Universität des Saarlandes
Max-Planck-Institut für Informatik

Thesis Tytple

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

Author Name

angefertigt unter der Leitung von / supervised by

betreut von / advised by

begutachtet von / reviewers

Saarbrücken, October 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, October 2015 Mohamed Gad-Elrab

Abstract

Stereoscopic and automultiscopic displays suffer from crosstalk. An 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 the nature of the stimuli. Moreover most of the previous work is carried on simple monochromatic scenes. Since the human visual system uses numerous other cues than disparity to estimate the depth of an object in a stereo scene, monochromatic scenes are poor choice for understanding the above mentioned relation. Moreover, the model for depth resolution via disparity as provided by the current literature fails to justify why and how the perceived depth is affected by the crosstalk. In this work, we improved and performed more generalized experimentation to see how the depth perception is affected by the crosstalk for different kinds of stimuli. Based on the result of these experiments, we derived a model for human visual system's resolution of depth from disparity that accurately measures the depth of a stimulus as perceived by the human in presence of cross-talk. Finally some improved algorithms for removal/compensation of crosstalk in automultiscopic are developed.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Discuss General idea behind crosstalk and why is it so bad. How it affects the depth and what current literature think about it. Context: 3D stereo, 3D displaysvision Problem + its importance Crosstalk relevance.

1.1 Goal and overview

1.2 Contributions

Experimentation, mitigation, HVS model.

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 or not[25].

Human visual system uses several monocular and binocular cues to determine the depth of objects in the view. 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)[15][25]. Figure 2.1 gives an outlook of the depth cues used by the HVS. These cues are then dynamically weighted according to their robustness by the HVS in order to estimate a depth value for each object in the view [12](Write details of those cues in Appendix).

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 (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

