STEREOSCOPIC 3D VIEWING SYSTEM USING A SINGLE SENSOR CAMERA

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Science in the

Graduate School of The Ohio State University

By

Deepak Prakash, B.E.

* * * * *

The Ohio State University
2007

Master's Examination Committee:

Dr. Yuan F. Zheng, Adviser

Dr. Ashok Krishnamurthy

Approved by

____

Adviser

Graduate Program in Electrical and

Computer Engineering

ABSTRACT

The usual practice for stereoscopic imagery has been to capture a scene from two different viewpoints from two cameras and careful consideration needs to be given when setting up such a system. Several factors like the lens focal length, the field of view of the lens and the time of the video capture have to be precisely aligned or synchronized for both the cameras. Practically this is a cumbersome task and in this thesis, we investigate the use of a single video camera on a moving platform that eliminates the above mentioned issues and still generates the same depth as perceived by the viewer. This is achieved by presenting different video frames to the two eyes with a time-slip between them.

Importance has been given to the comfort with which a viewer can perceive the 3D video. Out of the several factors that affect the viewer comfort like field of view, focal length and scene brightness, specific consideration has been given to the camera separation distance that plays the major role. Previous works show that the separation distance is a function of the focal lengths of the capturing and viewing cameras and the distances of the nearest and farthest points in the scene.

Applying this to our single camera system, we obtain an optimum frame delay that ensures a good comfort for the viewer while perceiving the depth from the videos. Our system also enables the user to adjust the frame difference in case he feels uncomfortable with the optimum delay or if he feels he wants to see the scene in more depth.

ACKNOWLEDGMENTS

I would like to express my gratitude to a number of people, without whom my work would not have been possible. First, I would like to thank my advisor, Dr. Yuan Zheng, who offered encouragement, motivation, and valuable career advice. Also, I would like to thank my sponsor for the financial support received by me through the duration of my project. I would be remiss in not thanking Dr. Ashok Krishnamurthy for agreeing to be in my examination committee and helping me prepare a good report of my work with his valuable suggestions.

Lastly, I would like to thank my family and friends for their constant support and encouragement. I will never be able to thank you all enough for instilling in me the confidence, desire, and toughness needed to reach my goals in life.

VITA

February 23, 1984…………………Born – Tiruchirapalli, India

May, 2005…………………………B.E., Anna University

2005 to present……………………Graduate Research and Teaching Assistant,
The Ohio State University

FIELDS OF STUDY

Major Field: Electrical and Computer Engineering

Studies in:

Signal Processing: Prof. Randolph Moses

Image Processing: Prof. Alex Martinez

Computer Vision: Prof. Jim Davis

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Stereoscopy or 3D Imaging is any technique capable of recording 3-dimensional imagery in order to create the illusion of depth. Stereoscopy is one of the growing research areas in computer vision and image processing, the study of which would be helpful in a large number of applications. It has been proved through various studies that stereoscopic images and videos provide more information and are much more helpful to human beings than their 2D counterparts. The benefits that stereoscopic displays can provide are widely understood [1],[2]–[4] and include: depth perception relative to the display surface; spatial localization, allowing concentration on different depth planes; perception of structure in visually complex scenes; improved perception of surface curvature; improved motion judgement; improved perception of surface material type. These benefits give stereoscopic displays improved representation capabilities that allow the user a better understanding or appreciation of the visual information presented [5]–[7].

Depth perception is obtained after two different perspective images of the scene acquired by two cameras are presented separately to the two eyes. Two cameras which capture the scene from different angles have their own disadvantages. Factors like Sensor alignment and time synchronization for the scene capture are difficult to obtain practically and have to be very precise in order to get the depth effect. So a single camera was used to capture the scene from which stereoscopic effect will be generated.

Though this method eliminated the alignment and synchronization issues of the two camera method, it has its own issues. Two different frames have to be presented to each eye and careful consideration needs to be made as of how and which frames need to be presented to the left and right eyes respectively.

1.1 Organization of this Thesis

The rest of this thesis is organized as follows.

Chapter 2 deals with explaining the physiological and psychological functions which are responsible for the depth perception in the human visual system. We start with explaining the depth cues, their influence on the observer and finally conclude with a theory on object distance calculation based on the retinal disparity cue.

Chapter 3 deals with the factors affecting the depth perception in a stereoscopic image pair. Various factors affecting the depth are discussed and specific importance is given to camera separation distance. A formula to calculate the camera separation distance for a given scene is derived which is used in the later chapters.

Chapter 4 extends the previous discussion to a stereoscopic video where other factors like frame rate and velocity of the platform also come into our picture.

Chapter 5 discusses the application of the formulae and the study in the previous chapters in an aerial platform scenario example. It also discusses the depth factor consideration that may vary depending on the observer viewing the video. The effects of increasing or decreasing the Field of View (FOV) are also studied.

Chapter 6 explains the Graphical User Interface (GUI) software developed as a result of this study. A sample video used for demo is shown and the various features developed in the application are discussed.

Finally, Chapter 7 gives a conclusion of this research work and also discusses possible future work on this study.

CHAPTER 2

HUMAN EYE DEPTH PERCEPTION

In order to understand the depth perception involved in viewing stereoscopic systems, we must first understand how the human visual system helps us in perceiving depth in real world. Sir Charles Wheatstone, a British scientist, discovered the depth sense of human beings in 1838. Though there were some earlier works on stereopsis before 1838, it was Sir Wheatstone's work combined with the discovery of photography that made stereoscopy hugely popular.

A lot of psychophysical studies have been done to determine the factors that aid us in estimating the depth of the objects in the scene. There are two main theories that describe the human depth perception in a broad sense. The first and the more popular one is the Depth Cue Theory which states that using various sources of depth information i.e. depth cues we are able to obtain the 3D information of the scene being observed [8]. The second one, namely the Ecological Theory states that depth perception is not only dependent on the depth cues but more importantly on the context of the environment as well [9].

In this work, we rely on the Depth Cue Theory and hence the stereoscopic viewing system designed does not depend on the context of the visual scene. The depth cue factors can be broadly classified into three types: monocular, binocular and oculomotor cues [10].

2.1 Monocular Cues

Monocular cues are those depth cues perceived with one eye. Most of these cues were first enumerated by the Renaissance painters. More information on these cues can be found in the book Sight and Mind [11]. The next few pages describe briefly about these different types of cues. There are six main types of monocular cues:

2.1.1 Relative Size

Retinal image size allows us to judge distance based on our past experience and familiarity with similar objects. A retinal image of a small car is interpreted as a distant car (Figure 2.1).

Figure 2.1: Relative Size (source: Internet).

This is due to the fact that the smaller the size of the retinal image, the greater will be the distance of the object from the observer. Though the retinal image size gets smaller we do not think the object to be shrinking rather we think that its distance from us is increasing because we have a definite idea of its constant size. This perception of a constant size of an object is called size constancy.

2.1.2 Interposition

Interposition cues appear when one object overlaps another. An object which occludes another object is perceived to be closer than the other object. Previous knowledge of the objects in the scene may be helpful in obtaining this information. In Figure 2.2, the green circle (one on the right) is perceived as closer to the observer since it overlaps the blue circle (one on the left).

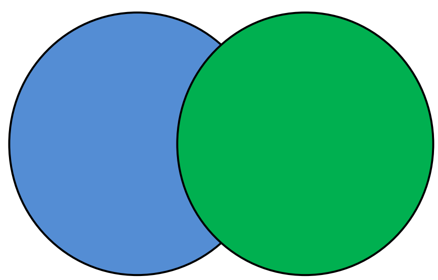


Figure 2.2: Interposition.

2.1.3 Linear Perspective

When a known object subtends a smaller angle, it is interpreted as being further away (relative size). In an image, parallel lines like roads and railway tracks appear to converge at some distant point (Figure 2.3). The locus of such kind of distant points where parallel lines seem to converge is called the horizon. Among the monocular cues, linear perspective is considered to be the most used depth cue in 3D computer generated imagery [12]. This cue was used even by the Renaissance painters several centuries ago.

Figure 2.3: Linear perspective (source: Internet).

2.1.4 Aerial Perspective

Relative color of objects gives us some clues to their distance. Due to the scattering of blue light in the atmosphere (Rayleigh's Effect), this effect can be used as a distance cue as well. Near objects appear less blue whereas distant objects appear bluer.

2.1.5 Light And Shade

Highlights and shadows can provide information about an object's dimensions and depth. Because our visual system assumes the light comes from above, a totally different perception is obtained if the image is viewed upside down.

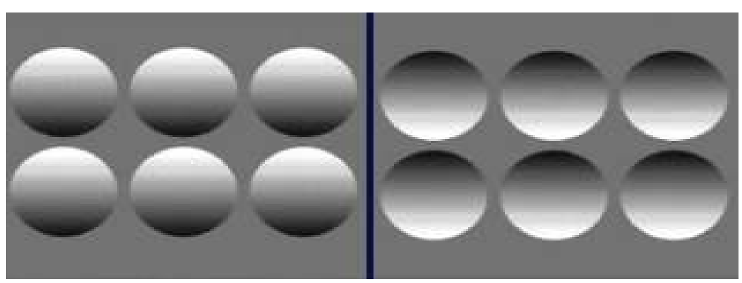


Figure 2.4: Highlights and shadows provide information about depth.

In Figure 2.4, one pattern of the shadows and the highlights on the left image make it appear convex whereas another pattern of them make the right image appear concave.

2.1.6 Monocular Movement Parallax

When our heads move from side to side, objects at different distances move at a different relative velocity. Closer objects move "against" the direction of head movement and farther objects move "with" the direction of head movement.

2.1.7 Texture Gradient

A surface is seen as coarse when seen from a close point as opposed to being perceived as smooth when seen from a farther point.

Thus these are the main types of monocular cues that help us in perceiving depth. In real world scenario, usually depth perception is achieved through a combination of two or more of these cues. There may also be cues like brightness (the brighter an object is, the closer it is to the observer), shading and shadowing. All these monocular cues are also referred to also psychological cues.

