprocessing for crosstalk compensation . However applying unsharp masking in spatial domain also boosts the unwanted noise. Since automultiscopic screens displays lightfields, and given the fact that crosstalk in any particular view is induced by the neighboring views, intuitively, it would make sense to applying unsharp masking in the view domain (also called as the epipolar domain) that would result in a crosstalk compensated lightfield with increased local contrast around the object. Let Ψ be a lightfield in which x, y denote the spacial dimensions on the displayed image and w denotes the epipolar domain (Figure 5.6)

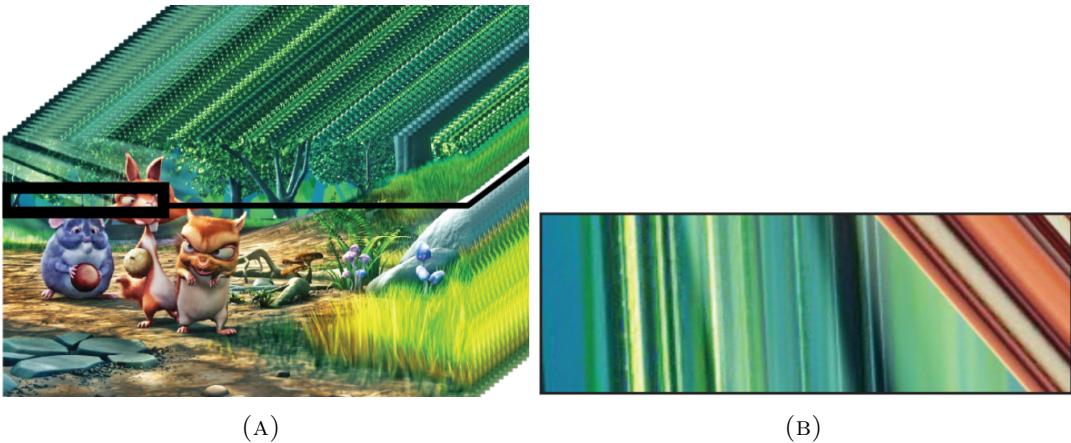


FIGURE 5.6: Representation of a 3D lightfield [6]. (A)(x, y) denotes the view dimensions while w denotes different perspective views for a light field. (B) Epipolar view of the squared region in (a)representing how every pixels shifts with respect to shift in perspective.

Then for an automultiscopic screen where only the light from the immediate neighbors is leaked, we blur the lightfield in epipolar domain with the following 1D kernel:

$$\boxed{\omega_1 \quad \omega_2 \quad \omega_3}$$

This will result in a epipolar domain blurred light field Ψ_{blr} . Next we subtract Ψ_{blr} from the original light field Ψ . This will give us Ψ_{hf} , a light field that only contains luminance-inverted ghosts in every view. i.e.

$$\Psi_{hf} = \Psi - \Psi_{blr} \quad (5.3)$$

The crosstalk-compensated light field is obtained as

$$\Psi_{opt} = \Psi + \Psi_{hf} \quad (5.4)$$

Hence, for every view image $I \in \Psi_{opt}$

$$\begin{aligned} I_{opt} &= I + (I - \omega_1 I_L - \omega_2 I - \omega_3 I_R) \\ I_{opt} &= (2 - \omega_2) I - \omega_1 I_L - \omega_3 I_R \end{aligned} \quad (5.5)$$

Where I is the view centered image and I_L, I_R are the immediate left and right neighbors respectively. Since the automultiscopic screen will display the optimized light field image as a weighted sum of I_L, I_R and I i.e.

$$\begin{aligned} I_{observed} &= \Phi_1 \cdot I_{L(opt)} + \Phi_2 \cdot I_{(opt)} + \Phi_3 \cdot I_{R(opt)} \\ I_{observed} &= \Phi_1 (2 - \omega_2) I_L - \omega_1 I_{LL} - \omega_3 I + \\ &\quad \Phi_2 (2 - \omega_2) I - \omega_1 I_L - \omega_3 I_R + \\ &\quad \Phi_3 (2 - \omega_2) I_R - \omega_1 I - \omega_3 I_{RR} \end{aligned} \quad (5.6)$$