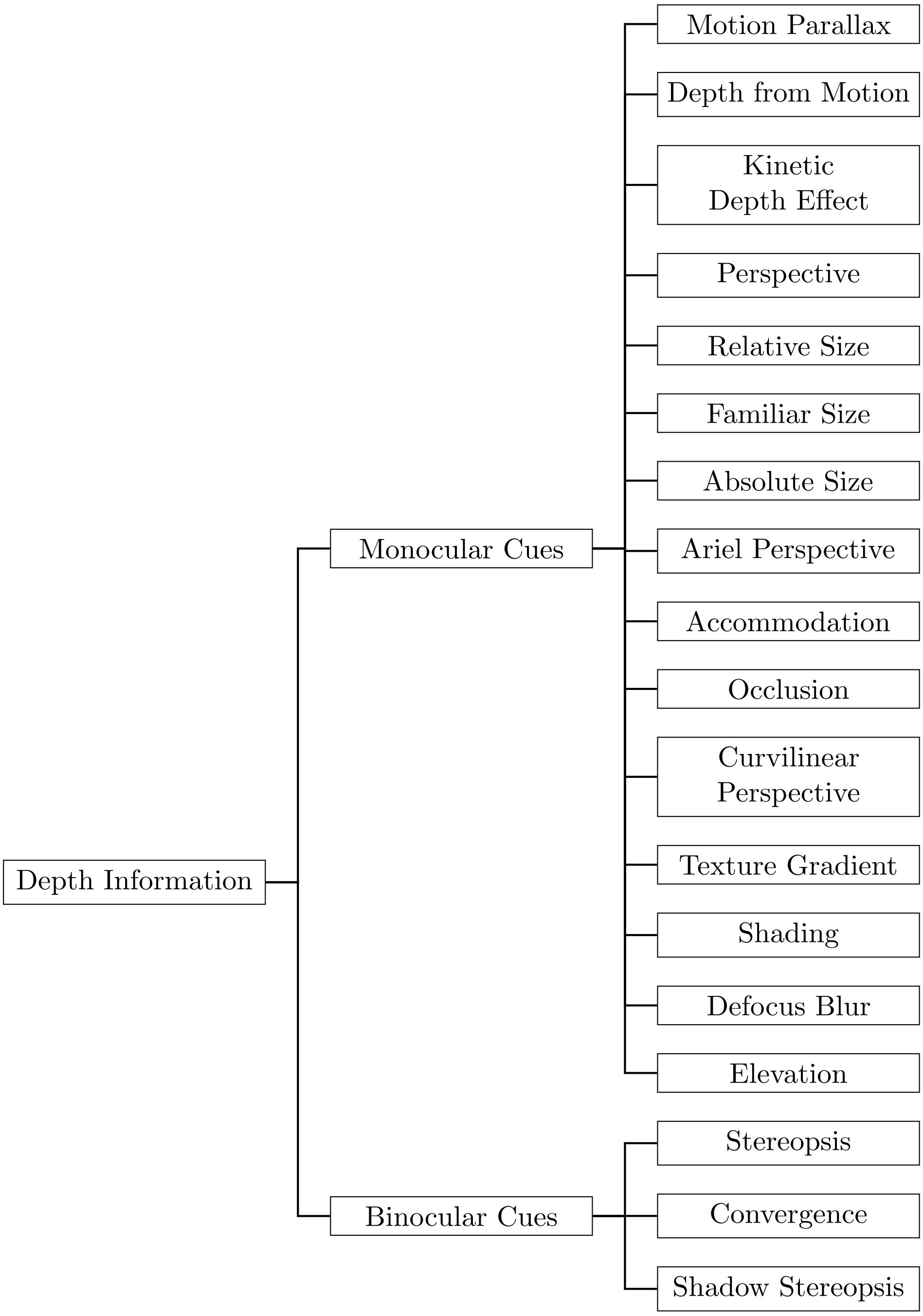


FIGURE 2.1: HVS Depth Cues

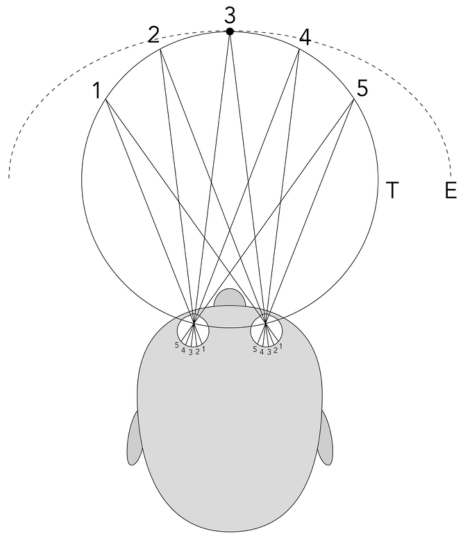


FIGURE 2.2: Representation of theoretical (T) and empirical (E) horoptor

binocular vision[8]. 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 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[29]. The process of obtaining a fused (Binocular fusion) 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 horoptor[26]. Theoretically, via geometrical principles, the horoptor is a circular segment that passes through the fixation point. However, Wheatstone in 1938 observed that the actual/emperical horopter is much larger than that. Figure 2.2 shows both the theoretical and empirical horoptor. Any object that is farther away from the horoptor 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 horoptor 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 fields 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[30]. It used two mirror both tilted at 45 degrees w.r.t eyes that reflected two different images from the sides. Currently all the stereoscopic screen present a different perspective image two both eyes with different technologies that will be discussed in later sections.

2.2 Crosstalk

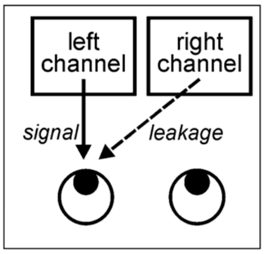
As discussed in the previous section, stereoscopy includes the process of displaying different perspective images to each eye in order to mimic the effect of depth in a scene. However, it is critical that the perspective of these two images should be segregated completely from each other. Currently, all the commercially available 3D displays (with exception of head mounted displays or Wheatstone setups) fail to isolate the two images completely. That means that a percentage of the image of one eye leaks to the other eye as well. This unintended leakage is called the crosstalk in stereoscopic screens[32]. Figure 2.3 Shows one such example where the screen has a simulated crosstalk of 14%. This 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:

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

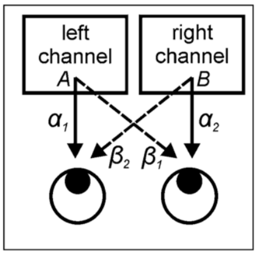
Here, "signal" is the luminance of the original image intended for an eye and "leakage" is the luminance of the light that leaks from the unintended image. Usually, on a stereoscopic screen, the amount of crosstalk is measured by observing the luminance of the left channel while a black (minimum luminance) 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 can not produce zero luminance as their minimum luminance hence resulting in a non-zero black level at its minimum. Another

(A) Original Image



(B) Original Image

(C) Original Image



(D) Viewers Crosstalk

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 Andrew J. Woods [32]

mathematical representation of crosstalk take into effect this non-zero black level and is defined as:

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

Throughout the rest of the thesis, we will be using crosstalk as mentioned in 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

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

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

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

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 law which means that leaked light will be greatly perceivable by the viewer on the dark image areas rather than bright areas. Also, the perceived crosstalk is dependent upon the contrast of the image and the binocular parallax of the stimuli i.e. crosstalk perception will increase with increase in contrast or increase in binocular parallax. It is commonly believed that ghosting plays the most critical role in determining the image quality. Wilcox and Stewart [31] observed that over 75% of the observers in their experiments reported ghosting to be the key feature that deteriorated the image quality. 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. Two objects 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 diplopic state. Crosstalk in still images is perceived to a larger extant 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 a lot of advices for the acceptable and unacceptable crosstalk for the viewers. Woods[32] points 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%.
- 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.

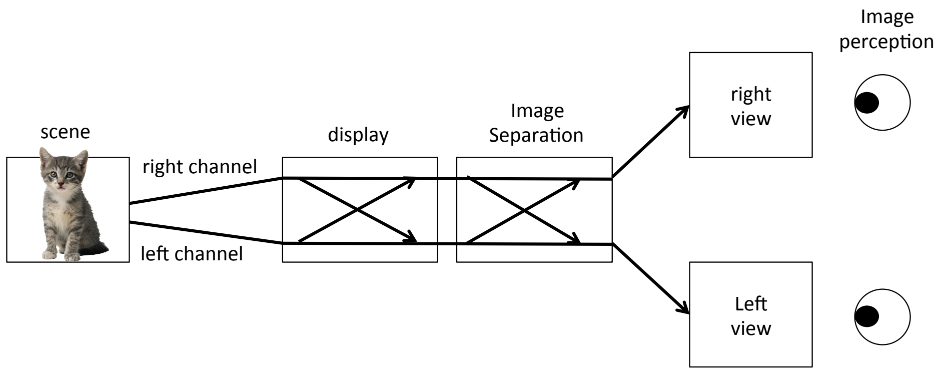


FIGURE 2.4: A flow diagram showing the process of stereo image perception starting from stereo scene capture. The crosstalk can be induced by the display (light generation) and by the image/view separation mechanism(3D glasses or autostereoscopic parallax barrier).

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 researcher 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 more complex scenes. Which is why we performed some experiments of our own to verify and expand the current knowledge in this area. This is also one of the main contributions of this thesis.

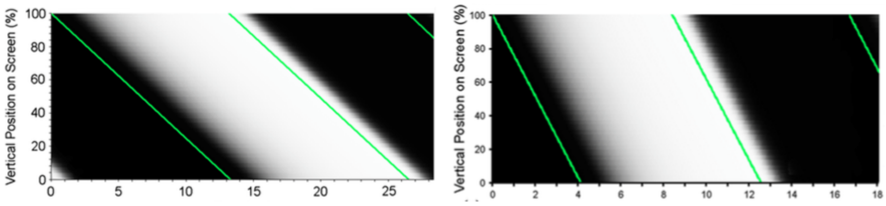
2.3 Stereoscopic/Automultiscopic Screens and its cross-talk

In this section we will review some of the stereo technologies and their associated crosstalk. The basic setup for a stereoscopic displays involves a display screen that has typically higher (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 has been separately delivered to the eyes). Various Stereo display technologies are available in the market such as CRT, DLP, Plasma, PDP and LCD screen. 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 separator.