2.2 Oculomotor Cues

To understand the binocular stereoscopic perception, one has to understand ‘accommodation’ and ‘convergence’ of the human eyes. Though convergence and accommodation are classified under oculomotor cues, they are closely related to stereopsis or binocular vision. Accommodation and convergence are discussed below.

2.2.1 Convergence

In order to look at a point, both our eyes have to converge to that point. This rotation of the eyes in order to concentrate on a single point in the scene space is called convergence shown in Figure 2.5. This point of concentration is called the fixation point. In order to focus at another point in the scene, the eyes may have to converge or diverge depending upon the location of the new fixation point with respect to the old point in the scene. The visual system may use this information to estimate the distance of the point from the observer or in other words the depth of that point in the scene.

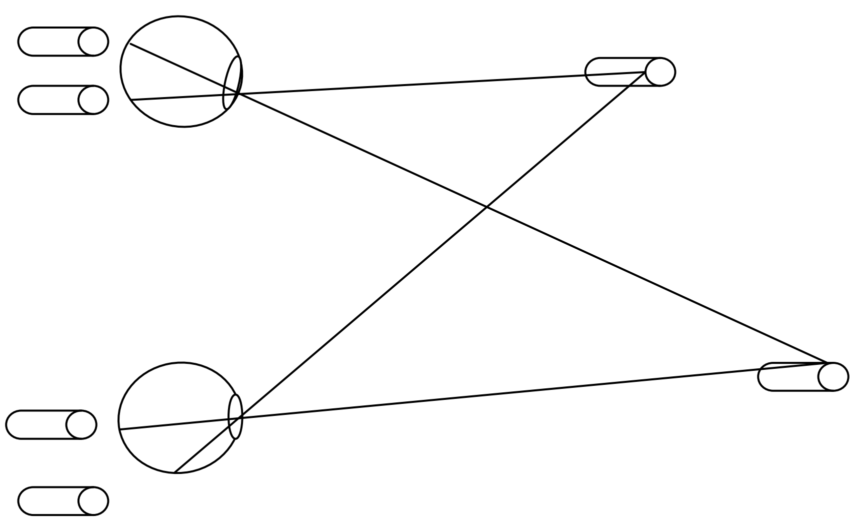


Figure 2.5: Illustration of convergence of human eyes.

2.2.2 Accommodation

Also to focus at the fixation point, apart from the convergence of both eyes the shape of the lens is also modified. The lens in each eye is either contracted or expanded by the eye muscles in order to bring the point of interest in sharp focus. This change in the shape of the lens is called accommodation. This is also a very important oculomotor cue used in estimating the depth of the scene.

2.3 Binocular Cues

Binocular means ‘two eyes’. Stereopsis or use of the binocular depth cue means the perception of depth of the scene using the discrepancy between the left and right eyes. ”The mind perceives an object of three dimensions by means of two dissimilar pictures projected by it on the two retinas” [13]. Binocular cues operate in a more complex way in depth perception than monocular cues. These cues are also called physiological cues and are more explicable with the anatomical function of the eyes.

2.3.1 Retinal Disparity

When the eyes converge on an object in space, it is seen as a single image and all other points in front of or behind it appears double. In Figure 2.6 below, the middle object is being focused and so the image of it appears in the center of both the retinas.

However the images of the closer and the farther object fall on different parts of the retinas and this difference in the location of each of the images in the two retinas is called retinal disparity.Disparity tells us the relative distance of that point with respect to the fixation point. The locus of all points that lie along with the fixation point and which appear to be at the same distance as the fixation point is called the horopter. Thus all the points on the horopter have zero disparity [14].

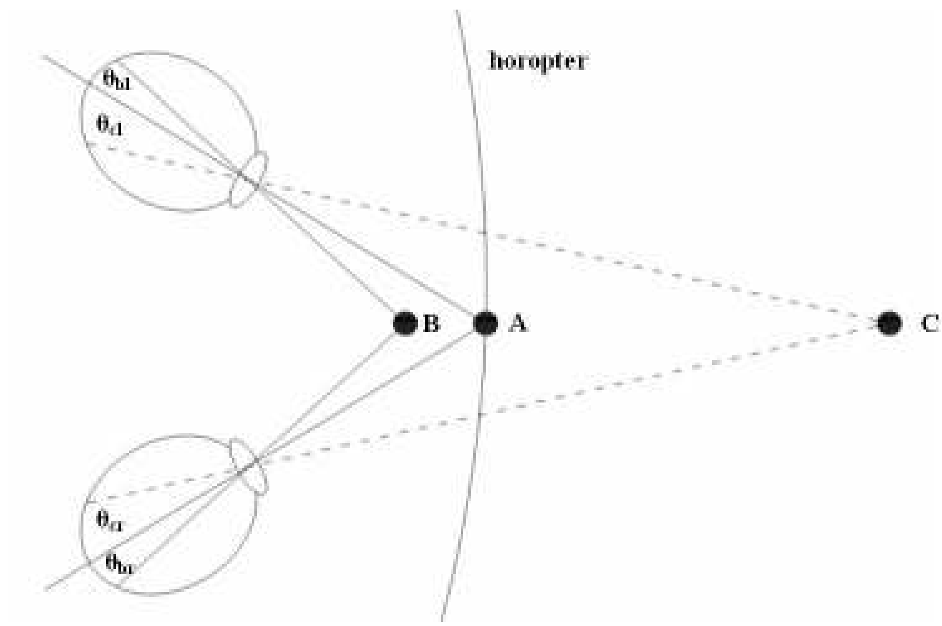


Figure 2.6: Fixation point and the horopter in the field of view of the human eyes.

The points which lie between the eye and the horopter produce negative disparities. Negative disparities are also called crossed disparities since the optical axes of the eyes have to cross in order to re-fixate this new point.

The points which lie beyond the horopter produce positive or uncrossed disparities in the retinas. Here the eyes diverge or uncross to focus on the new point.

2.4 Combination and Application of Depth Cues with Varying Distances

All the various depth cues we saw namely, monocular, oculomotor and binocular have various influences on the observer with respect to distance. As said earlier, depth perception is usually a combination of two or more of these cues in real world scenarios. In general, the more cues present in the scene, the better will be our depth perception and estimation. Some cues dominate the other cues in certain situations [15].

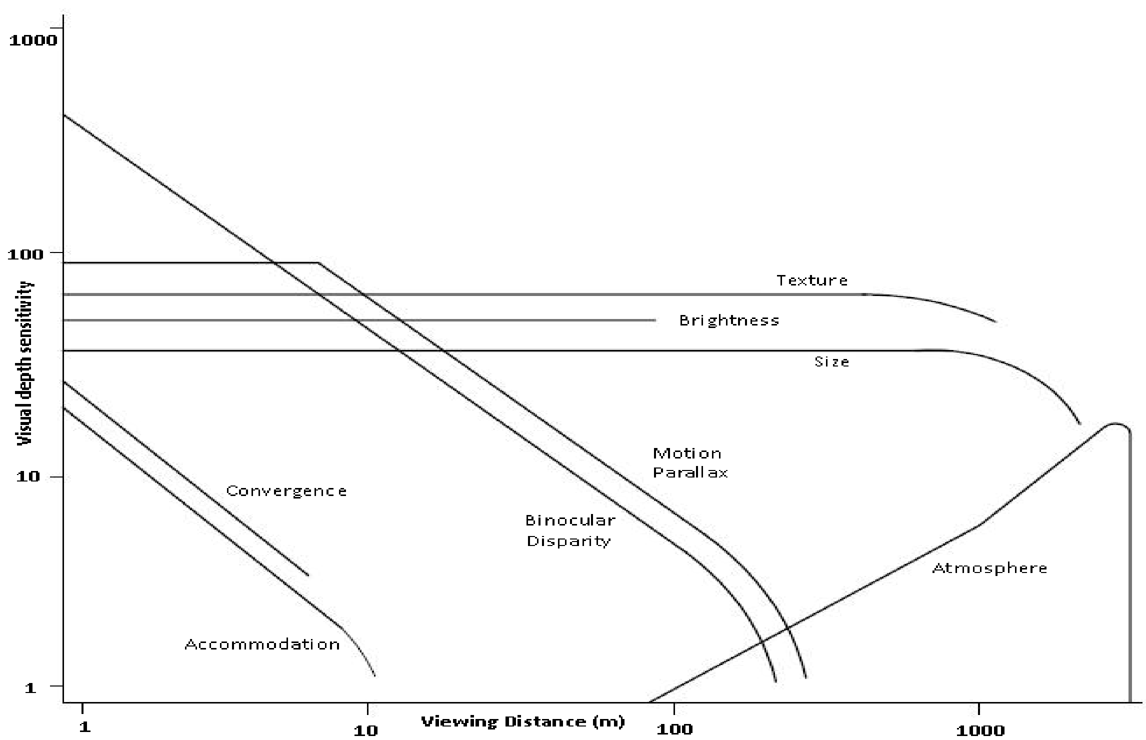


Figure 2.7: The effectiveness of depth cues as a function of distance. Adapted from [16].

For example, a person threading a needle primarily uses stereo cues to determine the location of the end of the thread and the eye of the needle, and usually brings the objects close to the eyes to increase the accuracy of stereo and oculomotor cues.

However, a submarine pilot is unlikely to use stereo or oculomotor cues to determine the distance to a far-off buoy, instead relying on multiple pictorial depth cues [16]. An important criterion for the dominance of one cue over another is the distance from the viewer to the objects of interest. In the above figure, Nagata has shown the influence of the different depth cues on the observer as a function of distance.

2.5 Retinal Disparity and Object Distance Calculation

From Figure 2.8 below, we can calculate the distance of the object from the observer. The convergence angle is the angle enclosed by the optical axes of the two eyes. Now, a relationship between the convergence angle and the distance can be derived.

At any moment in time, the eyes can converge only at a single point in the scene space (point F in our case). The other points in space namely P and Q will have a retinal disparity and can be represented as the difference in convergence angle between the fixation point F and the point P or Q.

