ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

2D TO 3D VIDEO CONVERSION

M.Sc. Thesis by
Özlem AYDOĞMUŞ

Department : Electronics and Telecommunication Engineering

Programme : Telecommunication Engineering

JUNE 2011

ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

2D TO 3D VIDEO CONVERSION

M.Sc. Thesis by
Özlem AYDOĞMUŞ
(504081326)

Date of submission : **06 May 2011**
Date of defence examination: **09 June 2011**

Supervisor (Chairman) : **Prof. Dr. Melih PAZARCI (ITU)**
Members of the Examining Committee : **Prof. Dr. Bilge ÜNSEL (ITU)**
Doç. Dr. Uluğ BAYAZIT (ITU)

JUNE 2011

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

2 BOYUTLU VİDEOYU 3 BOYUTA DÖNÜŞTÜRME

YÜKSEK LİSANS TEZİ
Özlem AYDOĞMUŞ
(504081326)

Tezin Enstitüye Verildiği Tarih : 06 Mayıs 2011
Tezin Savunulduğu Tarih : 09 Haziran 2011

Tez Danışmanı : Prof. Dr. Melih PAZARCI (İTÜ)
Diğer Jüri Üyeleri : Prof. Dr. Bilge GÜNSEL (İTÜ)
Doç. Dr. Uluğ BAYAZIT (İTÜ)

HAZİRAN 2011

FOREWORD

I would like to express my deep my gratitude to my advisor, Prof. Dr. Melih PAZARCI, for his valuable advices, comments, and guidance throughout this thesis.

Also I would like to give special thanks to my family for their generous understanding and moral support.

JUNE 2011

Özlem AYDOĞMUŞ

Telecommunication Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

2D **:** 2 Dimensional
3D **:** 3 Dimensional
LC **:** Liquid Crystal
MTD **:** Motion Time Difference
LCD **:** Liquid Crystal Display

LIST OF TABLES

Page

LIST OF FIGURES

Page

2D TO 3D VIDEO CONVERSION

SUMMARY

In recent years, with the development of 3D technology 3D video content has become a need for the consumer electronics market. 3D display technology has reached the quality that consumers can buy and use it in their homes. However, compared with 3D display devices, the development of 3D video content has remained behind. The reason of that is stereoscopic cameras, which are required for shooting 3D content of the video, are very expensive and the technical setup of these cameras is difficult. At the same time, for the problem of 2D to 3D conversion, 2D content information is manually converted to 3D graphics. However, these shooting and manually converting methods are expensive, time consuming and labor-intensive methods. Instead of these methods, there are automatic 2D to 3D conversion algorithms, which generate 3D effect from conventional videos without knowing specific camera parameters.

In this thesis, 3D effect generation is aimed by applying 2D to 3D conversion algorithm applied to a conventional 2D video with unknown camera parameters. For implementing this algorithm human depth perception and the relation between 2D and 3D content information (depth map) were investigated. For the 2D to 3D conversion algorithm, two different methods are used to obtain the 3D effect.

In the first method, the depth map was generated by using motion vectors that are obtained using motion estimation. In the second method, the depth map was generated by using edge information. With respect to depth information, both methods use the same shift method to create artificial stereo image pairs (left and right images).

Results were converted to the anaglyph format for viewing stereoscopic videos. The other reason to use the anaglyph format is that it is the cheapest way to view the stereoscopy.

Created stereoscopic videos are evaluated subjectively. Two different methods are used and compared to each other. Computational run time results are also calculated and compared to each other.

2 BOYUTLU VİDEOYU 3 BOYUTA DÖNÜŞTÜRME

ÖZET

Son yıllarda 3 boyut teknolojisinin gelişmesiyle birlikte tüketici elektroniği piyasasında 3 boyutlu video içeriği ihtiyaç haline gelmiştir. 3 boyutlu gösterim teknolojisi tüketicilerin sahip olup, evlerinde kullanabilecekleri kaliteye erişmiştir. Fakat 3 boyutlu video içeriği 3 boyutlu gösterim cihazlarına oranla geride kalmış bulunmaktadır. Bunun sebebi, 3 boyutlu içerik oluşumu için gerekli olan stereoskopik kameraların çekim maliyetinin yüksek ve kameraların teknik kurulumunun zor olmasıdır. Aynı zamanda 2 boyuttan 3 boyuta çevrim problemi için 2 boyutlu içerik, manüel olarak 3 boyutlu grafiğe çevrilmektedir. Fakat bu iki yol, pahalı, zaman alıcı ve emek gerektiren yöntemlerdir. Bu metotlar yerine bilinen yöntemlerle çekilmiş 2 boyutlu videoları kamera parametre bilgileri olmadan 3 boyut efekti verebilen otomatik 2D/3D dönüştürücü algoritmaları bulunmaktadır.

Bu tez çalışmasında, kamera parametleri bilinmeyen 2 boyutlu bir videoya 2D/3D dönüştürme algoritması uygulanarak 3 boyutlu görüntü efektinin oluşturulması hedeflenmiştir. Algoritmayı kurgulamak için, insan derinlik algısı ve 2 boyutlu içerik bilgisi ile 3 boyutlu içerik bilgisi arasındaki ilişki (derinlik haritası) araştırılmıştır. 2 boyuttan 3 boyuta dönüşüm algoritmasında iki farklı method kullanılarak 3 boyut efekti elde edilmeye çalışılmıştır.

İlk yöntemde, derinlik haritası, hareket kestirimi yöntemi sonucu elde edilen hareket vektörleri ile oluşturulmuştur. Ve ikinci yöntemde derinlik haritası kenar bilgisi kullanılarak oluşturulmuştur. Derinlik bilgisi sayesinde, iki metot da aynı kaydırma algoritmasını kullanarak yapay stereoskopik çift (sol ve sağ resim) oluştururlar.

Stereoskopik video sonuçları görebilmek ve aynı zamanda en ucuz yöntem olduğu için, sonuçlar anaglif formata çevrilmiştir. İzleyebilmek için anaglif gözlüklere ihtiyaç duyulmaktadır.

Oluşan stereoskopik videoların kalitesi sübjektif olarak değerlendirilmiş ve kullanılan iki farklı yöntem de birbirleri ile kıyaslanmıştır. Dahası hesaplama zamanları hesaplanmış ve birbirleri ile karşılaştırılmıştır.