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 florescent 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 flows through these individual filters which is regulated by the LCs. It should be noted that even at its best i.e 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/time. This response time is higher if the delta between the rotation is smaller and vice versa[32]. 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 is addressed individually from top to bottom.



(A) Time(ms) (B) Time(ms)

FIGURE 2.5: (a) Response of a conventional LCD display (in time domain). (b) Response of a high refresh rate LCD display [32].

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 images is being changed from complete black to complete white. It can be seen from 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 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 [32]. 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 grater leakage of light in the border areas of the shutters as compared to the center (provided the position of the eye is in the center of the shutters). It is important that the shutters are opened and closed with 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 that will result in 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

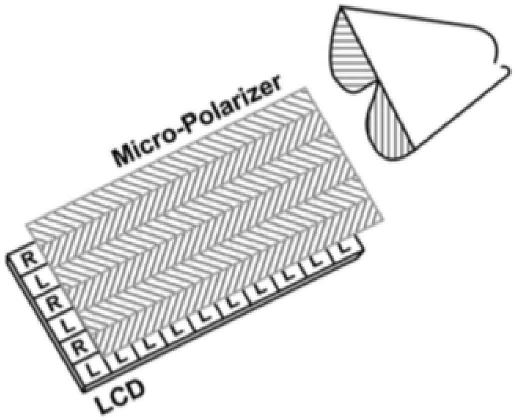
Polarization is a property dealing with the controlled oscillation of light (and other waves). Typically light waves oscillate in a way that their phase shifts over time are unpredictable. Passing the light through special polarizing filters forces them to oscillate in a controlled manner[28]. Light can be polarized in a linear manner where the light waves are forced to oscillate in a plane or is a circular manner i.e. 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 or 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 cheap polarized glasses that has 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, as with every piece of technology, 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). [32]. 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 up till now a material that perfectly preserves the polarization for all wavelengths has not been discovered. Lastly, it is hard to match the 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 tiled hence causing crosstalk [7]. Circular polarization however is less prone to this mismatch as compared to linear polarization hence it is more commonly used in 3D cinemas. In any case, filing 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 hence causing crosstalk.

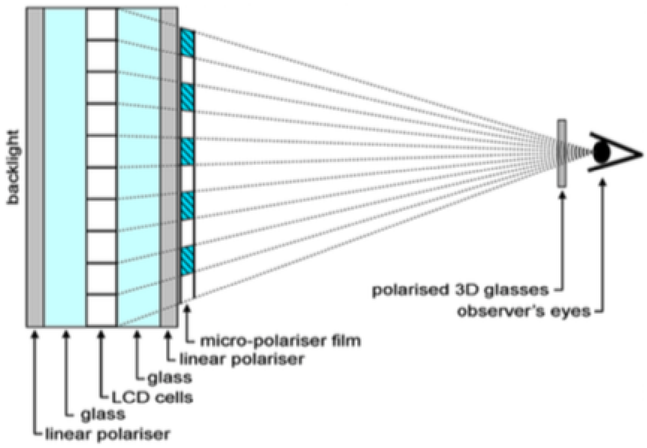
In summary, the important factors to consider when dealing with the crosstalk in a typical 3D cinema setup are as follows:

- 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.



(A)



(B)

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 [32] (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 that 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 an 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 vertical spatial resolution of the the displays as the light from half of the row pixels will be blocked by each of the lens 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 [32]. 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 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 needs 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[33]. The number of adjacent pixels columns occluded from one view determines how many different views can the screen display. 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 [27]. 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

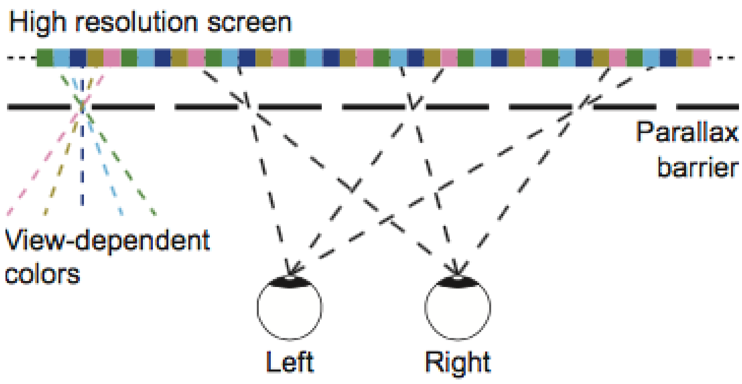
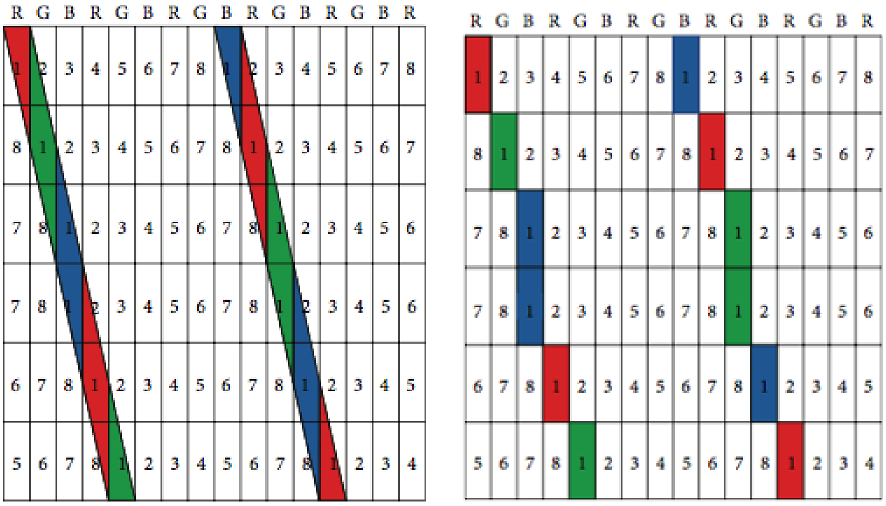


FIGURE 2.7: A five-view automultiscopic screen using a parallax barrier. The actual pixels on the LCD panel are called sub-pixels whereas the gaps in the parallax barrier can be thought of as a single view dependent pixel. Each of the observer's eyes can only see a specific set of sub-pixels through a view dependent pixel[33].