Here I_{LL} and I_{RR} are the immediate left and right images to left and right image of the view position in the light field. We know, from view-luminance intensity profiles for an automultiscopic screen 4.4, that:

$$\Phi_1 = \Phi_3 = 0.0351, \quad \Phi_2 = 0.8 \quad (5.7)$$

In order to cancel the effect of I_L and I_R for crosstalk compensation, we choose

$$\omega_1 = \omega_3 = 0.0421, \quad \omega_2 = 1.19 \quad (5.8)$$

Using the aforementioned values of $\omega_1, \omega_2, \omega_3, \Phi_1, \Phi_2$ and Φ_3 , the viewer observed image $I_{observed}$ can then be mathematically written as

$$I_{observed} = 1.42.I - 0.0014(I_{LL} + I_{RR}) \quad (5.9)$$

It can be seen from Equation 5.9 that even though the crosstalk effect induced by the immediate left and right image of a view has been eliminated. However, some over-subtraction of luminance due to the I_{LL} and I_{RR} propagates in the observed image $I_{observed}$. Even though compensating for crosstalk in this manner is quite efficient, one can see the average luminance of the view centered image I has been increased to 1.42. Since the maximum luminance value an automultiscopic screen can display is 1, our proposed technique will result in clipping of all the luminance values above 1 hence resulting in loss of contrast. Another drawback, as described in Figure 5.7, unsharp masking in view domain of a light field does not introduce any Cornsweet profiles.

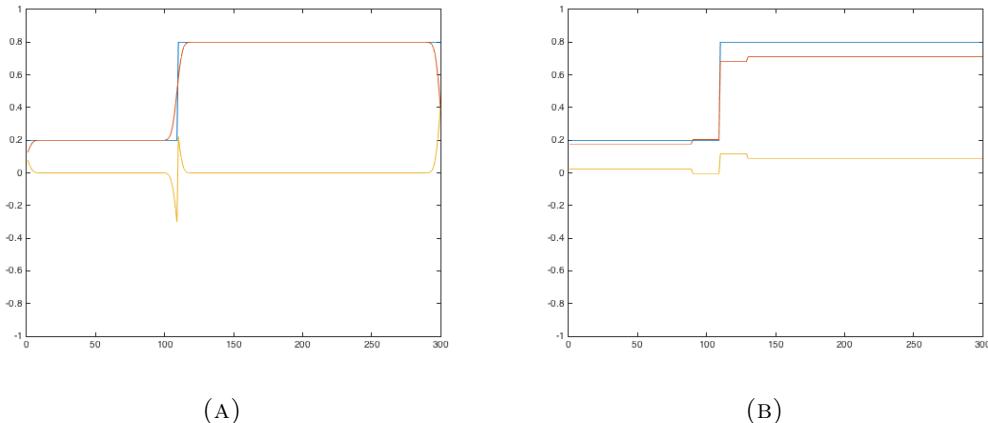


FIGURE 5.7: (A) 1D representation of luminance profile along an image row. Blue graph represents a luminance step and orange graph represents the luminance profile of the blurred image. Yellow graph represents corn-sweet profile centered at the luminance edge and obtained as the difference between the original and the blurred image. (B) Same procedure performed in view domain. Blue graph represents the luminance profile of a light field image and orange graph represents the non-energy preserving additive blurring due to neighboring views in an automultiscopic environment. The Yellow graph represents the difference between the view image and its view-domain blurred version that resembles closely to subtractive crosstalk cancellation rather than a corn-sweet profile.

5.2.2 Iterative crosstalk reduction

To the best of our knowledge, all the current subtractive crosstalk reduction techniques compute the pre-processed crosstalk compensated images by subtracting the amount of leaked light between neighboring views. The calculation of subtracted leaked light is always performed while considering the unmodified stereo image pair. We observe

that the amount of light leaked into a view image due to the unmodified alternative view image should be different when the alternate view image is already crosstalk compensated. This is the case in reality and hence we think that a view image should be compensated appropriately if its alternative view image is already crosstalk compensated. Inspired by this idea, we propose an iterative subtractive crosstalk reduction technique.