1. INTRODUCTION

Recently, 3D video signal processing has become popular in 2D-TV consumer markets. The 2D to 3D stereoscopic conversion technology has been used in various applications including broadcasting, communication, computer games, medicine, and education and so on. There are many developed 3D display systems. However, 3D imaging technology has not been successful in the market. The reason for this situation is the lack of 3D content. To solve this problem, with the help of stereoscopic cameras, the visual information can be captured, and this captured 2D content can be converted to 3D. This method brings a new technology to capture the video, and it is expensive and time-consuming. Instead of this, without a change in capture technology, a conventional 2D video can be converted to 3D. In this way, videos, which were not captured by stereoscopic cameras, can be watched with a 3D sensation. There are various 3D display devices; auto stereoscopic displays, LCD shutter glasses, polarization based separation, and anaglyphs. According to the use of these devices, there are many developed stereoscopic content generation algorithms.

1.1 Purpose of The Thesis

The main objective of this study is to propose an automatic stereoscopic conversion algorithm based on a computer vision technique. The implementation does not reconstruct the real 3D coordinates of any object. It is aimed to give 3D effect from a single view. In the 2D to 3D conversion system implementation, the anaglyph method was selected as the 3D visualization method. The reason of the selection is that it does not require a special display system, in other words, it is suitable for any conventional display system, and it is the cheapest solution between its counterparts. Moreover, the sub main objective is to make a comparison between depth estimation methods that are used to synthesize a stereoscopic pair for the anaglyph projection. The results of this thesis can easily be applied to LCD shutter glass displays by skipping the anaglyph formation step.

1.2 Historical Overview

The stereoscopic 3D viewing studies began nearly after the invention of 2D imaging technology, so stereoscopic 3D viewing is an old and known technique as 2D imaging technology. Except paintings and drawings, the first technological 2D image representation technique, i.e., photography was invented at the beginning of the 19th century [1]. Since then, 2D still imaging techniques have been improved. In 1867, William Lincoln invented the “zoopraxiscope” and this device was used to create movies [1]. Observing images from remote places was accomplished by the invention of television as early as 1920s (Edouard Belin and John LogieBaird) [1]. At almost the same time as the development of 2D imaging technology, in 1838, Sir Charles Wheatstone invented the first 3D display system “stereoscope” used to deliver stereoscopic 3D images. The stereoscope was developed and minimized by Sir David Brewster in 1844 [2]. With the help of this improvement, stereoscopic still photography was popular both in the U.S. and in Europe at the end of the 19th century. The first stereoscopic cinema appeared in 1922 in the red/green anaglyph format and it utilized dual strip projection. But, especially in 1950, 3D movies became quite popular with the release of the first color stereoscopic feature. In 1955, the interest was lost because of erroneous projection techniques and uncomfortable visualization [1]. Although the first known experimental 3D TV broadcast in the USA was in 1953, the first commercial 3DTV broadcast took place in 1980 in the USA [2]. Nowadays, 3D movies are very popular, because the display and capture technologies are quite qualified and give good results for 3D enhancement.

2. BACKGROUND KNOWLEDGE AND RELATED WORKS

In order to construct a 2D to 3D video conversion, background knowledge of generating stereo images must be known. For this reason, some information about human depth perception, the fundamentals of 2D images on 3D displays, related conversion algorithms, and general information about the proposed system are given in this chapter.

2.1 Human Depth Perception

The fundamental principle of stereoscopy is based on the human visual system and perception. Because of this reason, a clear understanding of how a stereoscopic 3D image is perceived by a user is required.

In 280, Euclid explained that if two eyes see the different images at the same time, it would provide the depth perception [3]. This event is called binocular parallax or binocular disparity and it is the most important cause of depth perception. The human brain fuses the two images that have small differences from each other, and it produces a 3D depth perception, in other words stereopsis. Stereopsis can provide information about depth relationships of objects in a scene.

Moreover, beside the binocular depth cue, the human visual system makes use of other depth cues to interpret the illusion of depth. They are monocular depth cues and oculomotor cues [4].

2.1.1 Monocular depth cues

Beside the binocular vision, using just one eye only can provide the depth perception [8]. There are some monocular depth cues (also known as pictorial depth cues):

Interposition: An object that occludes another is closer. It suggests a depth ordering to the human visual system.

Shading and Brightness: Light reflections from objects provide an understanding of their depth relationships. The shadows include shape information of the object, and far away objects seem dimmer.

Surface texture gradient: Closer objects have more detail. In other words, a texture of constant size on objects will vary in size on the retina with distance.

Relative Size: A large object seems closer and a smaller object is judged further away than the same object which has a large image on the retina.

Linear Perspective: Parallel lines converge at a single point far away. With the help of the perspective, the same size object at different distances projects a different size image on the retina.

Aerial Perspective: Because of atmospheric effects, such as fog, dust, rain, further away objects are blurrier.

Motion parallax is another important depth cue that provides the depth perception. While passing in front of a landscape, the objects close to us move faster than the objects that are further away. In addition, motion parallax does not make stereopsis redundant; instead, stereopsis and motion parallax combined result provides a better depth perception result [5].

Accommodation is another oculomotor depth cue for the monocular vision. When the distance of an object changes, eye accommodation provides focus on the object. In other words, it enables maintaining a clear image [8].

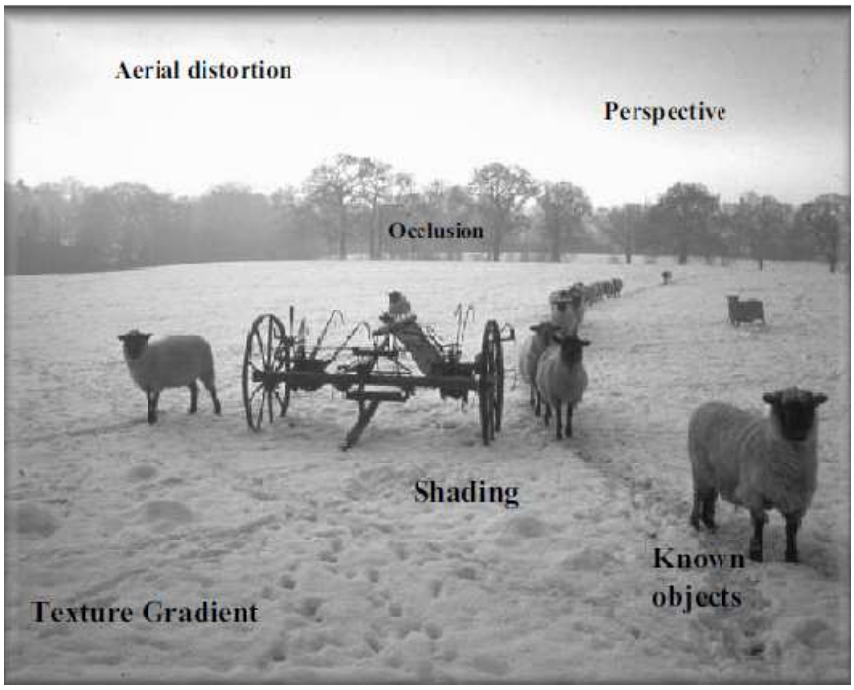

Figure 2.1 : Pictorial depth cues [5].