Hence,

Retinal Disparity for point Q: $\delta_q = \alpha - \alpha_q$

where α and α_q are the convergence angles subtended by points F and Q respectively.

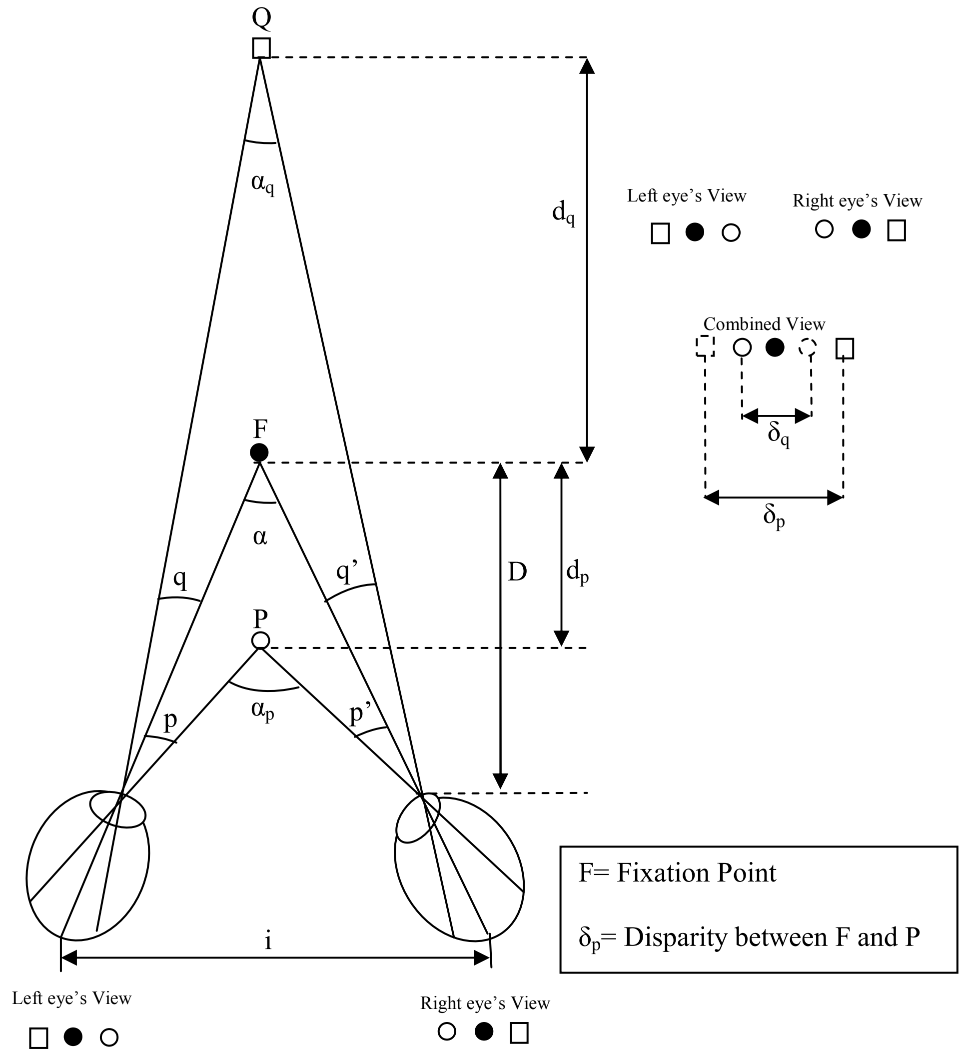


Figure 2.8: Convergence and Disparity for direct viewing. Adapted from [25].

In order to focus on the point F in the scene space, both the eyes converge to subtend an angle α at the center. The point P is closer to the observer and the point Q farther away from the observer compared to point F. The relationship between the distance of the fixation point (D) and the convergence angle (α) of the optical system is given as:

$$\alpha = 2 \times arc \tan \left(\frac{i}{2 \times D}\right)$$

where i is the inter-ocular distance.

For small angles of α,

$$\alpha = \frac{i}{D}$$

The difference in convergence angle between the fixation point F and that of a second point (say Q) is called retinal disparity. Thus in Figure 2.8 when we consider the small angle approximation, we have

$$F: \quad \alpha = \frac{i}{D}$$

$$Q: \ \alpha_q = \frac{i}{D + d_q}$$

Thus the retinal disparity is given as

$$\delta = \alpha - \ \alpha_q = \frac{i \times d_q}{D \times (D + d_q)}$$

This equation establishes geometrically that the disparity is related to the fixation point distance (D) and the relative distance d_q.

With the above derivation, Foley [17] argues that using the angle of convergence of the eyes and the distance between the eyes (separation of eyes for adults is approximately 65 mm), the human visual system calculates the absolute depth of the object. Then using the relative distance from the retinal disparity, the depth of other points in the scene is calculated. However, empirically, this idea has not been verified so far.

2.6 Panum's Area, Near Point and Far Point Distances

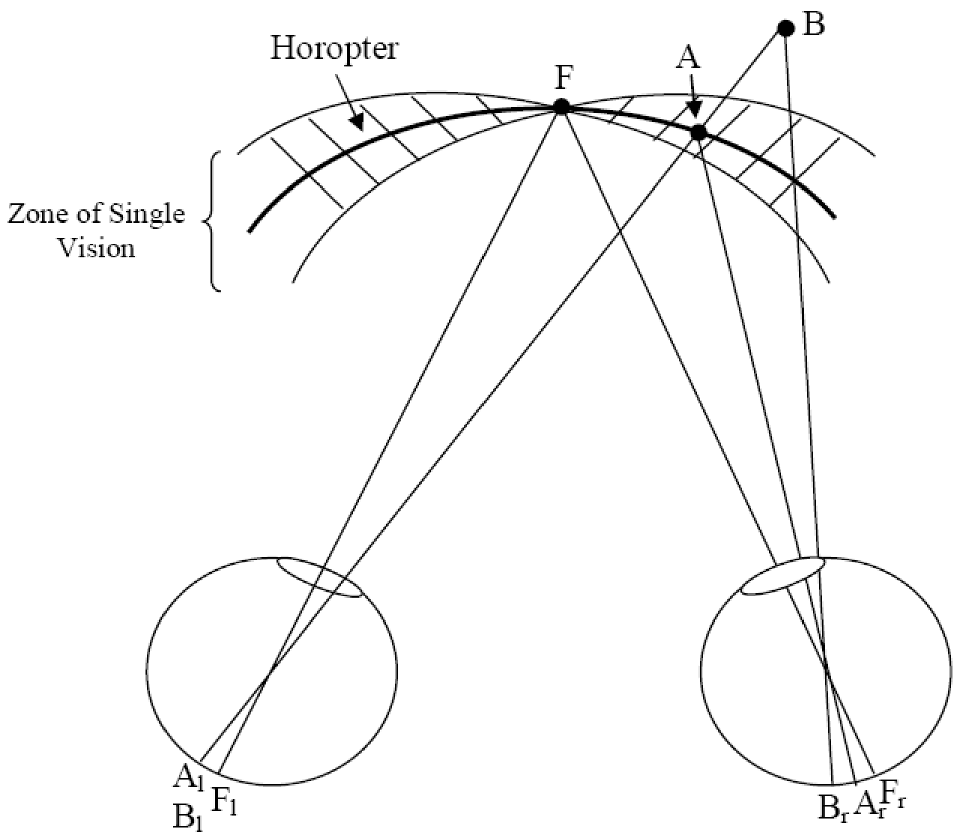


Figure 2.9: Panum's Area with the horopter and single vision region.

While fixating on an object in the scene, we can observe two phenomena. One, as the objects get farther away from the fixation point, they become increasingly blurred. Two, in the horopter region, there is a range of distances within which the brain can fuse the two retinal images into a single vision image. This range of distances in the area of single vision is called Panum's area.

For a defined field of view of the eyes, after fixating a point in a scene, the physical point which is nearest to the observer is called the Near Point. The physical point which is farthest from the observer in the same field of view is called Far Point. Hence the distance of the near point from the observer is called the Near Point Distance whereas the distance of the far point is called the Far Point Distance. These near point and far point distances in a scene play a very important role in the depth perceived of the scene and will be discussed in chapter 3.

CHAPTER 3

FACTORS AFFECTING DEPTH IN A STEREOSCOPIC IMAGE PAIR

When capturing a scene for creating stereoscopic images or video, several parameters affect the perceived stereoscopic depth of the observer. Some factors arise from the camera we use and the others depend on the scene space. Both these types of factors need to be considered to obtain a good stereoscopic image/video pair.

3.1 Factors Affecting Depth

Figure 3.1 shows some of the important factors used in a stereoscopic image capture. They are as follows:

i. The camera separation distance,
ii. The convergence distance of the camera lenses,
iii. The convergence angle of the camera lenses,
iv. The focal length of the lenses, and
v. The object distance from the camera.

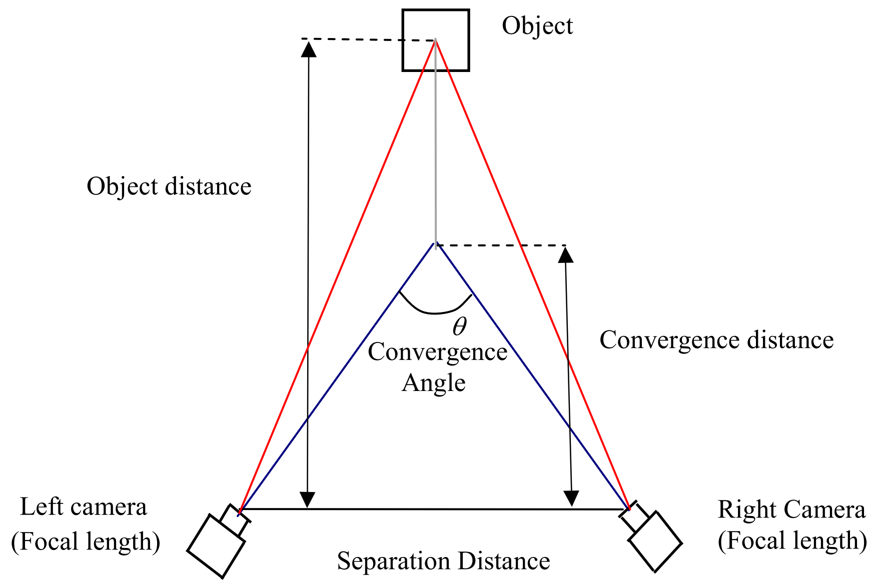


Figure 3.1: Factors affecting the capture of a Stereoscopic Image/Video Pair.