Consider a light field Ψ consisting of a set of view images f_1, f_2, \dots, f_N . When the i^{th} view is displayed on an automultiscopic screen, its crosstalk $\phi(f_i)$ from the neighboring views can be written as

$$\phi(f_i) = a_1.f_1 + a_2.f_2 + \dots + a_{i-1}.f_{i-1} + a_{i+1}.f_{i+1} + \dots + a_n.f_n \quad (5.10)$$

Where the light intensity leakage values $a_1 \dots a_n$ are given by the curves in Figure 4.4. The crosstalk compensated images $\{\gamma_1 \dots \gamma_n\}$ can then be computed iteratively as:

$$\begin{aligned} \gamma_1^{n+1} &= f_1 + \phi(\gamma_1^n) \\ \gamma_2^{n+1} &= f_2 + \phi(\gamma_2^n) \\ &\vdots \\ \gamma_N^{n+1} &= f_N + \phi(\gamma_N^n) \end{aligned} \quad (5.11)$$

Where n is the number of current iteration. After each iteration, clipping of the values exceeding the display's dynamic range is performed. The iterations terminate when the error ϵ falls below a threshold.

$$\epsilon = \sqrt{(\gamma_1^{n+1} - \gamma_1^n)^2 + \dots + (\gamma_N^{n+1} - \gamma_N^n)^2} \quad (5.12)$$

This technique compensates for the crosstalk more appropriately. However, the successive subtraction of luminance from the light field view images between iterations might result in lower average luminance in the crosstalk compensated light field (Figure 5.8).