2.1.2 Binocular depth cues

2.1.2.1 Stereopsis

The distance between left and right eye is about 65 mm [6]. By the disparity of images from the two eyes, the human brain can perceive and determine the depth of the observed objects. In figure 2.2, the left and right eyes see the pyramid, and get different images. It is the binocular disparity. The brain combines the two images and the composite stereo shape of the pyramid occurs.

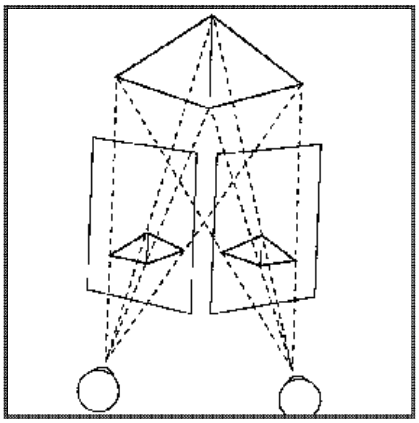

Figure 2.2 : Image generation from binocular disparity [3].

2.1.2.2 Convergence

Convergence is an oculomotor depth cue of binocular vision [8], and it is the angle between the left and right eye sights.

For nearer distances the convergence will be larger and on the contrary, if the distance is further away, the convergence will be smaller.

To define the distance of the object, the angle of the two eyesights need to be modified by extra ocular muscles. Especially when it cooperates with eye accommodation, the change of convergence has significant contributions to the depth perception in near distances. However, when the distance is over 10 m, human visual system cannot perceive the depth of objects using convergence. The change of angle of the two eyes can be classified as follows:

Positive-Parallax: Two eyes focus behind the screen, and the image will be observed behind the screen.

Zero-Parallax: As shown in figure 2.3, two eyes focus on the screen and the image will be observed on the screen.

Negative-Parallax: Two eyes focus in front of the screen, and the image will be observed in front of the screen.

In this way, while observing a scene, the human visual system divides the scene into three levels: far, middle, near, and these parallax situations are used artificially to create a 3D effect illusion for viewers.

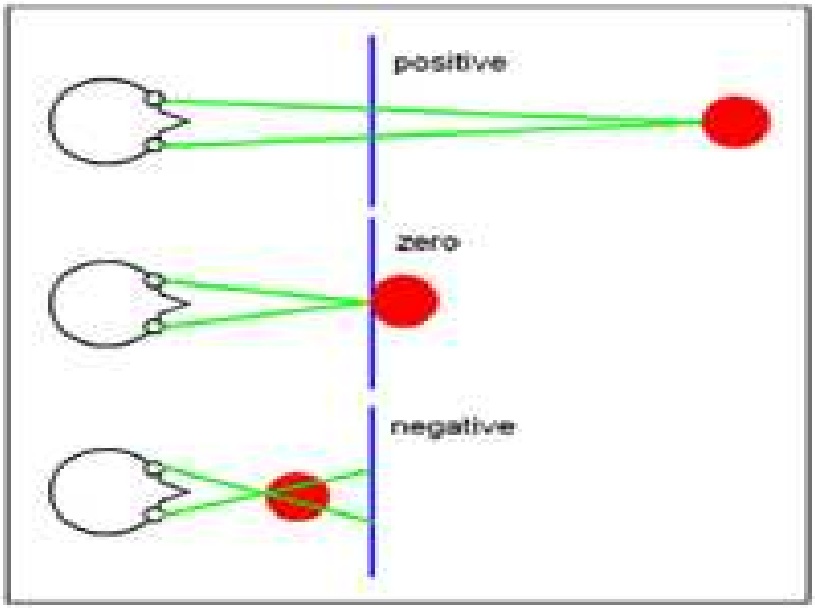

Figure 2.3 : The positive, negative, and zero parallax [7].

Knowing these methods and binocular disparity will be helpful for the 2D to 3D image conversion algorithm to simulate binocular vision.

2.2 Methods of the 3D Display

The subject of this thesis is stereo perception in electronic devices. At present time, there are three ways of displaying in 3D.

- Computer Graphics
- Stereoscope
- Stereo perception in electronic devices.

First, a 2D image can be modeled with the use of pictorial depth cues and it will give the depth perception to the brain. Here, two eyes see the same image at the same time, and it can be done with the drawing method or using some special mathematical representation of any three-dimensional surface [10]. This method is the base of the 3D computer graphics.

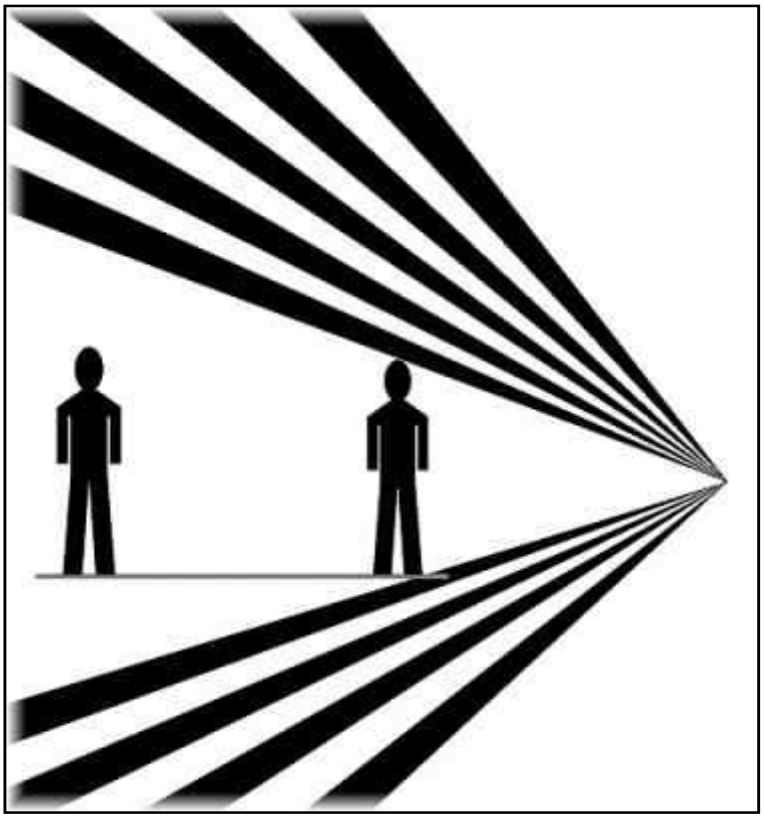

Figure 2.4 : Converging lines create depth in the image [9].