The convergence distance and convergence angle can also be considered as a single factor instead of two separate factors as they are dependent on each other. Studies on these factors have been done before and they are summarized in table 3.1.

It is understood from the previous studies [18] that when the focal length of both cameras is equally increased, the perceived depth is increased, the size of the objects is increased, but the field of the view is decreased; when the convergence distance is increased, the perceived depth is shifted closer to the eyes; when the separation between the two cameras is increased, the perceived depth is increased.

Parameter	Effect of Increase	Effect of Decrease
Camera Separation Distance	Increase in disparity values Increase in perceived depth Constant size of objects Constant field of view	Decrease in disparity values Decrease in perceived depth Constant size of objects Constant field of view
Camera Focal Length	Increase in disparity values Increase in perceived depth Increase in size of objects Decrease in field of view	Decrease in disparity values Decrease in perceived depth Decrease in size of objects Increase in field of view
Convergence Distance	Decrease in disparity values Shift in perceived depth Constant size of objects Constant field of view	Increase in disparity values Shift in perceived depth Constant size of objects Constant field of view

Table 3.1: Effect of Changing the Stereoscopic Camera Parameters. Adapted from [18].

The other factors of our interest are, the distance of the nearest and farthest objects in the scene (in case of multiple objects in the scene which is usually always the case) and the maximum disparity able to be perceived by the human visual system.

The camera separation distance or the camera base distance is a very important factor in the stereo perception of a scene as it directly affects the retinal disparity of the images and hence the stereoscopic depth perceived. So now, we need to derive a relationship between these stereo parameters: camera base distance, near point distance, far point distance and the focal length of the lens.

The maximum permissible on-film deviation (in case of analog cameras) or the sensor deviation (in case of digital cameras) for the human eyes also needs to be considered. In this work, a digital camera was used and hence the on-sensor deviation term will be used henceforth. For now we will not consider the convergence distance and angle factors the reason of which will be seen in the next chapter.

3.2 Camera Base Distance Calculation

We set the camera in a scene and capture images to obtain a stereoscopic view of the scene. We take note of (or measure) the near and far point distances in the scene. Far point distance can be very large but not infinity and the near point distance can be as small as possible so that the camera lens can still maintain the image in focus. This derivation has been taken from [19] and has been presented again in order to clarify some steps and to give the reader a better picture of the stereo geometry.

The parameters used in the camera base distance calculation are as follows:

- a_n: the distance from the lens center to the near point object in the scene.
- a_f: the distance from the lens center to the far point in the scene.
- f: the focal length of the camera lens.
- a: the distance along the lens axis from the lens center to the plane of maximum sharpness in the scene.
- a': the distance along the lens axis from the lens center to the plane of maximum sharpness in the sensor gate (i.e.) the sensor plane.

For simplicity purposes, we will call (a) as sharp object distance and (a') as sharp image distance. The relationship between these two variables is given by a common Gaussian form of lens equation which is

$$\frac{1}{sharp\ object\ distance} + \frac{1}{sharp\ image\ distance} = \frac{1}{focal\ length}$$

$$\frac{1}{a} + \frac{1}{a'} = \frac{1}{f}$$

This is rewritten as,

$$\frac{1}{a'} = \frac{1}{f} - \frac{1}{a} \qquad --- (1)$$

We can note that as the value of (a) approaches the value of (f), the distance from the lens to the image sensor must become infinite to maintain sharp focus.

The other remaining variables in this geometry are:

- b_0: the lens separation or the distance between left and right lens centers.
- b_n: distance between the left and right images of the near point on the sensor.
- b_f: distance between the left and right images of the far point on the sensor.
- d: this is the on-sensor deviation, and is equal to the difference between b_f and b_n (i.e. $d = b_f - b_n$).

Figure 3.2 shows the general geometry of a stereo camera with two objects in the field of view- a near object and a far object with the camera's maximum focus at a point between them.

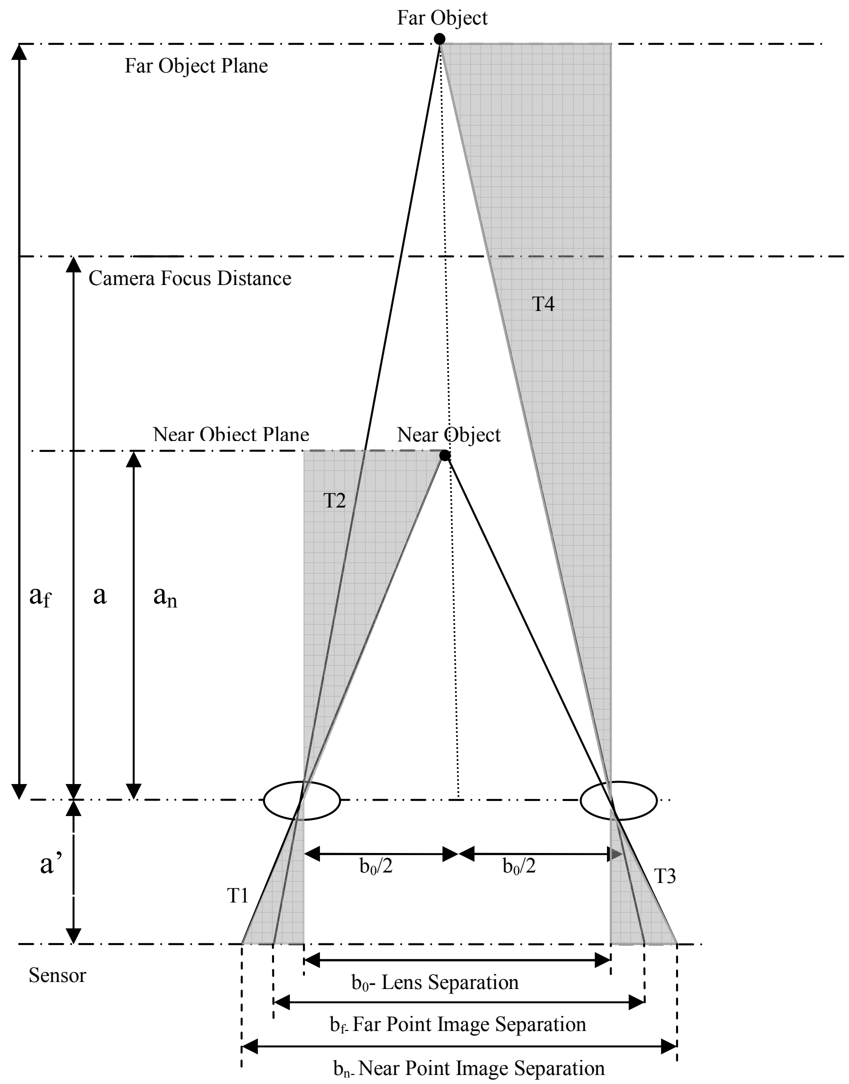


Figure 3.2: Stereo camera geometry.

The images of the two objects captured by the lens are then projected onto an image sensor plane. In Figure 3.2, using similar triangles shown as shaded areas we can derive a relationship between the near and far point objects.

For similar triangles T2 and T1:

$$\frac{b_0/2}{a_n} = \frac{(b_n - b_0)/2}{a'}$$

which gives us,

$$\frac{b_0}{a_n} = \frac{b_n - b_0}{a'}$$

$$b_n = \frac{a' \times b_0}{a_n} + b_0 \qquad ---(2)$$

Equation (2) gives the relationship between the left and right near-point images captured on the sensor. From the similar triangles T3 and T4 we have,

$$\frac{\frac{b_0}{2}}{a_f} = \frac{\frac{b_f - b_0}{2}}{a'}$$

which simplifies to,

$$\frac{b_0}{a_f} = \frac{b_f - b_0}{a'}$$

$$b_f = \frac{a' \times b_0}{a_f} + b_0 \qquad ---(3)$$

This equation gives the distance between the left and right far-point images on the sensor. Now the on-sensor deviation (d) is by definition the difference between the near point image spacing (b_n), and the far point image spacing (b_f).

$$d = b_n - b_f \qquad --- (4)$$

Substituting Eqn. 2 and Eqn. 3 into Eqn. 4 we get,

$$d = \frac{a' \times b_0}{a_n} + b_0 - \frac{a'.b_0}{a_f} - b_0$$

Cancelling out the b_0 term,

$$d = a' \times b_0 \times \left(\frac{1}{a_n} - \frac{1}{a_f}\right)$$

$$b_0 = \frac{\frac{d}{a'}}{\frac{1}{a_n} - \frac{1}{a_f}}$$

$$b_0 = d \times \frac{(a_f \times a_n)}{(a_f - a_n)} \times \frac{1}{a'}$$

Substituting Eqn. (1) here,

$$b_0 = d \times \frac{(a_f \times a_n)}{(a_f - a_n)} \times \left(\frac{1}{f} - \frac{1}{a}\right) \qquad --- (5)$$

Equation 5 enables us to calculate the lens separation or in other words, the camera base distance. If we know the necessary variables namely, the near point distance 'a_n', the far point distance 'a_f', the on-sensor deviation limit 'd' and the distance from the sensor plane to the lens center 'a', we can calculate the optimum camera base separation distance needed to obtain a stereoscopic set of pictures.

Out of the four variables, the near point and far point distances are calculated from the scene space. The variable 'a' is calculated from the camera settings. Experimental studies on psychophysics reveal that the average deviation limit for the human eyes is 1.2mm for the 35mm sensor format and it is approximated as 1/30 and is also called the depth factor. The relationship between the various camera parameters for a stereoscopic image pair derived in this chapter will be later applied to that of our stereoscopic video scenario in the next two chapters.