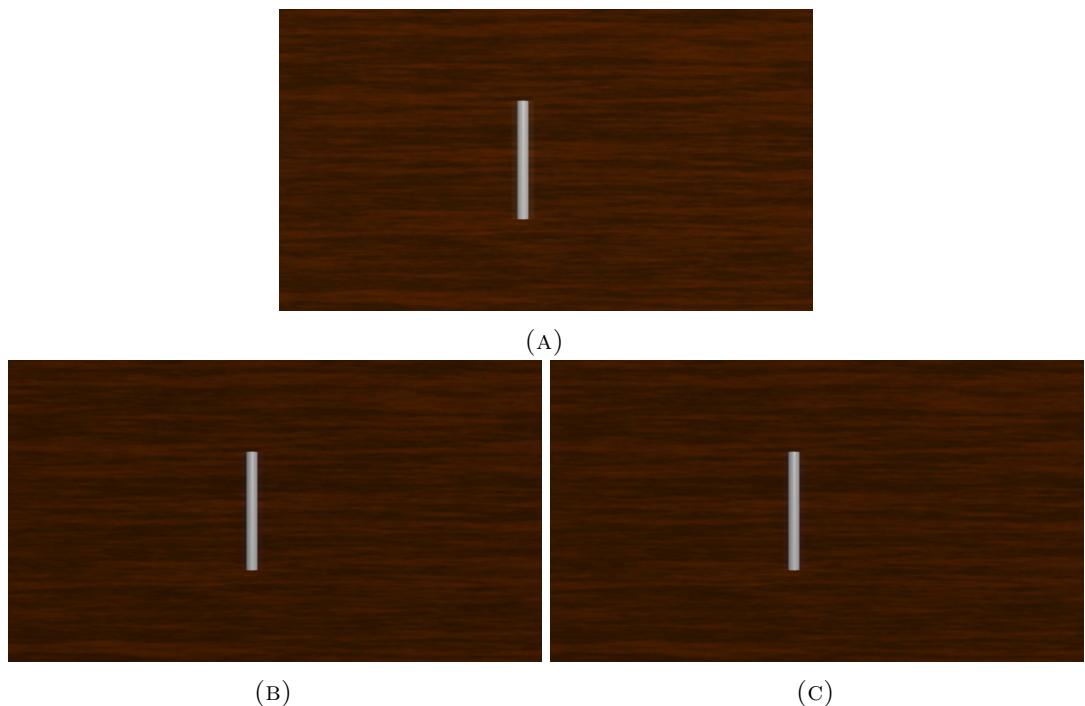


FIGURE 5.8: Naive subtractive crosstalk compensation compared to our proposed iterative crosstalk compensation. (A) An automultiscopic display's view image containing crosstalk given by the curves in Figure 4.4. (B) Compensated view image using naive crosstalk subtraction. (C) Compensated view image using iterative crosstalk subtraction.

Chapter 6

Conclusion

6.1 Summary

In this work, we show for the first time, how the crosstalk affects the perceived depth in automultiscopic displays. In addition to that, we also show how the perceived depth is affected for natural stimuli of different dimensions and geometry in stereoscopic displays. From our experiments we verified that in stereoscopic environment where each object has one ghost located on either side, crosstalk had an adverse effect on the perceived depth magnitude. The perceived distance of the objects from the fixation point appeared to be lower. This loss of depth in general, is directly proportional to the crosstalk level and the disparity of the object, i.e., higher levels of crosstalk or disparity resulted in a greater loss of observed depth. This loss of depth is more severe for thin objects rather than thick ones or ones with complex geometry. In an automultiscopic environment however, where two copies of identical ghosts are located on each side of the object, we found no significant loss of depth. The reason for this might be the identical nature of the ghost-object combination in both eyes, the symmetry of the configuration leading to a maximum cross-correlation at the real depth disparity.

On the contrary, in the stereoscopic case, the HVS has to make a confused decision between matching the object in one retinal image to either the ghost (located at zero disparity) or the original object (located at actual disparity) in the other retinal image. After the experiments, a debriefing session was held with the test subjects regarding the overall experience. Even though human observers reported no loss of perceived depth in the automultiscopic case, they did report that having to deal with a greater number of ghosts added to their viewing discomfort more than the stereoscopic case. Also from our experiments for the stereoscopic environment, we found that depth loss due to crosstalk was negligible for all stimuli as long as the crosstalk levels were kept under 3%. Due to

this we predict that crosstalk level below 3% should not have any effect of the perceived depth of the objects when viewed on a stereoscopic display.

In order to predict the viewer's observed depths of stimuli when presented on a 3D display, understanding how the HVS resolved for disparity between two retinal images is important. Hence, a faithful HVS model is required. Current HVS disparity estimation model, that estimates the disparity by computing the cross-correlation of patches between two retinal images works well in estimating the disparity of stimuli when no crosstalk is present. It however, fails to correctly estimate the disparities when crosstalk is added to the stimuli. In this work, we propose a modified HVS disparity estimation model that weighs the computed cross-correlation between patches with respect to the crosstalk level and the disparity appropriately. This way, the HVS model estimates the disparity of stimuli that is closer to the actual human observed disparities/depths we showed in our experiments.

6.2 Future work

From our experiments it is clear that crosstalk affects the HVS's stereopsis and, in order for these effects to be negligible, the crosstalk levels in a stereoscopic display should be no more than 3%. We also know that 5% crosstalk in stereoscopic displays is enough to induce viewing discomfort. However, we do not know what the minimum threshold for crosstalk in an automultiscopic display should be. A user study would be beneficial to find this limit. More importantly, it is imperative to fully understand how the HVS resolves depth information from a 3D scene. This will not only help us better predict the observed depth of a scene when viewed through a stereoscopic display, but will also help us improve the current crosstalk reduction techniques or help us come up with new ones. Thus, work in more comprehensive models of depth perception is required.