The second way is generating a stereo perception. There are two images, which are prepared for the left and right eyes separately, and these left and right images placed side by side are seen at the same time. To generate stereo vision, a stereoscope is used. The first stereoscope, which was invented by Sir Charles Wheatstone, use mirrors to refract the light, and the second stereoscope, which was developed and minimized by Sir David Brewster use prisms to refract light.

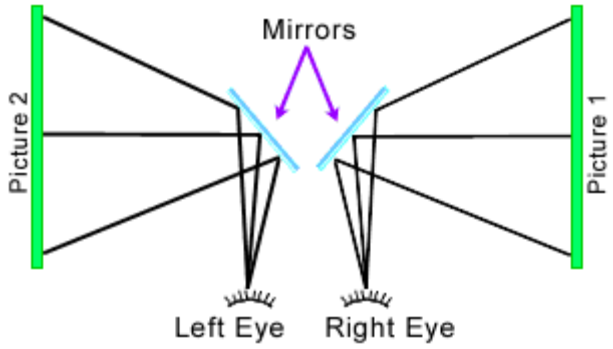

Figure 2.5 : Wheatstone's stereoscope [11]

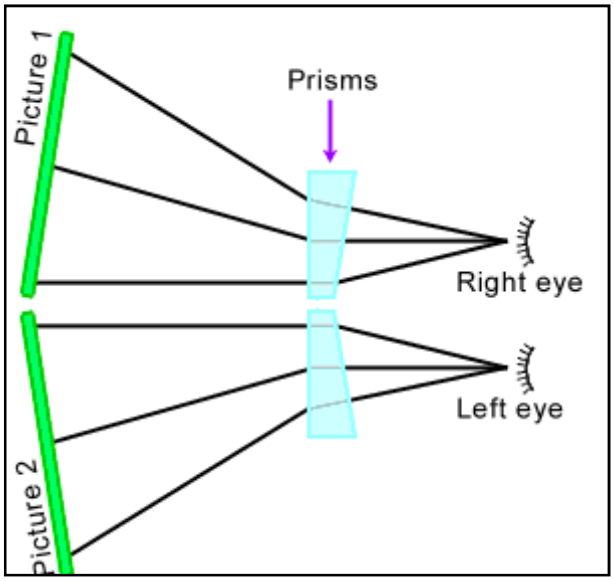

Figure 2.6 : Brewster's stereoscope [11].

Figures 2.5 and figure 2.6 show the stereo image of simulating 3D view by using stereoscopes.

The third way of viewing in 3D space is inspired by the stereoscope. There are four techniques, in order to see the left and right images simultaneously. Four different devices are used in these methods.

- ✓ Anaglyph glasses
- ✓ Active Shutter Glasses
- ✓ Polarized Glasses
- ✓ Auto Stereoscopic displays

Color Multiplex (Anaglyph 3D Vision): The observer wears a glasses where two lenses are different colors, such as red for left and cyan for right (chromatically opposite) [12]. Left and right images whose color content is specially processed for 3D sensation are superimposed, and because of the offset difference between left and right images, the display produces a depth effect to the viewers. Overall, using different color filters on the left and right images construct a 3D effect image. This method's display format is single frame colored.

Figure 2.7 shows an anaglyph image which is created by NASA during the exploration of Mars. It provides perception of the depth of land.

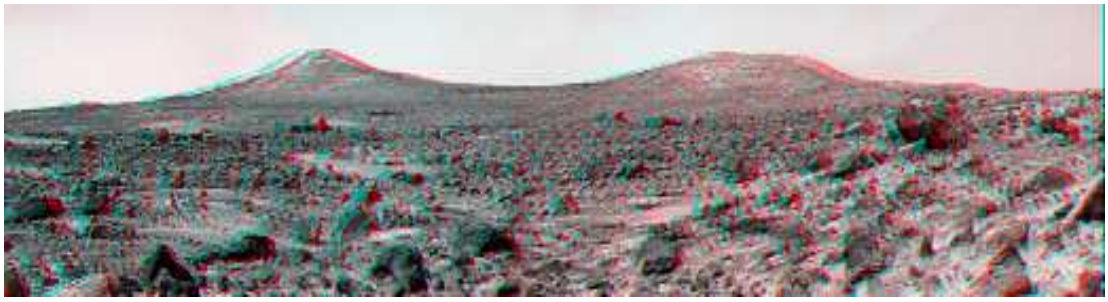

Figure 2.7 : Two color-coded anaglyph [12].

Time Multiplex (Liquid crystal shutter glasses): The left image for the left eye and the right image for the right eye are shown in sequence at a known screen refresh rate; according to the synchronization of the scene refresh rate, lenses of the glasses darken in an alternating sequence. In left image display, the left lens of the glasses gets transparent and the right lens gets opaque (vice versa for the other state). This method's display format is frame sequential and generally frame sequential synchronized glasses are used. Moreover, in HD format the 3D content can be broadcast as interlaced 1080 lines video signals in 30 fps. In this case, each eye will see progressive 540 lines. To view this alternating option, field synchronized sequential shutter glasses are used. Because of this reason, display format can be frame sequential or row/column interleaved. In addition, an infrared, radio frequency, or bluetooth transmitter sends a synchronization signal to control these special glasses. At the refresh rate, our brain will combine the left eye image and the right eye image. Because of this combination, this event produces a stereo effect.

Polarization Multiplex (Polarization): Left and right images pass through a polarized light plate; one of the images is polarized horizontally and the other one is polarized vertically. The observer wears passive polarized glasses where the left filter is horizontally polarized and the other one is vertically polarized. Each of these filters passes only the light which is polarized in its direction. This is the linear polarization method [13]. This way, left and right eyes see the left and right images simultaneously. Again, combining these two images gives a stereo image.