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 swear 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 [33]. 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 exhibits light leakage (crosstalk) between adjacent views. One reason for this is that the lenticules of a lenticular sheet are not perfect and does 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 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 later 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. I.e. it helps smoothen the view transition while switching the position of the viewer. Hence ideally it



(A) (B)

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 [24]

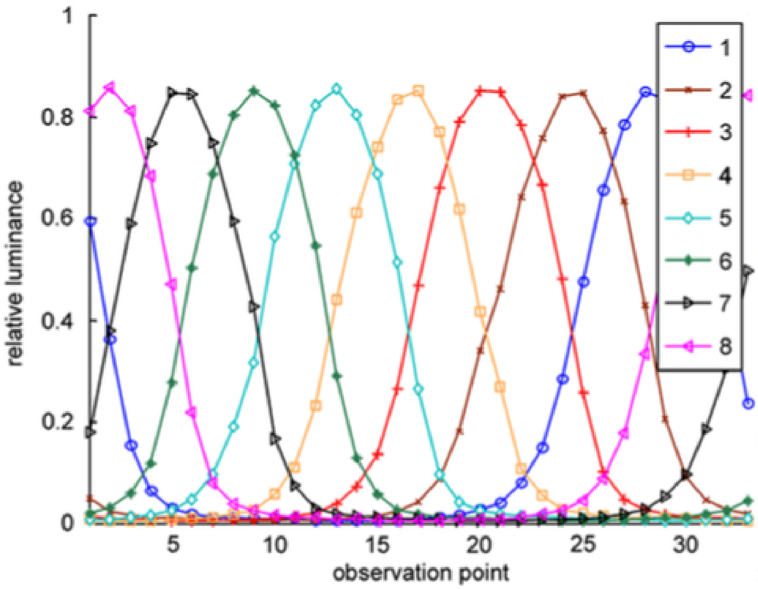


FIGURE 2.9: The experimentally observed luminance distribution of all views across different view points. It can be seen that some portion of the adjacent views is leaked even at the sweet spots. The light leakage increases as the observation point is moved away from the sweet spot [32].

would be desired if there is 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 believed that crosstalk in stereoscopic or automultiscopic screen results in reduced visual comfort, reduced disparity range, reduced contrast and most importantly, reduction of perceived depth. However, the literature is not thorough about how the crosstalk effects the perceived depth. Wilcox et al [31] performed a large scale experiment in which 77 observers were shown a small stereo footage in two commercial large-formate 3D theaters. The typical stereo degradations in theaters that were considered were crosstalk (ghosting), brightness, and the other qualities of 3D. 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 the perceived depth.

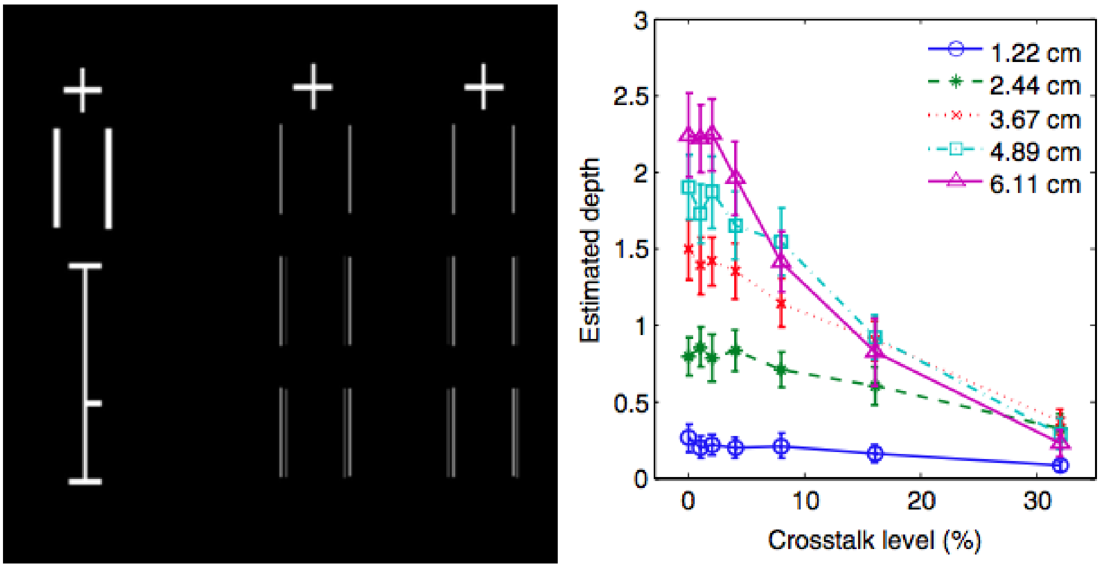
3.1 Effects of crosstalk on perceived depth

K.C.Huang et al [9] performed intensive experiments in order to obtain a threshold for the system crosstalk that on average will not mitigate the effects of the depth from disparity. Since the viewer crosstalk is dependent upon the system crosstalk and the local contrast of the desperate object 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]. Which means that compared to areas of the image with high local contrast, the HVS is more tolerable 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 (similar to ours) 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 desperate object. Rohaly et al [16] determined the effects that contrast exhibits on the stereoscopically perceived depth. They concluded that decrease in contrast made the objects (crossed and uncrossed) appear to be farther from the viewers. Which means that the crossed objects with decreased contrast appeared to be moving towards the horoptor 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 Note to self: Mention this in the discussion of the results. It was also argued that the contrast exerts its effect before or at the extraction of depth.

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 [20], 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. 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 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 swear. In 2012 [21], 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 a uniformly luminant objects on a black background where there is no other depth cue present. This is usually not the case with the complex scene 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 also observed that the perceived depth degraded with disparity even



(A) (B)

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

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 determined which we think is also necessary.

Tsirlin et al [19] again performed similar experiments but this time with complex natural scenes. The subjects were shown a complex scenes with the objects of interest marked by arrows. Later, 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 it does not tell us anything about how the HVS will respond to the geometry of the stimulus. Moreover, the users rating can be deviated from each other.