CHAPTER 4

A STEREOSCOPIC 3D-VIEWING SYSTEM USING A SINGLE SENSOR CAMERA

There have been many ideas in designing stereoscopic vision systems. Some of them include the traditional two-camera system (fixed stereo base), the two camera system with varying stereo base distances, and the attractive single-camera stereo system.

Conventionally, people use the two-camera system and there are a lot of issues in using that. They are as follows:

i. Since there are two cameras, the focal lengths and the fields of view of the two lenses would not be exactly the same and these differences would cause a lot of trouble in perceiving the depth of the captured images.

ii. The alignment and the characteristics of the two sensors (film in the case of analog cameras and CCD/CMOS sensor in digital cameras) would not be precise and would cause intensity differences between corresponding points in the stereo pair.

iii. The two video streams would not be in perfect time synchronization and hence would distort the perceived depth effect.

iv. Even if the two video sensors are aligned perfectly, it is not suitable for frequent adjustments of the stereo base in order to get varying depths of the scene.

The practical implementation of such a system in which we need to eliminate all the above errors is a cumbersome task. These unnecessary geometric and chromatic aberrations [20] can be avoided if we use a single-camera stereo system.

The fundamental issues of stereoscopic 3D view were studied by many works before. As mentioned in the previous chapter, the depth perception of the stereoscopic view is affected mainly by three parameters which are camera focal length, camera convergence distance, and separation of two cameras.

4.1 The Relationship between the Separation and Frame Rate

Instead of two cameras, a single video camera mounted on the window of a moving car was used to capture the scene and later the 3D video was generated which is discussed in this section. The angle of the video camera with respect to the scene, which was mostly 90°, was maintained constant throughout the capture. For the single-sensor video, the effects of the first two parameters remain the same. But the separation of two cameras factor has a new set of issues because the stereoscopic view is achieved by a single-sensor video, not by two separated cameras. That is, the disparity for two eyes is provided by two frames with a *time-slip* between them. How the delay between the two frames affects the depth perception and how to control it deserve a careful investigation in this study.

4.2 Controlling the Inter-Ocular Distance by Frame Delays

The principal factor that affects the depth perceived in the final image is the camera separation which needs to be determined for a scene. A number of approaches have been experimented to imitate, including exact modeling of the user's eye separation. However, depending on the scene content, this may result in a huge or very little perceived depth on the display due to the very large or small retinal disparities respectively. This would cause a lot of eye strain and discomfort for the viewer. Using exact eye spacing is only reasonable for the orthoscopic case, where the object size and depth matches the target display size and is in a comfortable depth range.

In practice this is rarely the case and the camera separation is often determined by trial and error, which is tedious and can easily result in an image suited only to the creator's binocular vision.

The main focus of this study is on the separation of two frames or the inter-ocular distance (*IOD*) between two frames for generating the stereoscopic view because it is different from that in the two-camera systems.

Consider that there are *N* frames for a video. We will name a video frame as

$$frame\ n, \qquad \text{where} \quad n = 0, 1, .. i .. j, .. N.$$

Let the time-period between two consecutive frames be *T*. Then the *time-slip TS* between *frame-i* and *frame-j* for generating the stereoscopic perception will be

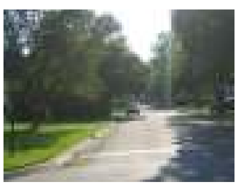
$$TS_{j,i} = T \times (j - i) \qquad --- (1)$$

If the speed of the moving platform be V, one can calculate the separation distance of the two frames as:

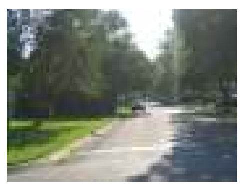
$$I_{j,i} = V \times TS_{j,i} = V \times T \times (j - i) \qquad --- (2)$$

where $I_{j,i}$ is the *inter-ocular distance (IOD)*. IOD, *time-slip TS* and frame difference $(j - i)$ all are equivalent representations of the same quantity. Frame difference $(j - i)$ is a unit less quantity. If TS is expressed in seconds and V is expressed in meters/second, then IOD will be expressed in meters.

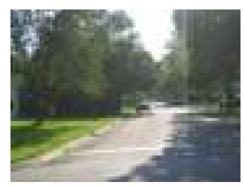
Let $d_{j,i} = j - i$ where d represents the *frame-distance* (*FD*), or equivalently the time-slip between the two video frames for the inter-ocular delay. It is clear that the range of *IOD* or *FD* will affect the stereoscopic view. To understand the issue, consider the five video frames as shown in Figure 4.1.



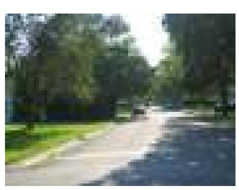
frame-0



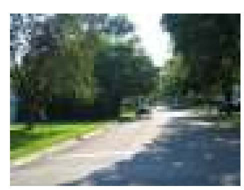
frame-1



frame-2



frame-3



frame-4

Figure 4.1: Five consecutive video frames taken from a single camera.

The above video frames were taken at a rate of 10 frames per second while the platform moves at a speed of 35 miles per hour. Accordingly, the separation distance between two consecutive frames is approximately 1.7 yard. If one can manage to focus his/her left eye on *frame-0* and right on *frame-1* as shown in Figure 4.2, a stereoscopic view of the scene can be obtained. Stereoscopic viewer apparatus such as Loreo 3d lite viewer (discussed in Section 6.2) may also be used to aid in obtaining the stereoscopic view for the image pair.

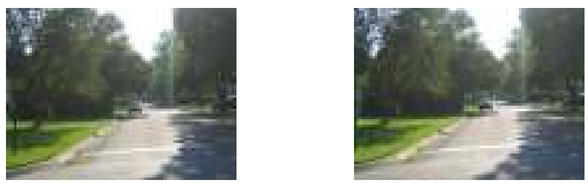


Figure 4.2: Using *frame-0* and *frame-1* for stereoscopic view.

Similarly using *frame-0* and *frame 4*, another stereoscopic view of the same scene is obtained (Figure 4.3) in which the perceived depth is more than that obtained from the pair of *frame-0* and *frame-1*. This means that the depth of the stereoscopic view is increased as the time delay between the two video frames increases.

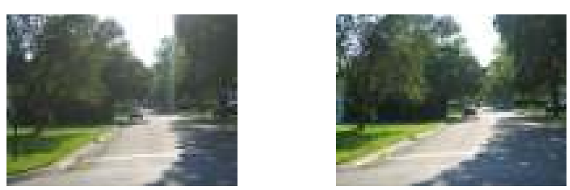


Figure 4.3: Using *frame-0* and *frame-4* for stereoscopic view.

Figure 4.4: Using *frame-0* and *frame-6* for stereoscopic view.

If the *IOD* of the two video frames or *equivalently the FD* becomes even larger, such as *frame-0* and *frame-6* shown in Figure 4.4 where the *IOD* is 8.5 yards, one will find it very difficult to obtain the stereoscopic view from the two frames. This is because the disparity between the two frames exceeds the capability of the human eyes for generating a cue of depth perception. With an extra effort, one may obtain the depth perception for objects which are far in the scene (the parked vehicle, e.g.). Objects which are closer to the sensor (the tree on the left side of the image, e.g.) appear as not related in the two frames.

The above discussion would lead us to the question: how far could the *IOD* be such that two video frames can generate a stereoscopic view with as great a depth as possible? The answer to this question will determine the maximum delays between the two video streams for the left and right eyes, and for the range of the delays. As we saw in Section 3.2, we can conclude that this maximum delay range would be specific to the object in the scene being captured. Consequently, we can design the controls to deliver the required delays between the two frames by adjusting the magnitude of the time-slip between the left and right frames.

To have a solid understanding of the relationship between *IOD* and the depth of perception, we have to perform experimental studies to see how most people behave in obtaining the stereoscopic perceptions from pairs of frames. Apparently, the time delays depend on the speed of the moving platform which could be a ground vehicle or an unmanned air vehicle (UAV). Therefore the time delay is a function of also the speed.

The frame rate is another parameter to consider. Here the NTSC standard is used as an example by which the video is captured at a rate of 30 frames per second. If 10 yards is the maximum *IOD* allowed for a particular application, the frame rate of the video should generate at least two video frames within the *IOD*. By a simple calculation, the moving platform can have a speed up to 300 yards per second which is more than enough for a grand vehicle but not for the UAV. For the latter case, the frame rate must be increased. The video frame rate also depends on the kinds of applications in which the speed of the moving platform is different.

In this chapter we discussed the various new parameters coming into our picture when we consider a stereoscopic video capture. Apart from the camera parameters we saw in Chapter 3, the frame rate, the scene content and the velocity of the moving platform also play a major role in our depth perception of the stereoscopic video.

CHAPTER 5

DEPTH OF FIELD

Stereopsis is a psychological depth cue. By simulating stereopsis, a 3D display can improve the user's depth perception of the virtual scene. Clearly it is desirable to maximize the effectiveness of the stereoscopic effect. Based on geometric considerations and the limits of visual acuity, Valyus [21] calculates that the minimum discernible difference in depth between two adjacent planes increases with distance from the eyes.

He also calculates that beyond 1350 meters it is difficult for an average person to stereoscopically distinguish depths. That is, as far as stereopsis is concerned, different depth planes would appear the same from 1350 meters out to infinity. Modern texts consider binocular stereo vision to be ineffective past 20 meters. This limiting distance is called the radius of stereoscopic vision. In order to improve an observer's ability to discriminate distant depths in the real world a binocular viewing apparatus such as a telestereoscope can be used. This will have the effect of looking at a small model of the scene which brings the stereo image closer to the viewer [22].

When observing a huge structure or a small mountain at a distance of 3 km, stereo vision contributes almost nothing to our understanding of the spatial shape. However, if we create a stereo pair of images with the viewpoint separated by 500 m (say), we will obtain a useful enhanced “hyper stereo” image. This technique is used extensively in stereo photogrammetry. Conversely, if we wish to understand the spatial structure of an object only a few meters away, we will be better off with our normal eye separation.