Spatial Multiplex (Autostereoscopy): In this method, the observer does not use a glasses or a filter. Here, left and right images are shown individually on the display. Examples of auto stereoscopic displays include parallax barrier, lenticular, volumetric, electro-holographic, and light field displays [13]. For example in the barrier display, the monitor with barriers will let the left eye and right eye see the different images at the same time.

Table 2.1: The classification of 3D display methods [15].

Viewing Technique		Multiplex. Method	Number of views	Advantages	Disadvantages
Stereoscopic Aided viewing	Passive Glasses	Color multiplex	2	Suitable with 2DTV	Color issue
				Used with DVD/Blu-ray	
				Cheap Colored glasses	Poor quality 3D
		Polarization multiplex	2	Good quality 2D	New 3D LCD TV set with micro-polarization sheet
				Good quality 3D	Vertical spatial resolution
				Low cost polarized glasses	reduced by half in 3D mode
	Active Glasses	Time multiplex	2	Good quality 3D	High end plasma or DLP
				Full quality 2D	New fast rate LCD (100/120 Hz)
				Low cost 3D glasses more expensive than the polarized glasses	Temporal resolution reduced by half in 3D mode
					Expensive glasses
Auto stereoscopic Free viewing	none	Spatial multiplex	>2(i.e., 5, 9...)	No glasses	Expensive new display
					Poor 3D experience for now
					Limited viewing angle
				"Coarse" motion parallax	Reduced horizontal resolution in 3D mode
					Degraded 2D viewing without 2D/3D switch

3D display methods are shown in Table 2.1 with their advantages and disadvantages.

2.3 Depth Perception in Electronic Stereoscopic Images

After the explanation of 3D viewing methods, the creation of depth sensation in electronic stereoscopic images will be described in this section.

In a stereoscope, photos, which belong to the same scene, are taken at two different viewpoints, and these differences in the viewpoints create a disparity in images. When the observer sees the two images at the same time, the image disparity creates a retinal disparity. However, this created retinal disparity is not identical to the natural retinal disparity that occurs during natural vision. Because of this reason, the created 3D scene does not give a natural 3D scene sensation [5].

The other important thing for the depth perception in a planar stereoscopic display is the viewing geometry and the formation of the generation of depth perception with parallax situations [5]. The screen disparity as can be seen in figure 2.9 creates a 3D effect illusion. In the case of positive parallax, the viewer will think that the object is behind the screen and in the case of negative parallax; the viewer will think that the object is in front of the screen. In zero parallax, the object seems that it stands on the screen. In either case, eyes focus to the screen distance. With the use of these parallax situations, a 3D illusion effect is created.

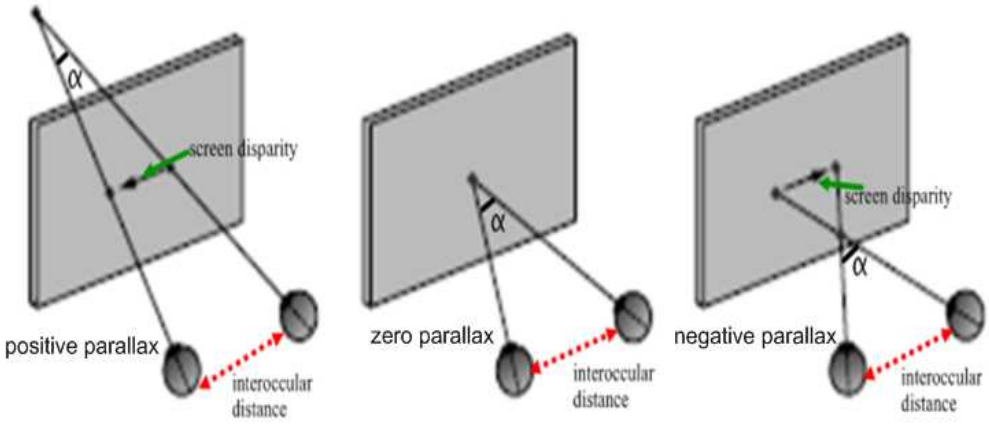

Figure 2.8 : Types of parallax.

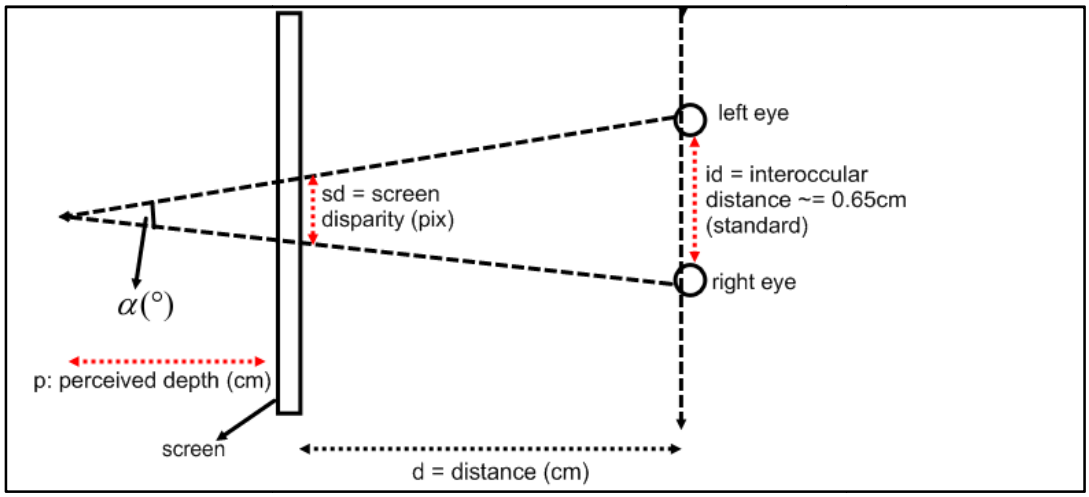

Figure 2.9 : Positive parallax angle (α).

p: Perceived depth.

id: The distance between two eyes, in other words the interoccular distance, Children's interoccular distance is around 55 mm and adult's interoccular distance is 75 mm maximum. As a result of the measurements, the average interoccular distance is taken as 65 mm [8, 9].

d: The distance between the viewer and the display.

sd: Screen disparity is the difference between left and right images, and this variable is dependent on the size of the screen.