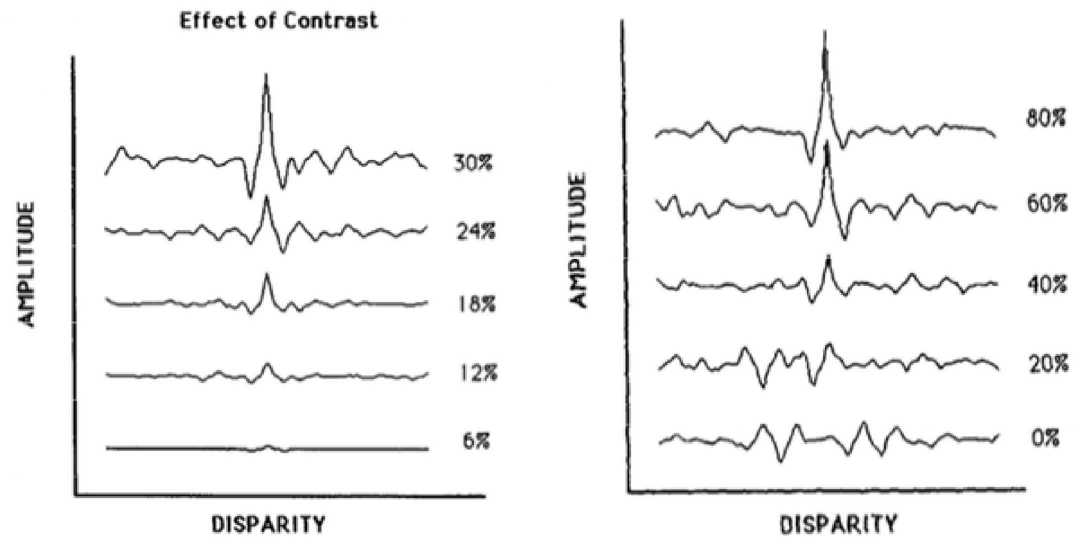
3.2 Depth resolution mechanisms in HVS

It is clear that crosstalk in stereoscopic displays reduce the perceived depth proportional to the level of crosstalk and the disparity of an object. However, in order to understand why the perceived depth is degraded, we had to look into some 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 IOC) profiles at different contrasts has been plotted. The peaks in these profiles represents 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. 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 [6] explained the limits of stereopsis and modeled them by using windowed interocular cross-correlation. The neighborhood of each pixel in one stereo images (say left) 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 patched was computed. The disparity was obtained at the displacement value where the maximum correlation for the patches was obtained. The 2D cross-correlation formula used was as seen in equation 3.1.

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



(A) (B)

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

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 an HVS model that can simulate not only the correct disparity estimation but also the degraded disparity when crosstalk is added to the images. note to self: Mention the disparity gradient as explained in this paper in the discussion.

3.3 Reduction of crosstalk

As mentioned in the previous sections, crosstalk severely hinders the perceived depth and reduces the overall visual quality and 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 have some amount of crosstalk present in them. Fortunately it is possible to calculate the amount of leakage between views at any position for any display. 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 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 other perspective view before displaying. 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. Which 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 hence 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 as well as maintaining the overall image quality.

3.3.1 Stereoscopic screens

Konard et al [11] 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. Limpscomb and Wooten [14] took the specially 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 above mentioned techniques assumes static scenes and hence ignore the temporal aspects. In a time sequential displays, crosstalk is induced not only by the view image from the opposite aspect (i.e. left and right view) but also from the views in the previous frame. Smit et al [17] took this phenomenon into consideration as well. However, as the problem with every subtractive approach, this technique also suffered from the over subtraction. Sohn and Jung [18] proposed a technique 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

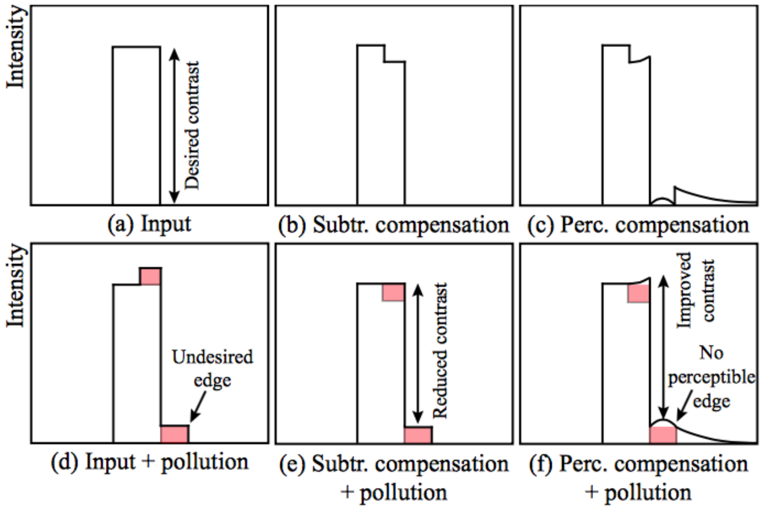


FIGURE 3.3: (a) Input image. (b) Subtractive compensation. (c) CSF weighed compensation. (d-f) Observed images [22].

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 step. 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, Baar and Poulakos et at [22] proposed a crosstalk reduction minimization that utilized the contrast sensitivity function CSF to steer the minimization in such a way that only perceptually important areas in a scene are corrected.

$$\underset{x}{\operatorname{argmin}} ||\lambda_l \, . \, r||^2, \quad s.t. \quad 0 \leq x \leq 1. \tag{3.2}$$

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

where λ_l is a diagonal matrix of the CSF spectral coefficients, x and $\bar{x}$ denotes the original and the compensated images respectively and $\phi()$ denotes the non-linear function representing the crosstalk. The constraints on the optimization are necessary so that the minimized image values does 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 presence of another signal with similar pattern. Baar and Poulakos further improved the optimization by incorporating into equation 3.2 an additional weighing with visual masking operator.

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

Here λ_n is a visual masking binary operator that will help ignoring the optimization for non-perceivable areas of the ghosted image due to visual masking. This idea can also be used for optimization of light fields in automultiscopic displays.