5.1 Depth Of field of a scene from an aerial platform

Hyper Stereoscopy means creating a stereoscopic pair of images/video with the camera separation distance that is far greater than the human eye separation distance. This would help us perceive the depth of an object which is very far away and that whose depth cannot be obtained with naked eyes or when the camera separation is closely equal to the human eye separation.

As stated in the previous section, modern texts cite the limiting stereo vision distance as approximately around 20 meters. With the help of hyperstereoscopy and this work, this number can be extended way further.

5.2 Practical Example

Now let us apply the formulae we derived in Chapter 2 and Chapter 3 to a practical example of a stereoscopic system.

Let us say we want to capture a video of a mountain from an aircraft. Let us say that the nearest point (the peak of the mountain in this case) seen from our camera is about 10,000 ft away and the farthest point (or the foot of the mountain here) is about 25,000 ft away. We will first calculate the stereo base required to obtain an optimum depth effect.

Using Eqn. (5) we get

$$b_0 = d \times \frac{a_f \times a_n}{a_f - a_n} \times \left(\frac{1}{f} - \frac{1}{a}\right)$$

This can be approximated as

$$b_0 = \frac{1}{30} \times \frac{a_f \times a_n}{a_f - a_n} \times \left(\frac{f_v}{f_c}\right)$$

where d – the on sensor deviation has been approximated as 1/30.

F_v- the focal length of the viewing lenses.

F_c- the focal length of the camera lenses.

Here in our case we have,

$$a_n = 10000\, ft \; and \;\; a_f = 25000\, ft$$

And here let us assume that the focal length of the camera lens and the viewing lenses are the same.

$$f_c = f_v$$

Therefore we have,

$$b_0 = \frac{1}{30} \times \frac{a_f \times a_n}{a_f - a_n} \times \frac{f_c}{f_v}$$

$$b_0 = \frac{1}{30} \times \frac{25000 \times 10000}{25000 - 10000} \times 1$$

$$b_0 \cong 555.55ft$$

So a camera base distance or inter-ocular distance (IOD) of approximately 555 ft is needed to obtain a good stereoscopic view of the mountain from the aircraft. The depth factor which was assumed to be 1/30 can in fact take values from 1/20 till 1/50 or 1/60 depending on the user's comfort. A smaller depth factor (1/60) would decrease the perceived depth whereas a larger depth factor (1/20) would increase the perceived depth and hence the discomfort in viewing the video. Plugging in these values gives a range of camera base distance,

$$277\ ft < b_0 < 833ft$$

This implies that we have the freedom of varying the camera base distance within this range. Now if we consider the velocity of the aircraft too we can have an idea of the frame difference needed for our system.

A velocity of 100 ft/second would give the frame difference as a unit of time,

$$\frac{277}{100}\ s < Time\ Slip < \frac{833}{100}\ s$$

$$2.7\ s < Time\ Slip < 8.3\ s$$

Thus the frame delay between the two video frames can take a value between 2.7 seconds and 8.3 seconds. The optimum time slip for this case would be 5.5 s for the average human eyes ($\frac{555ft}{100\times\left(\frac{ft}{sec}\right)} = 5.55sec$).

5.3 Considerations

In this example scenario we discussed now or any other applicable situation, there are several things to be considered for such an implementation of a stereoscopic system. They are as follows:

i. Many of the videos tested had an angle of 90° of the optical axis of the camera with respect to the scene. The video capture was also tried with different angles such as 0°, 15°, 45°, 135°, and 180° were used to capture the scene. Stereoscopic depth was still perceived in all these angles except for the two parallel angles 0° and 180°. This gives us more freedom in the sense that during a video capture the camera angle can be varied in order to concentrate more on our object of interest and still perceive the depth in the scene.

ii. Distance to near-point object has to be measured. One can use optical or laser range finders to determine this.

iii. Distance to far-point object which is a bit more difficult to measure also needs to be known. Long range finders or at least a best guess estimate has to be used.

iv. The field of view of the two frames being presented needs to be given careful consideration as well. This is discussed in more detail in the next section.

5.4 Field of view

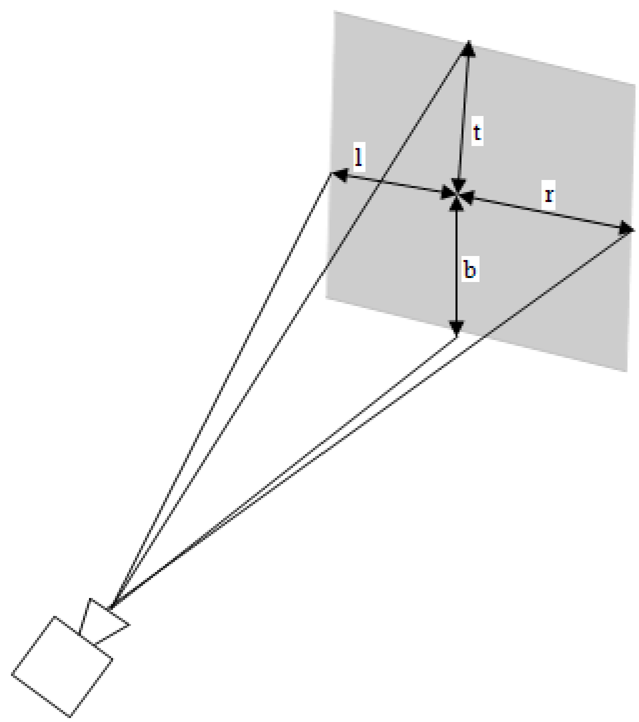


Figure 5.1: Field of view of a camera.

As stated earlier, the field of view is also an important factor in the stereoscopic perception of a scene. Since the camera system employed here in this work is a parallel single camera system, the field of view has a different effect on the depth perception when compared to that of the usual convergent two camera system.

Let us consider a camera system capturing a scene and let us define the field of view of this system. The field of view (FOV) is determined by six parameters: the near clipping plane, the far clipping plane, the left boundary l, the right boundary r, the top boundary t, and the bottom boundary b as shown in Figure 5.1. In this work, the near and far clipping planes are nothing but the near point and far point distances that we saw earlier in Chapter 2 and Chapter 3.

5.4.1 Field of view in a parallel camera system

Figure 5.2 shows a single camera system with its field of view used to capture a stereoscopic image from a scene. Since in this system, the single camera is moved

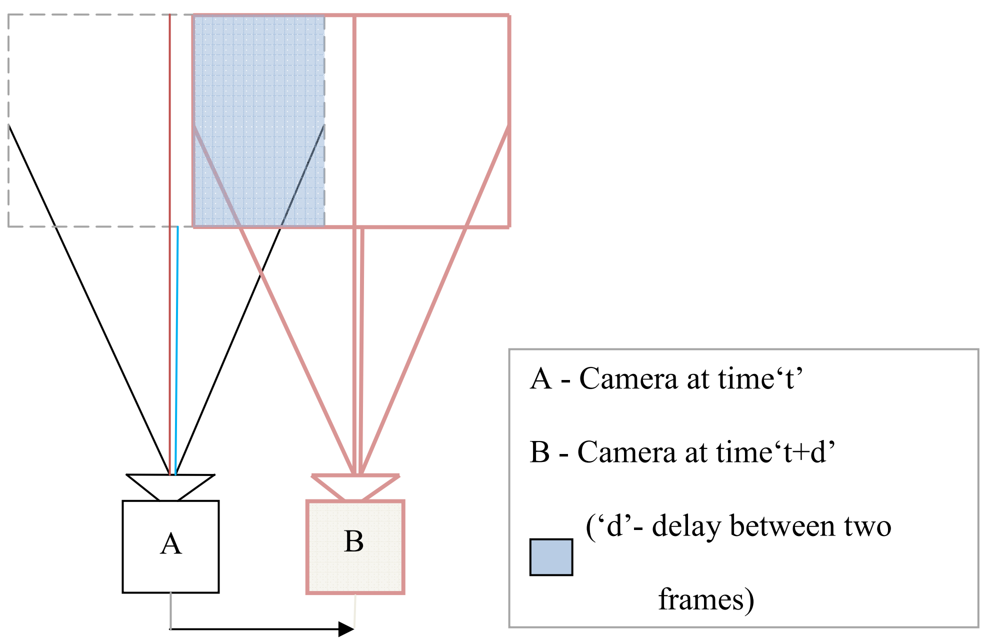


Figure 5.2: Single Parallel Camera System with different FOVs

horizontally while capturing the video, there would be no change in the vertical parameters of the field of view namely ‘t’ and ‘b’. So the only two parameters that affect the FOV in our case are the left boundary ‘l’ and the right boundary ‘r’. The effects of altering these two parameters are discussed in the next section.

5.4.2 Effects of Increased and Decreased FOV

The effects of a large FOV as well as a small FOV have been studied extensively before and this section gives a brief summary of the results of previous works. Pfautz [16] has well summarized the studies done previously about the effects of varying the FOV in his dissertation.

A wide FOV has proved useful in a variety of tasks in the real world and computer-generated worlds. Various experiments have been done regarding FOV using virtual environments and flight simulators to perform different kinds of geographic orientation and target detection functions .A lot of psychophysical experiments too have been conducted with human subjects. These studies show that an increased FOV is necessary for the user of a virtual environment to adequately perform simple tasks like estimating distance or locating objects in space.

Hence we can conclude that for a better stereoscopic perception, a larger FOV would be more helpful though some cases of nausea and dizziness arise because of the wide FOV.

5.4.3 Cropping and Zooming In the Image

Though a large FOV gives the viewer a better understanding of the scene when compared to a smaller FOV, the trade-off would be a lesser depth of the scene. This is because for a larger FOV, the overlap between the two frames has to be large and hence the frame difference has to be small. In case of a small overlapping FOV, a point or an object in one image area may not have its corresponding point in the other image area. This means such a point or an object will be of no use in stereoscopic perception. Hence the cropping of the non-overlapping area in the images before presenting them to the observer would enhance the depth. Also it would be convenient if the user has the control to zoom in a particular object in the scene that he wants to perceive in depth in greater detail.