In positive parallax, for positive values for *sd*, the perceived depth indicates on equation 2.1e.

$$\tan\left(\frac{\alpha}{2}\right)=\frac{sd}{2p}=\frac{id}{2(p+d)} \qquad \textbf{(2.1a)}$$

$$\frac{p}{sd}=\frac{p+d}{id} \qquad \textbf{(2.1b)}$$

$$\frac{p+d}{p}=\frac{id}{sd} \qquad \textbf{(2.1c)}$$

$$\frac{d}{p}=\frac{id}{sd}-1 \qquad \textbf{(2.1d)}$$

$$p = \frac{d}{\left(\frac{id}{|sd|}\right) - 1} \qquad \textbf{(2.1e)}$$

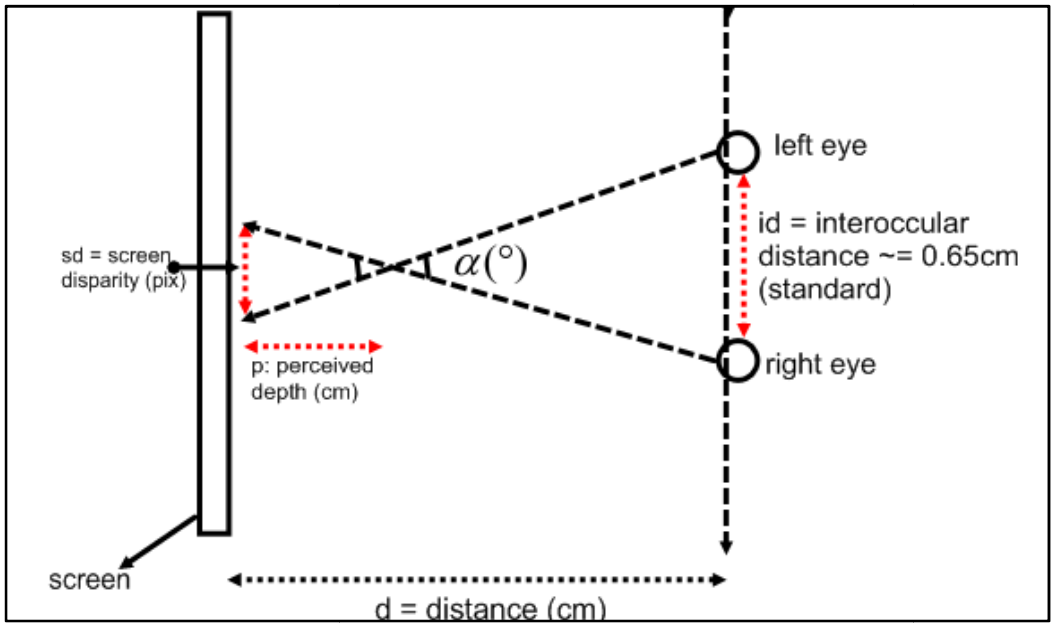

Figure 2.10 : Negative parallax angle (α).

In negative parallax, for a negative value for *sd*, the perceived depth indicates on equation 2.2e:

$$\tan\left(\frac{\alpha}{2}\right) = \frac{sd}{2p} = \frac{id}{2(p-d)} \qquad \textbf{(2.2a)}$$

$$\frac{p}{sd} = \frac{p-d}{id} \qquad \textbf{(2.2b)}$$

$$\frac{p-d}{p} = \frac{id}{sd} \qquad \textbf{(2.2c)}$$

$$\frac{d}{p} = \frac{id}{sd} + 1 \qquad \textbf{(2.2d)}$$

$$p = \frac{d}{\left(\frac{id}{|sd|}\right) + 1} \qquad \textbf{(2.2e)}$$

From these equations, the perceived depth is directly proportional to the viewing distance *d*. Observers who are in different viewpoints have different depth sensations. In addition, perceived depth is directly proportional to the screen disparity *sd* and screen disparity changes with the size of the screen. According to the size of the display, varying depth perceptions can be achieved. Moreover, the third variable, the interoccular distance *id* is inversely proportional to the perceived depth *p*. Because of this inverse proportion, children who have smaller interoccular distance than average will have more perceived depth than adults will.

Against uncomfortable viewing experience, the angle of parallax will be taken below 1.5° [14]. According to this constraint, the magnitude of the screen disparity and the viewing distance will be changed.

The magnitude of the perceived depth will be smaller than the distance between the screen and the viewer. Because of this reason, to calculate the screen disparity, the viewing geometry may be approximated by equation 2.3.

$$d \cong d + p \tag{2.3}$$

With this approximation, the parallax angle equation can be rewritten as in equation 2.4.

$$\alpha \cong 2\arctan\frac{sd}{2d} \tag{2.4}$$

2.4 The Automatic Stereoscopic Conversion Algorithm and Related Works

2D videos obtained from satellite broadcasts, cable TV, and DVD/Blu-ray players, can converted into stereoscopic image sequences by using 2D to 3D conversion technique and by using a suitable 3D display device.

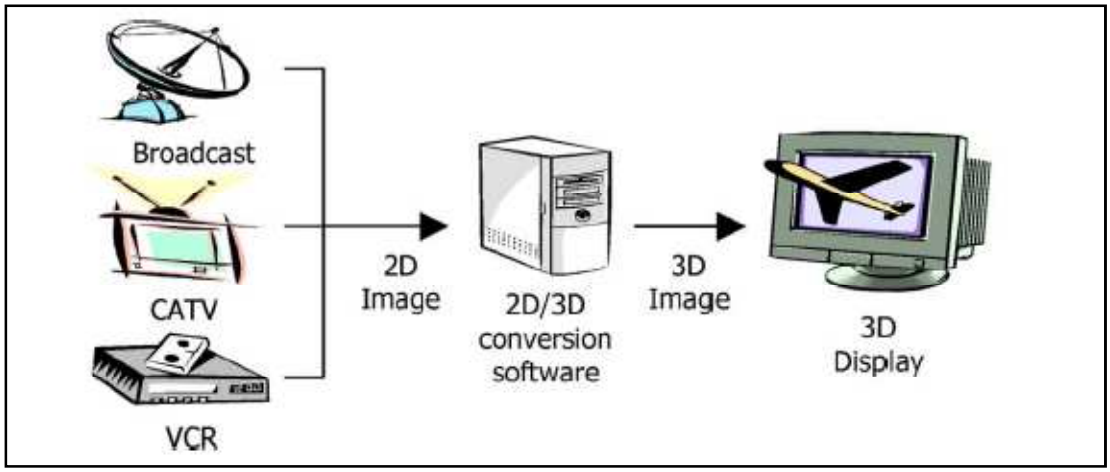

Figure 2.11 : Overview of the stereoscopic conversion system [16].