3.3.2 Automultiscopic screens

As discussed in chapter 2, crosstalk in an automultiscopic screen is introduced by the neighboring views i.e. the horizontal neighbors of a pixel in case the lenticular sheet is not tilted as well as the vertical neighbors of the same 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. Due to this reason, eliminating crosstalk from automultiscopic displays is trickier than stereoscopic displays.

Barkowsky and Campisi [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 8 views, the perceived views can be modeled as

$$L_p = (E \, + \, A) \,.\, L_d \tag{3.5}$$

where L_p and L_d are the vectors of displayed and perceived images in luminance domain, A is the circulant Toeplitz matrix approximating the crosstalk coefficients and E is 8x8 identity matrix. A generic reduction of crosstalk was derived from equation 3.5 for each channel

$$L_{rd} = (E \, + \, A)^{-1} \,.\, L_{ri} \tag{3.6}$$

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) hence equation 3.6 can be modified in order to increase the average contrast of the image.

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

here 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 hence for each view, the value of beta was computed iteratively that 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 [10] 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 Weiner inverse filtering to preprocess the images in order to mitigate the crosstalk. Li [13] 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 and Hou [24] 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 the tilt of 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.

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

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

Even though previously some work has been done by Tsirlin [20][19][21] in order to understand these effects, we felt that the experiments were incomplete and needed some more thorough investigation.. The reason for that is that firstly Tsirlin's experiments were performed on monochromatic images where no other depth cue other than disparity was present. As mentioned in chapter 3, we conducted such a pilot experiment on ourselves (the author and one colleague) with a thin rectangular white bar on a black background using a similar Wheatstone setup and found that it took some getting used to to sense any depth of the stimulus even at high disparities. Even after spending some time on it, it was extremely hard and unreliable to guess the apparent distance of the stimulus from the screen in length units. This would mean that the naive test subjects would also have encountered the same problems. This observation is backed by the fact that the test subjects in [21] and [20] 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. Coupled with the fact that 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, it is likely that generated results could have been unreliable. The scenes of the natural world usually have at least some other monocular cues e.g. proper light shading, texture gradients and defocus blur etc. Moreover, 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.

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 correct location (or disparity) of the correct match for an object in the corresponding binocular retinal image. To the best of our knowledge, it is yet unclear whether this HVS confusion is elevated or reduced when the geometric complexity of the object (stimulus) is increased. Which is why we thought it would be important to understand the relation of crosstalk on observed depth on stimuli of different widths and geometric complexities. Current studies suggests that for a certain crosstalk level and considerably smaller width of a desperate object, the extent of observed depth degradation increases as the disparity increase. 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, then in such case the HVS should be able to find the proper match relatively easily because it is easy to distinguish between the ghost and the object 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 completely separates. Hence the graphs of figure 3.1b should be “U shaped”.

In stereoscopic displays, the arrangement of object-ghost pair is antisymmetric between the eyes maybe make a diagram for this from the notebook. This however is not the case with automultiscopic screens. 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 screen 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. Again, 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.

4.2 stimuli

Our stimuli consisted of an object (cylinder or a dragon) placed in front of a plane that had a light wooded texture. The reason for choosing a textured background was the HVS's efficiency in distinguishing between textures. This would make it easy for the observer to correctly guess the distance between the background and the foreground object. In order to make sure that our stimuli contains most of the basic monocular cues, we properly rendered them on computer using blender 2.73. The blender scene as shown in figure 4.1 consisted of a an object of interest placed at origin. The background plane with wooden texture was placed 3 meters (world units) 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.

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. This however would have changed the object or the texture size which could have been used as a 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 meaning the cylinder always appeared to be in front of the plane by some distance. In order to avoid resizing, images were rendered at the resolution of 768x432. With the smallest baseline, the observed distance between the cylinder and the plane was negligible where as 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 repository. The details of each of those objects can be found in table 4.1.

Table 4.1: Description of objects used as stimuli.

Object	Scene Dimensions (xyz)	Width on rendered image (pixels)
Cylinder(thin)	5cm x 2m x 5cm	9
Cylinder(medium)	10cm x 2m x 10cm	18
Cylinder(think)	45cm x 2m x 45cm	27
Cylinder(thickest)	80cm x 2m x 80cm	36
Dragon	0.31m x 0.22m x 14.17cm	181(max)

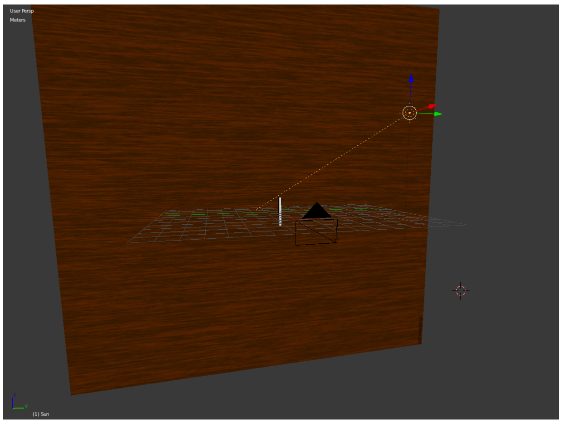
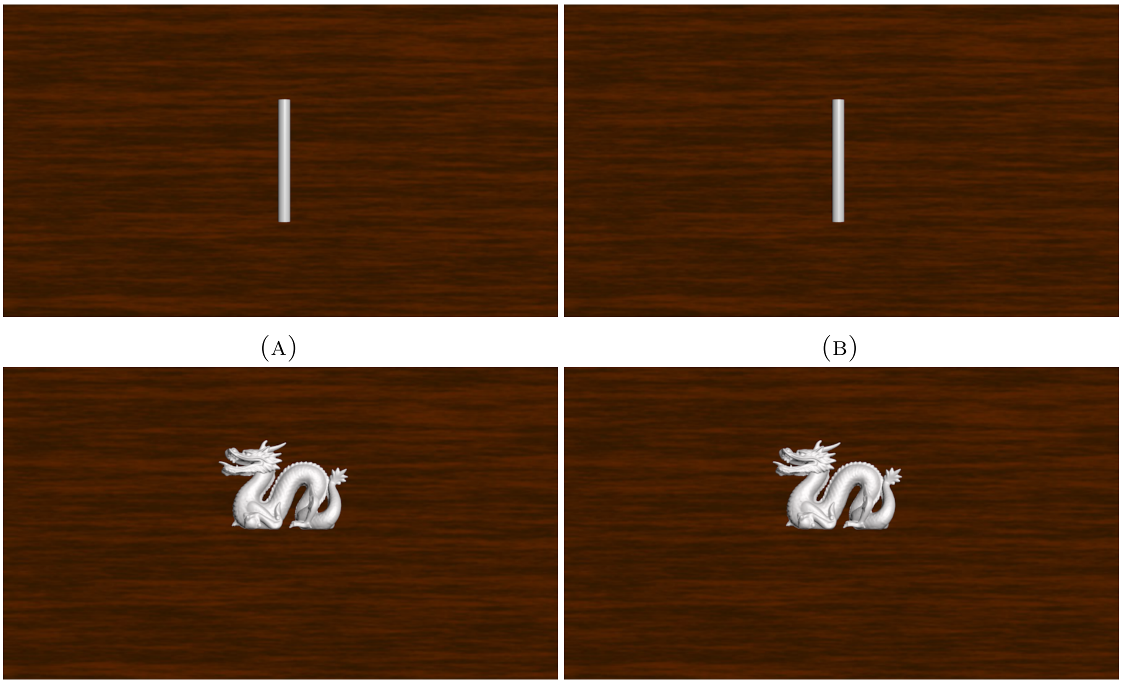