In this chapter we discussed the application of our results and study done in Chapter 3 and Chapter 4 to a practical example scenario. The considerations of implementation of such scenario were also discussed after which the effect of varying the field of view was studied.

CHAPTER 6

SOFTWARE EXPERIMENT

The purpose of this work was to design an user-friendly stereoscopic viewing software that helps the viewer in perceiving the depth of the scene comfortably. The developed software should also be robust and stable in all conditions. Hence, it was decided that a graphical user interface (GUI) application in Visual C++ would be the best choice.

6.1 OpenCV and MFC library

This application was developed using the MFC Library, a set of classes constituting an application framework, which is the framework of an application written for the Windows API (Application Programming Interface).OpenCV library developed by Intel Corporation was used for this work. OpenCV is an open source computer vision library developed for usage in C or C++. It is optimized and intended for applications like real-time image and video processing. Example applications of the OpenCV library are Human-Computer Interaction (HCI); Object Identification, Segmentation and Recognition; Face Recognition; Gesture Recognition; Motion Tracking, Ego

Motion, Motion Understanding; Structure From Motion (SFM); and Mobile Robotics.

The advantage in using OpenCV is that it is hardware and operating system independent and hence the application developed can be used in a variety of platforms. The main four modules which comprise OpenCV are as follows:

- cv- the main OpenCV functions
- cvaux- experimental OpenCV functions
- cxcore- data structures and linear algebra support, and
- Highgui- GUI functions.

As the work involves capturing a video stream, manipulating the video sequence and displaying it to the user, this open source library was used.

6.2 StereoVideo – The Stereoscopic Video Viewer

StereoVideo is the 3d software developed as a result of this study. This software takes a single video stream as input and displays two different frames of the video in the GUI: one on a left video window and the other on a right video window.

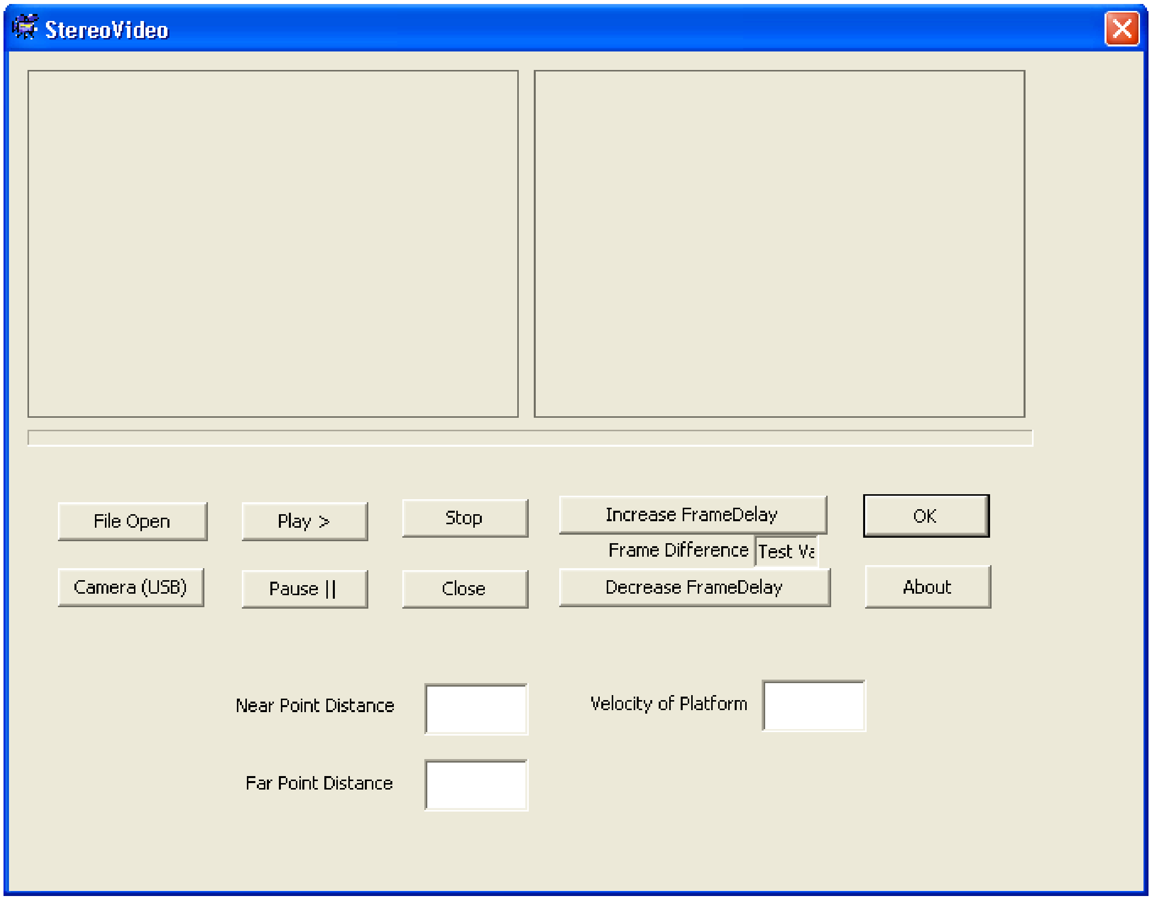


Figure 6.1: StereoVideo – The Stereoscopic 3D Viewing Software.

The user wears a 3d glass to see the video. In this study, Loreo 3d Lite Viewer was used. It is a parallel format 3d viewer intended for viewing side-by-side stereoscopic image pairs. Apart for viewing 5x7 inch wide prints on a monitor screen, it can also be used for viewing postcard stereo images. The figure below is a snapshot of the StereoVideo software.

6.3 Results Used in StereoVideo

6.3.1 Phase I

The discussion explained in the section 3.2 was the first phase used in developing this software. To recount that, two frames f and (f+n) would be presented to the left and right eyes (where f, n are integers: f= 0, 1...etc and n=1, 2…etc) through the respective windows on the developed software. Two controls would enable the user to adjust the stereoscopic frame difference; one is the Increase FrameDelay button and the other is the Decrease FrameDelay button.

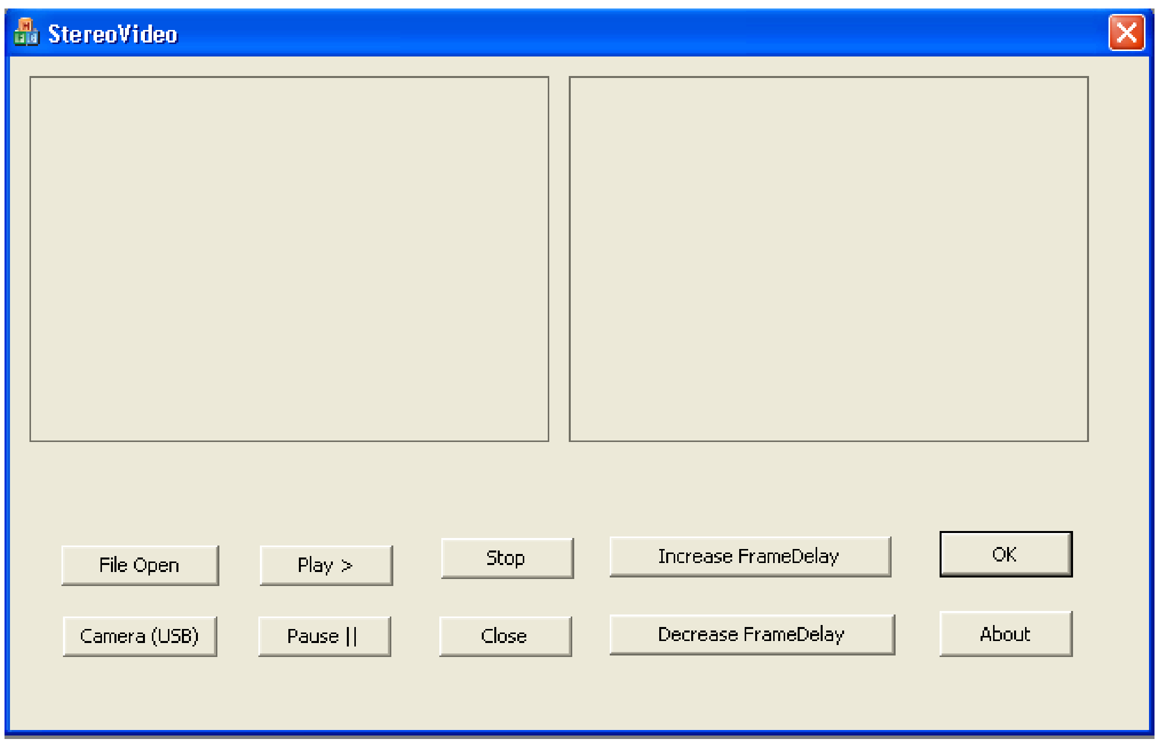


Figure 6.2: Phase I of StereoVideo with Increase and Decrease FrameDelay buttons.

6.3.2 Phase II

After the controls to change the frame difference were implemented, the need to determine the optimum frame delay required for a scene arose. So if the distance of the nearest point and the farthest point in the scene and the velocity of the camera platform are known and entered by the user, the software would determine the optimum frame difference based on equation in section 5.2.

Figure 6.3: Default frame difference (fd=1) for a sample video in StereoVideo.

For example, in the above figure the default frame difference of fd=1 is used when a video is played. Let us say that the nearest point distance in the scene is found as 50 ft and the farthest point distance is found as 300 ft. Also let us assume the velocity of the camera platform as 10 ft/second and the depth factor as 1/30. The software calculates the desired frame delay according to equation in section 5.2 and adjusts the frames accordingly.