A general stereoscopic conversion system is shown in Figure 2.11. The generalized approach of 2D-3D conversion algorithm contains the following steps:

- Conventional 2D Video
- Depth Map Formation
- Synthesis of Stereo Frames
- Stereo Image Projection (according to the display device)

Conventional 2D video is the input of the conversion algorithm, and it is obtained from monoscopic captured still image sequences. The depth map determines the position of the object in the scene. The main difference between a 2D image and a 3D image is the depth information, and the depth map is a key factor for 3D perception quality. After the depth map generation, with the help of this information, a stereoscopic image pair is formed via shifting pixels or using a depth image based rendering way. Finally, according to the selected display device, stereo image projection will be done. In addition to this projection, selected display device will influence the synthesis stereo image pair, because for example it may require more than two views for auto-stereoscopic displays.

The first commercial 2D to 3D image conversion TV was developed using the Modified Time Difference (MTD) method [17]. This method takes the N^{th} frame of the 2D video as the left image and $(N-2)^{th}$ image as the right image when object moves to the right direction. Due to movement in the right direction, the developed conversion algorithm will fail in stationary scenes and complex scenes that include camera motion, or at high-speed motion or at slow motion. In order to solve this

problem, at scenes with high-speed motion, $(N-1)^{th}$ frame was taken as the right image, and at slow motion scenes $(N-5)^{th}$ frame was taken as the left image. Moreover, according to the velocity of the frame, the time-delayed image was selected as the left or the right image.

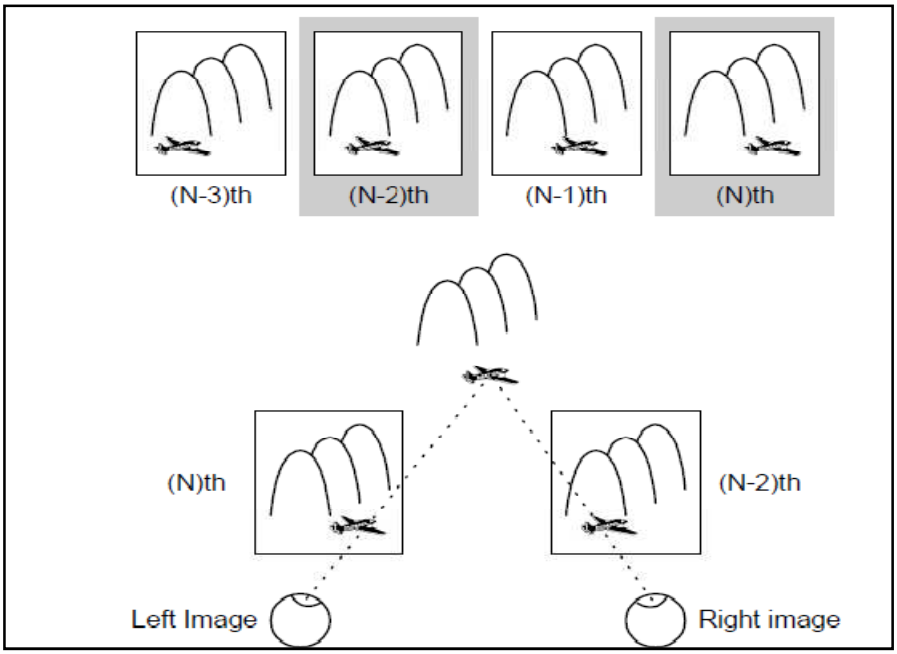

Figure 2.12 : Principle of the MTD [14].

However, in these methods there is a lack of information for complicated scenes and for stationary scenes. Because of this reason, the depth information is required for these problematic scenes. In fact, the previous MTD method uses a depth map, and it does not require any special computational method. At a horizontal motion, the time delay creates a horizontal disparity between frames due to the motion, and the motion dependent algorithm will fail at stationary scenes. Moreover, there are many depth map generation algorithms which retrieve the depth map from 2D videos. These depth extraction algorithms use various monocular depth cues. For example, in [18], a novel architecture is developed via extracting depth information using image segmentation based on independent component analysis (ICA) method, and it uses a smoothing filter for occlusions that are formed during the synthesis of stereo image pair. By using the blur information and the optical flow motion information of the 2D video, it can generate a depth map, because usually the focused pixels are foreground objects and vice versa [19]. In other methods, using color segmentation and motion difference information obtained via a variety of motion algorithms, a depth map can be generated [20, 22, 23, 24, 25, and 26]. In addition, the depth map retrieved from

the geometric perspective and from edge information [21, 28, and 29]. The different depth extraction methods that come from monocular depth cues can be used together in the step of depth map extraction [27].

Philips has developed algorithms that derive a depth map for each video frame automatically. The method is based on static depth cues such as color and texture, and on dynamic depth cues such as frame difference or motion [35].

After the depth map extraction, the depth information and the 2D monocular video must be synthesized to obtain a stereo image pair. There are two ways to construct the second image. One of them is the shift algorithm, which based on the relativity between binocular vision and the image depth, and the other one is the depth image-rendering algorithm, which is dependent on the camera parameters [29, 30].

2.5 The Proposed 2D to 3D Conversion System

In this section, the general information about the proposed 2D to 3D conversion system will be given.

First, to visualize 3D perception, anaglyph is the cheapest stereoscopic projection technique. Only color filters are used to view the 3D display. It can also be implemented at television sets and computer screens without the need of complicated hardware, such as special screens, controlled shutter glasses. Because of this reason, the anaglyph method was selected to test the results of work done in this thesis for a 3D effect. The created results can easily be applied to other display techniques by replacing the anaglyph creation step.

For the synthesis of the stereo-image pair, the proposed system needs to extract the depth map. First, the block matching method will be used for creating a depth map based on motion vectors. The edge detection will be used as a second way to obtain the depth map.

For the synthesis of stereo pair image generation, the shift algorithm will be used.

Figure 2.13 shows the general scheme of two implementations. The only difference between the two methods is at the step of the depth map formation; remaining steps are common to both.