FIGURE 4.1: Typical arrangement of the scene rendered with blender.

4.2.1 Stereo and Automultiscopic stimuli

For the stereo experiments, captured the scene with the camera baseline starting from 0.002 meters till 0.12 meters with the step of 0.002 meters. This gave us sixty sets of stereo images where the theoretical depth of the cylinder ranged from 0.13 cm to 3.45 cm mention the degrees as wellin front of the plane of focus. Figure 4.2 shows stereo images of one of the cylinder and the dragon. The step size was chosen small enough so that the transition from the minimum depth to the maximum for an object would seem continues to the observer.



(A) (B)

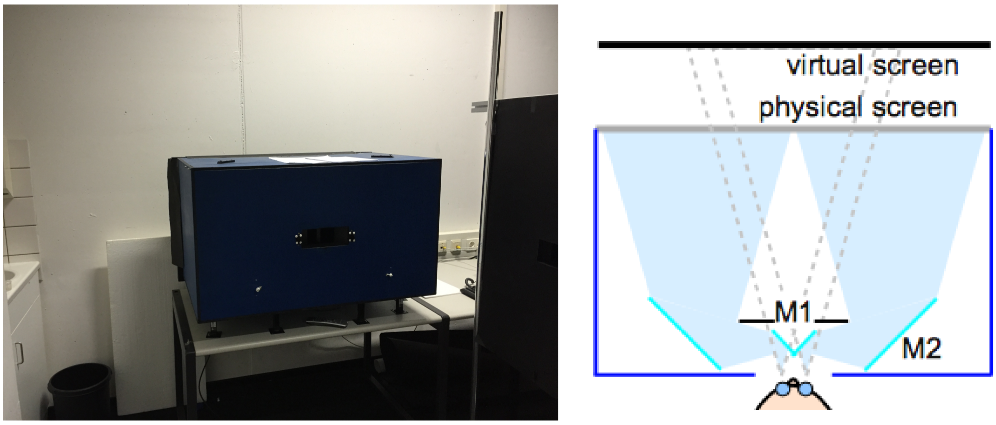
(C) (D)

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.

The images for automultiscopic experiments were rendered similarly. The only difference being that the range of baseline started from 0 until 0.24 meters. The reason for this is explained in the next section.

4.3 Apparatus and simulation procedure

The experiments were conducted on a commercial 47" SIMS2 HDR47E LCD display with a resolution of 1920x1080. The left and right half of the screen each with a resolution of 960x1080 were gated to the left and right eye using a custom built Wheatstone setup [30] that consisted of two mirrors M1 and M2 set at 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. The display even though build 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 the traditional LCD display hence we chose to use this mode.



(A) (B)

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 [23]

Since on an LCD panel, all the rgb channels equally and additively contribute to the crosstalk, we simulated the ghosted images simple 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, the automultiscopic case was a little tricky. First of all we simulated the view intensity profiles in figure 2.9 using Gaussians with $\sigma = 2$ and μs located at the sweet spots of the automultiscopic displays (figure 4.4). To match the profiles in figure 2.9, the gaussians were upscaled such that the value 0.8 was attained at the mean. Using our simulated profiles, we found that at any

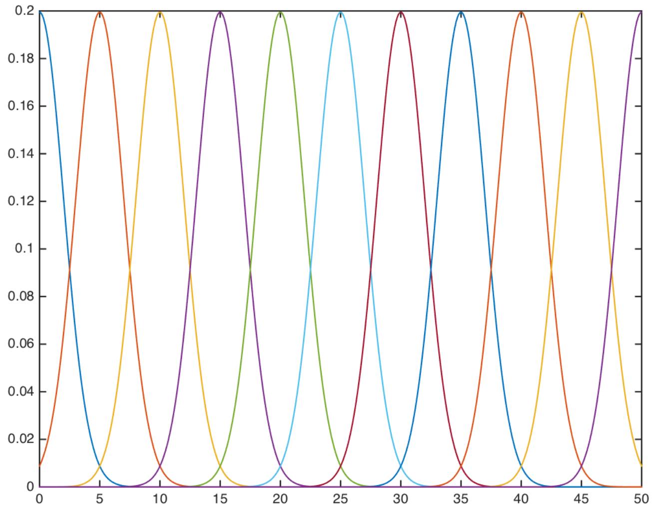


FIGURE 4.4: Luminance intensity profiles for an 11 view automultiscopic display simulated with Gaussians of $\sigma = 2$ centered at the sweet spots. X-axis denote the viewing position.

sweet spot viewing position, the amount of light leaked from views other then the two adjacent views was negligible. Hence in our simulation we only considered light leakage from the adjacent views. 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