Bibliography

- [1] BARKOWSKY, M., CAMPISI, P., LE CALLET, P., AND RIZZO, V. Crosstalk measurement and mitigation for autostereoscopic displays. In *IS&T/SPIE Electronic Imaging* (2010), International Society for Optics and Photonics, pp. 75260R–75260R.
- [2] BOEV, A., GOTCHEV, A., AND EGIAZARIAN, K. Crosstalk measurement methodology for auto-stereoscopic screens. In *3DTV Conference* (2007), pp. 1–4.
- [3] BOUMAN, C. A. Visual perception, contrast, gamma, and mtf. Accessed: 2015-09-29.
- [4] BURT, P., AND JULESZ, B. A disparity gradient limit for binocular fusion. *Science* 208, 4444 (1980), 615–617.
- [5] CORMACK, L. K., STEVENSON, S. B., AND SCHOR, C. M. Interocular correlation, luminance contrast and cyclopean processing. *Vision Research* 31, 12 (1991), 2195–2207.
- [6] DU, S.-P., DIDYK, P., DURAND, F., HU, S.-M., AND MATUSIK, W. Improving visual quality of view transitions in automultiscopic displays. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 192.
- [7] FILIPPINI, H. R., AND BANKS, M. S. Limits of stereopsis explained by local cross-correlation. *Journal of Vision* 9, 1 (2009), 8.
- [8] HONG, H., JANG, J., LEE, D., LIM, M., AND SHIN, H. Analysis of angular dependence of 3-d technology using polarized eyeglasses. *Journal of the Society for Information Display* 18, 1 (2010), 8.
- [9] HOWARD, I. *Binocular vision and stereopsis*. Oxford University Press, New York, 1995.
- [10] HUANG, K., YUAN, J.-C., TSAI, C.-H., HSUEH, W.-J., AND WANG, N.-Y. How crosstalk affects stereopsis in stereoscopic displays. In *Electronic Imaging 2003* (2003), International Society for Optics and Photonics, pp. 247–253.
- [11] JAIN, A., AND KONRAD, J. Crosstalk in automultiscopic 3-d displays: Blessing in disguise? In *Electronic Imaging 2007* (2007), International Society for Optics and Photonics, pp. 649012–649012.
- [12] KANE, D., GUAN, P., AND BANKS, M. S. The limits of human stereopsis in space and time. *The Journal of Neuroscience* 34, 4 (2014), 1397–1408.
- [13] KONRAD, J., LACOTTE, B., AND DUBOIS, E. Cancellation of image crosstalk in time-sequential displays of stereoscopic video. *Image Processing, IEEE Transactions on* 9, 5 (2000), 897–908.

- [14] KOOI, F. L., AND TOET, A. Visual comfort of binocular and 3d displays. *Displays* 25, 2 (2004), 99–108.
- [15] KROL, J. D., AND VAN DE GRIND, W. A. The double-nail illusion: experiments on binocular vision with nails, needles, and pins. *Perception* 9, 6 (1980), 651–669.
- [16] LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., AND YOUNG, M. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision research* 35, 3 (1995), 389–412.
- [17] LI, D., ZANG, D., QIAO, X., WANG, L., AND ZHANG, M. 3d synthesis and crosstalk reduction for lenticular autostereoscopic displays.
- [18] LIPSCOMB, J. S., AND WOOTEN, W. L. Reducing crosstalk between stereoscopic views. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology* (1994), International Society for Optics and Photonics, pp. 92–96.
- [19] MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. Visible difference predictor for high dynamic range images. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on* (2004), vol. 3, IEEE, pp. 2763–2769.
- [20] REICHELT, S., HÄUSSLER, R., FÜTTERER, G., AND LEISTER, N. Depth cues in human visual perception and their realization in 3d displays. In *SPIE Defense, Security, and Sensing* (2010), International Society for Optics and Photonics, pp. 76900B–76900B.
- [21] ROHALY, A. M., AND WILSON, H. R. The effects of contrast on perceived depth and depth discrimination. *Vision Research* 39, 1 (1999), 9–18.
- [22] SMIT, F. A., VAN LIERE, R., AND FROEHLICH, B. Non-uniform crosstalk reduction for dynamic scenes. In *Virtual Reality Conference, 2007. VR’07. IEEE* (2007), IEEE, pp. 139–146.
- [23] SOHN, H., JUNG, Y. J., AND MAN Ro, Y. Crosstalk reduction in stereoscopic 3d displays: Disparity adjustment using crosstalk visibility index for crosstalk cancellation. *Optics express* 22, 3 (2014), 3375–3392.
- [24] TSIRLIN, I., ALLISON, R. S., AND WILCOX, L. M. Crosstalk reduces the amount of depth seen in 3d images of natural scenes. In *IS&T/SPIE Electronic Imaging* (2012), International Society for Optics and Photonics, pp. 82880W–82880W.
- [25] TSIRLIN, I., WILCOX, L. M., AND ALLISON, R. S. The effect of crosstalk on the perceived depth from disparity and monocular occlusions. *Broadcasting, IEEE Transactions on* 57, 2 (2011), 445–453.
- [26] TSIRLIN, I., WILCOX, L. M., AND ALLISON, R. S. Effect of crosstalk on depth magnitude in thin structures. *Journal of Electronic Imaging* 21, 1 (2012), 011003–1.
- [27] VAN BAAR, J., POULAKOS, S., JAROSZ, W., NOWROUZEZHRAI, D., TAMSTORF, R., AND GROSS, M. Perceptually-based compensation of light pollution in display systems. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (2011), ACM, pp. 45–52.