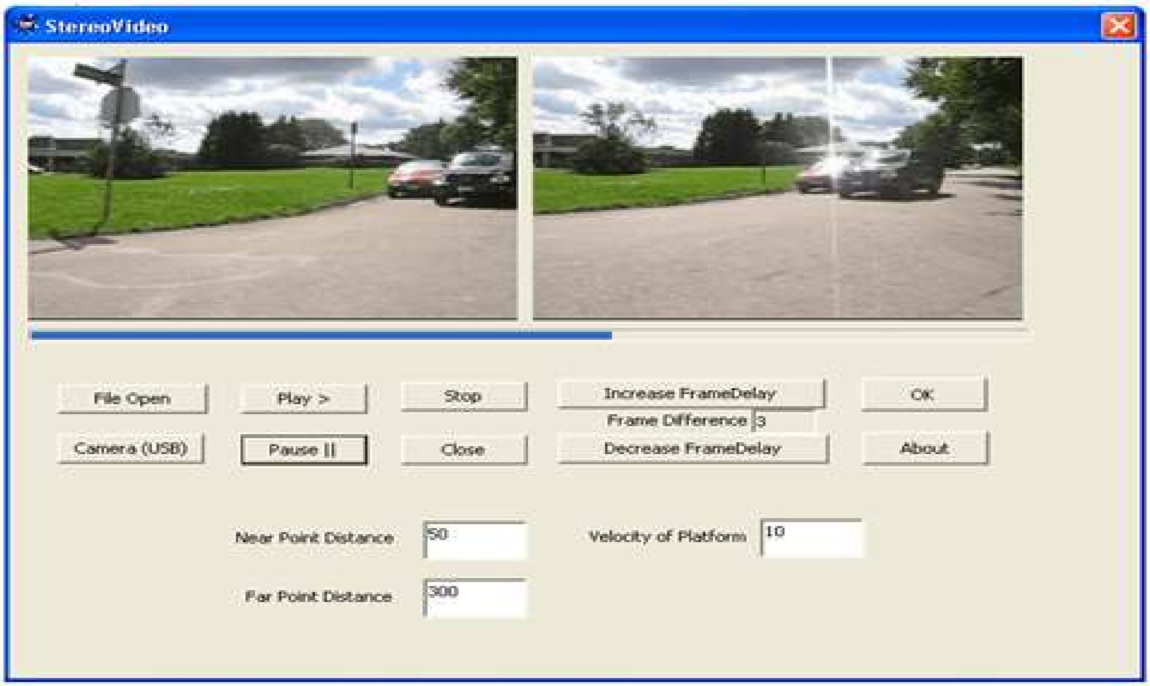


Figure 6.4: Optimum frame difference (fd = 3) for a sample video in StereoVideo.

The above figure 6.4 shows that a frame difference of 3 has been suggested by the software for optimum viewing in these conditions. If the user feels uncomfortable with this frame difference he can increase or decrease (using the Increase and Decrease FrameDelay buttons) it according to his needs. One has also to note that a larger frame difference for a prolonged time may cause eyestrain after sometime.

6.4 Components in the StereoVideo GUI

This section describes the functions of the various components in the GUI. They are as follows:

i. File Open

   This button selects the source video file to be played. File types can be .avi or .mpeg and also it has to be noted that OpenCV does not support some codecs of .avi especially the DV codec. Cinepak and MJPEG codecs work fine. In this work, Cinepak codec .avi files were used.

ii. Camera (USB)

   This button allows us to use a live video from a USB camera port. This would be useful in many real-time applications like enemy territory surveillance.

iii. Play

   After the video source has been selected, pressing this button starts playing the video.

iv. Pause

   This button allows us to freeze the video. This button will be useful for the user to initially get accustomed to the 3D view by viewing the paused video first and then maintaining that depth perception through the rest of the video.

v. Stop

Pressing this button would stop the video and reset it back to the initial position.

vi. Close

This button closes the connection between the application and the source file.

vii. Increase/Decrease FrameDelay

This button allows us to increase or decrease the frame delay between the frames in the left and right windows. The user can change it according to his/her comfort.

viii. Ok

This button closes the application.

ix. About

This button gives information such as the version and copyrights of the software.

x. Frame Difference

This static text box displays the current frame difference with which the two video frames are being played. This is a read-only text box.

xi. Near point, Far point and Platform Velocity

These edit boxes allow the user to enter values for the near point distance, far point distance, and the platform velocity respectively. After the software gets these values, an optimum frame difference is calculated and the videos are automatically adjusted and displayed. The user if necessary can change this frame difference according to his comfort using the Increase/Decrease FrameDelay buttons.

StereoVideo, the 3D viewing system with all its user-friendly and nice features will be of great use for stereoscopic applications and research. Also because of its robustness and stability in various kinds of operating system environments, this can be used in real-time stereoscopic applications with great ease.

CHAPTER 7

CONCLUSION AND FUTURE WORK

The use of a single camera system to capture a stereoscopic video has been presented in this thesis. First the system was considered to capture a stereoscopic pair of images where the relationship between the camera separation distance and the other scene parameters were derived. This relationship was then applied to the stereoscopic video pair of frames in which new parameters namely, the video frame rate and the platform velocity were also incorporated in our study.

Then an optimum frame delay for a specific scene was calculated and implemented in a Graphical User Interface (GUI) application. Several sample videos were captured and tested using the software developed. Subjects were able to perceive good 3D depth for the sample video sequences. StereoVideo, the 3D software developed in this study has user-friendly features and can be used for real time or even offline applications.

Future work on this thesis can be a consideration of the field of view which is different in the two frames presented to the user. Eliminating the non- overlapping areas in the two frames as well as the ability to zoom in the object of interest would enable more comfort for the user. These features can be added to StereoVideo which would also increase the depth of the scene greatly. Though different convergent angles of video capture were tested, a detailed study on the effects of perceived depth in these cases can be done.

REFERENCES

1. L. Lipton, *Stereographics: Developers Handbook*, Stereographics Corporation, 1997.

2. D. B. Diner and D. H. Fender, eds., *Human engineering in stereoscopic viewing devices*, Plenum Press, ISBN 0-306-44667-7, 1993.

3. I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis*, Oxford University Press, ISBN 0-19-508476-4, 1995.

4. D. McAllister, ed., *Stereo computer graphics and other true 3D technologies*, Princeton University Press, ISBN 0-691-08741-5, 1993.

5. T. Mitsuhashi, "*Subjective Image Position in Stereoscopic TV Systems - Considerations on Comfortable Stereoscopic Images*," in Proceedings of the SPIE, vol. 2179, pp. 259 – 266, Mar. 1994.

6. R. Sand and A. Chiari, eds., *Stereoscopic Television: Standards, Technology and Signal Processing*, European Commission, Directorate General XIII-B, Brussels, 1998.

7. C. Ware and G. Franck, "*Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions*," tech. rep., University of New Brunswick, Technical Document #TR94-082, 1994.

8. E. Goldstein, *Sensation and Perception* (3rd Edition), Belmont, California: Wadsworth Publishing, 1989.

9. J. Gibson, *The Ecological Approach to Visual Perception*, London: Lawrence Erlbaum Associates, 1986.

10. B. Gillam, *The Perception of Spatial Layout from Static Optical Information in W. Epstein & S. Rogers (Eds.),* Perception of Space and Motion, pp. 23-67. New York: Academic Press, 1995.

11. L. Kaufman, *Sight and Mind: An Introduction to Visual Perception*, Oxford University Press, USA, 1974.

12. G. Hone & R. Davis, *Brightness and Contrast as Cues to Depth in the Simulator Display: Cue Combination and Conflict Resolution,* in Proc. SPIE – Human Vision, Visual Processing and Digital Imagery VI, vol. 2411, pp. 240-249,1995.

13. C. Wheatstone, *Contributions to the physiology of vision. Part the first: On some remarkable and hitherto unobserved phenomena of binocular vision*, Phil. Trans. R. Soc.II, 371-394, 1838.

14. P. Buser & M. Imbert, *Vision,* Cambridge, Massachusetts: The MIT Press, 1992.

15. J. Cutting & P. Vishton, *Perceiving Layout and Knowing Distance: The Integration, Relative Potency and Contextual use of Different Information about Depth* in W. Epstein & S. Rogers (Eds.), Perception of Space and Motion, pp. 69-118. New York: Academic Press, 1995.

16. J. Pfautz, *Distortion of Depth Perception in a Virtual Environment Application*. Master's Dissertation, Massachusetts Institute of Technology, 1996.

17. J. M. Foley, *Binocular distance perception*, Psychological Review, 87(5), 411-434, 1980.

18. M. Krueger, *Adaption in depth perception using stereoscopic TV display*, M.S. Thesis, Department of Industrial Engineering, University of Toronto, 1991.

19. http://home.vicnet.net.au/~vic3d/bases.html

20. G. Jang, S. Kim, I. Kweon , "*Single-camera Catadioptric stereo system* "- Proc. of Workshop on Omnidirectional Vision, 2005.

21. N. A. Valyus, *Stereoscopy*, The Focal Press, London and New York, 1966.

22. Z. J. Wartell, *Stereoscopic Head-tracked displays: Analysis and Development of Display Algorithms*, PhD Dissertation, Georgia Institute of Technology, 2001.

23. J.D. Pfautz, *Depth perception in computer graphics*, Ph.D. Thesis, Department of Computer Science, University of Cambridge, 2002.

24. S. Nagata, *How to Reinforce Perception of Depth in Single Two-Dimensional Pictures,* in S.Ellis (Ed.), Pictorial Communication in Virtual and Real Environments, pp. 527-545.London: Taylor & Francis, 1993.

25. K.R. Boff & J.E. Lincoln, *Depth Perception: Engineering Data Compendium: Human Perception and Performance*, AAWRL, Wright- Patterson AFB, pp.1053-1128, 1988.

26. J. Fewerda, Section 24.4, "*The World of 3D*", Second Edition, 1987.

27. R. Mannle, "*An Analysis of Depth Perception and Composition*", Stereoscopy, Series 2 No. 15, June 1993.

28. J. Bercovitz, *Image-Side Perspective and Stereoscopy*, Lawrence Berkeley National Laboratory, Berkeley, CA, 1998.