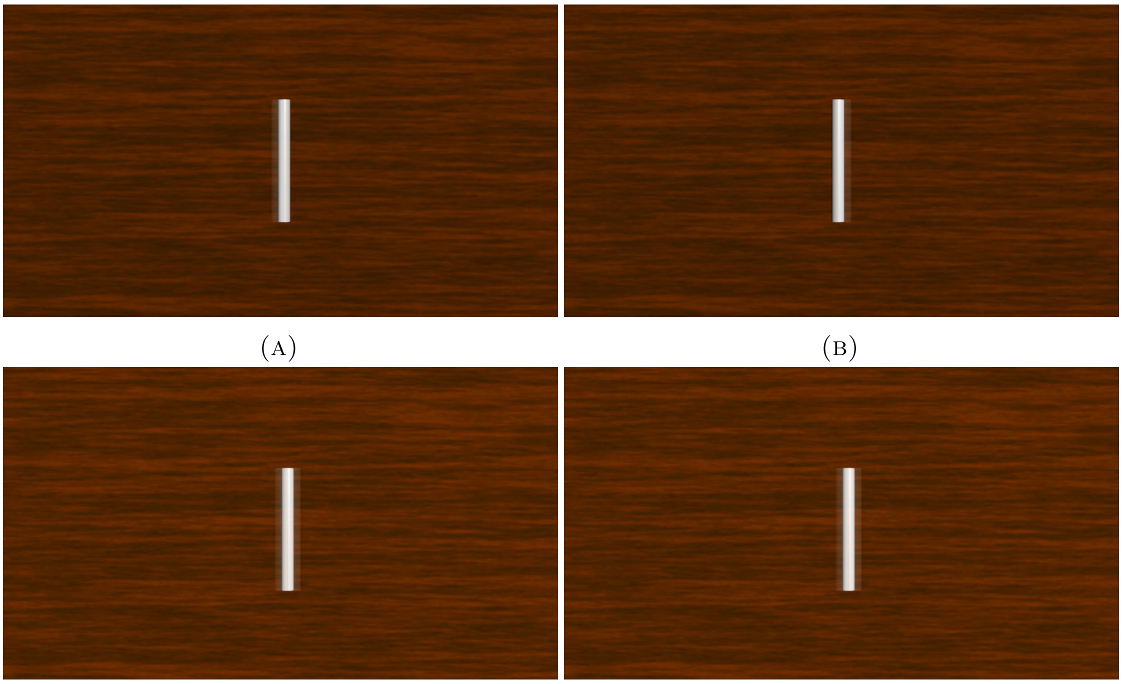
$$l(y, x) \simeq a_y(x) \,.\, f(y) \tag{4.1}$$

Where $a_y(x)$ is the Gaussian(corresponding to location y) value at position 'x' and $f(y)$ is the light field image at position 'y'. And the observe image "g" at any viewing position 'x' can be mathematically modeled as

$$g(x) \simeq ... + l(x - 2\delta_x, x) + l(x - \delta_x, x) + l(x, x) + l(x + 2\delta_x, x) + l(x + \delta_x, x) + ... \tag{4.2}$$

This would mean that the amount of ghost separation at any disparity of the object would be constant on either side. In order to keep the nature of the automultiscopic experiments similar to that of the stereo (where the ghost separation increase as the disparity increase), we changed the sampling rate δ_x with respect to the disparity of the object. i.e. For a disparity 'd', the the the observed left and right images can be written as

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



(A) (B) (C) (D)

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)

4.4 subjects

4.5 Results

4.6 Discussion

4.6.1 Stereo

4.6.2 Automultiscopic

Chapter 5

Model

5.1 Hypothesis and their results

5.1.1 Simple correlation model

5.1.2 Results

5.1.3 Windowed correlation model

5.1.4 Results

5.1.5 Simple correlation model

5.1.6 Results

5.1.7 PDF model

5.1.8 Results

5.1.9 Mathematical model

5.1.10 Results

5.2 Discussion

Chapter 6

Applications

6.1 Depth prediction application

6.2 Crosstalk mitigation

6.2.1 Proposed optimizations

6.2.2 Unsharp masking in view domain

6.2.3 Iterative subtraction

Chapter 7

Conclusion

7.1 Summary

7.2 Future Work

7.3 Open Questions

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] Filippini, H. R., and Banks, M. S. Limits of stereopsis explained by local cross-correlation. *Journal of Vision 9*, 1 (2009), 8.

[7] 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.

[8] Howard, I. *Binocular vision and stereopsis.* Oxford University Press, New York, 1995.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Li, D., Zang, D., Qiao, X., Wang, L., and Zhang, M. 3d synthesis and crosstalk reduction for lenticular autostereoscopic displays.

[14] 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.

[15] 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.

[16] Rohaly, A. M., and Wilson, H. R. The effects of contrast on perceived depth and depth discrimination. *Vision Research 39*, 1 (1999), 9–18.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] van Baar, J., Poulakos, S., Jarosz, W., Nowrouzezahrai, 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.

[23] 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.

[24] Wang, X., and Hou, C. Improved crosstalk reduction on multiview 3d display by using bils algorithm. *Journal of Applied Mathematics 2014* (2014).

[25] Wikipedia. Depth perception — wikipedia, the free encyclopedia, 2015. [Online; accessed 20-August-2015].

[26] Wikipedia. Horopter — wikipedia, the free encyclopedia, 2015. [Online; accessed 21-August-2015].

[27] Wikipedia. Light field — wikipedia, the free encyclopedia, 2015. [Online; accessed 7-October-2015].

[28] Wikipedia. Polarized 3d system — wikipedia, the free encyclopedia, 2015. [Online; accessed 8-September-2015].

[29] Wikipedia. Stereopsis — wikipedia, the free encyclopedia. [Online; accessed 21-August-2015].

[30] Wikipedia. Stereoscope — wikipedia, the free encyclopedia, 2015. [Online; accessed 23-August-2015].

[31] 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.

[32] Woods, A. J. Crosstalk in stereoscopic displays: a review. *Journal of Electronic Imaging 21*, 4 (2012), 040902–040902.

[33] 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