- [28] VANGORP, P., MANTIUK, R. K., BAZYLUK, B., MYSZKOWSKI, K., MANTIUK, R., WATT, S. J., SEIDEL, H.-P., ET AL. Depth from hdr: depth induction or increased realism? In *ACM Symposium on Applied Perception* (2014), ACM, pp. 71–78.
- [29] WANG, X., AND HOU, C. Improved crosstalk reduction on multiview 3d display by using bils algorithm. *Journal of Applied Mathematics 2014* (2014).
- [30] WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on* 13, 4 (2004), 600–612.
- [31] WIKIPEDIA. Contrast (vision) — wikipedia, the free encyclopedia, 2015. [Online; accessed 27-November-2015].
- [32] WIKIPEDIA. Cornsweet illusion — wikipedia, the free encyclopedia, 2015. [Online; accessed 29-October-2015].
- [33] WIKIPEDIA. Depth perception — wikipedia, the free encyclopedia, 2015. [Online; accessed 20-August-2015].
- [34] WIKIPEDIA. Horopter — wikipedia, the free encyclopedia, 2015. [Online; accessed 21-August-2015].
- [35] WIKIPEDIA. Light field — wikipedia, the free encyclopedia, 2015. [Online; accessed 7-October-2015].
- [36] WIKIPEDIA. Peak signal-to-noise ratio — wikipedia, the free encyclopedia, 2015. [Online; accessed 25-October-2015].
- [37] WIKIPEDIA. Polarized 3d system — wikipedia, the free encyclopedia, 2015. [Online; accessed 8-September-2015].
- [38] WIKIPEDIA. Stereopsis — wikipedia, the free encyclopedia. [Online; accessed 21-August-2015].
- [39] WIKIPEDIA. Stereoscope — wikipedia, the free encyclopedia, 2015. [Online; accessed 23-August-2015].
- [40] WILCOX, L. M., AND STEWART, J. A. Determinants of perceived image quality: ghosting vs. brightness. In *Electronic Imaging 2003* (2003), International Society for Optics and Photonics, pp. 263–268.
- [41] WOODS, A. J. Crosstalk in stereoscopic displays: a review. *Journal of Electronic Imaging* 21, 4 (2012), 040902–040902.
- [42] ZWICKER, M., MATUSIK, W., DURAND, F., PFISTER, H., AND FORLINES, C. Antialiasing for automultiscopic 3d displays. In *ACM SIGGRAPH 2006 Sketches* (2006), ACM, p. 107.

Index

CSF, [27](#)

DoS, [37](#)

HVS, [1](#), [5](#)

IOC, [24](#)

LC, [11](#)

POF, [2](#), [53](#)