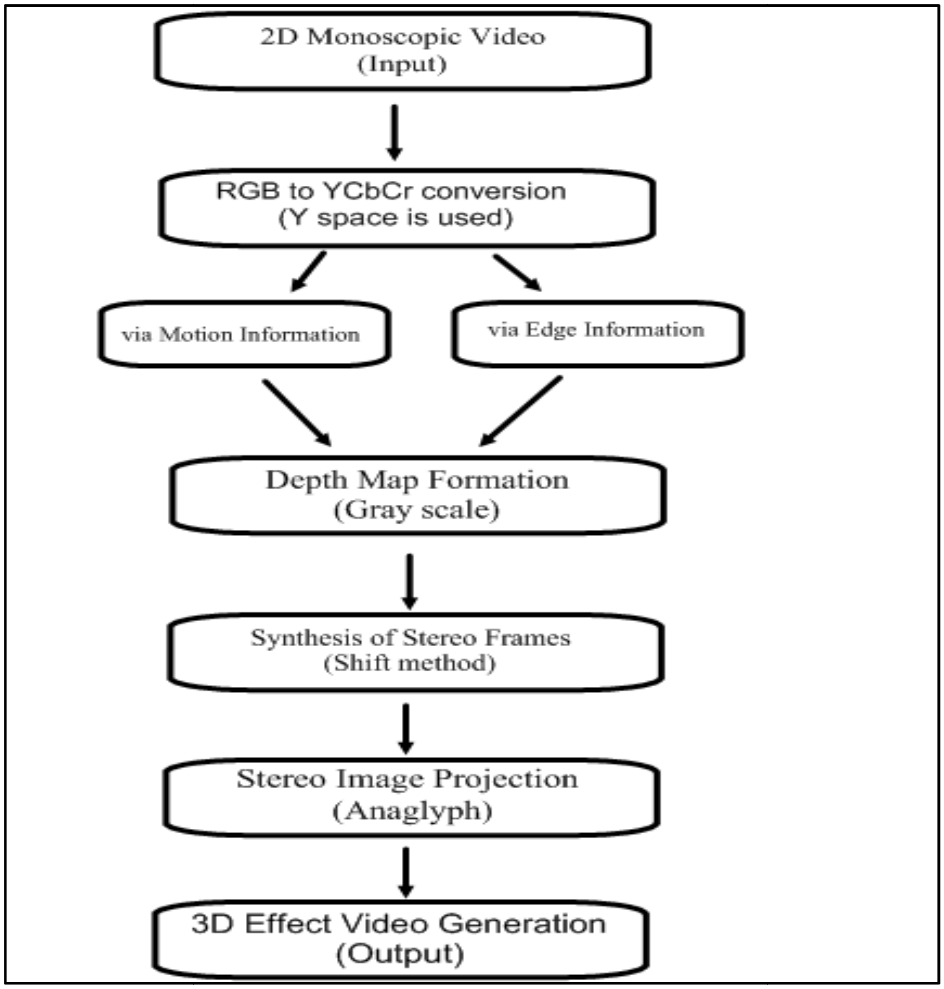

Figure 2.13 : The general scheme of the implementation.

3. THE DEPTH ESTIMATION FROM 2D VIDEO

3.1 Objectives

The objective of this section is to extract and analyze depth information in 2D images. The depth information is used to convert a single 2D image into a 3D effect version. The mapped depth information can be represented via a gray-level image where each gray level is related to a depth value. The depth map allows calculating the amount of shift for each pixel of an image. In other words, the first part is to process the image to get a depth map, and the second part is to use the depth map to generate left and right eye images.

3.2 The Depth Map Generation via Motion Vectors

3.2.1 Motion vector extraction

Stereoscopic video generation using the magnitude of motion vectors assigns depth to moving objects.

Instead of using optical flow method results [19] to synthesize a depth map, the block-matching method, which is a popular method for practical motion estimation due to its lesser hardware complexity as compared to the optical flow methods [34], can be considered.

For extracting motion vectors, motion estimation via the block matching method was used. This algorithm attempts to divide both the previous and current images into square blocks and then compute the motion for these blocks. Obtained motion vectors are assigned to their corresponding blocks. In other words, each motion of the block is represented by its motion vector. Because of this reason, the size of the storage of motion vectors is lower than the original size of the video.

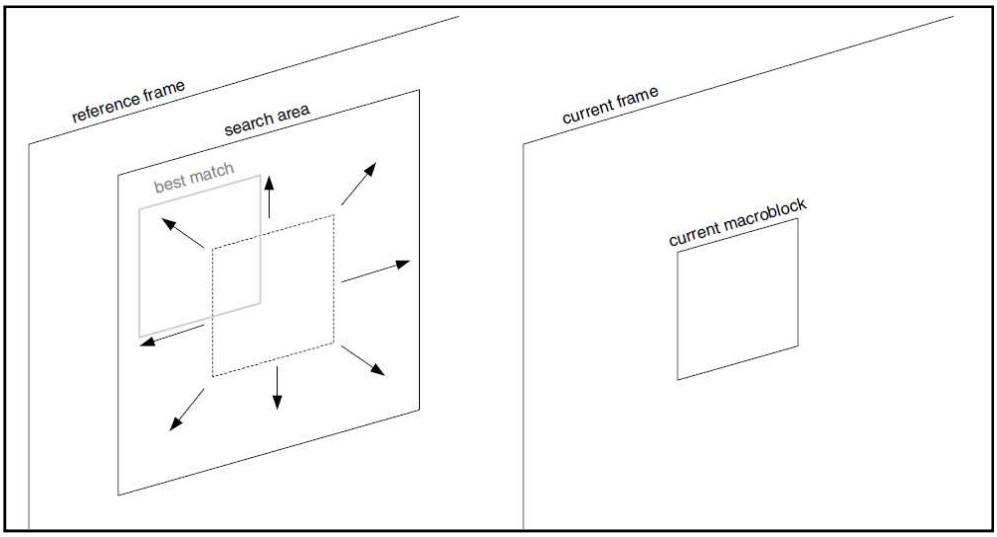

Figure 3.1: Motion estimation with block matching method [39].

For obtaining motion vectors from the video sequence, OpenCV V.1.0 block matching function was used. The implementation in OpenCV uses a spiral full search that works out from the location of the original block (in the previous frame) and compares the candidate new blocks with the original.

As in figure 3.1, the difference between the current block and a set of neighboring regions in the reference frame is calculated. The region that gives the lowest error is selected. The horizontal and vertical position difference gives motion information. This difference calculation process is based on the sum of absolute differences (SAD) of the pixels. If a good enough match is found, the search is terminated. If there is no motion, there is a high correlation between the two blocks at their original position.

Because more search operations increase complexity, the block size is taken as 16x16. After motion estimation operation, the obtained horizontal and vertical vectors are stored in an array. Motion vector of each block is assigned to all pixels of the particular block. In this way, every pixel of the frame will have horizontal and vertical vectors. Moreover, a similar method and approximation is used in literature [23], and for gaining knowledge about the expected results, the results of this study are observed subjectively.

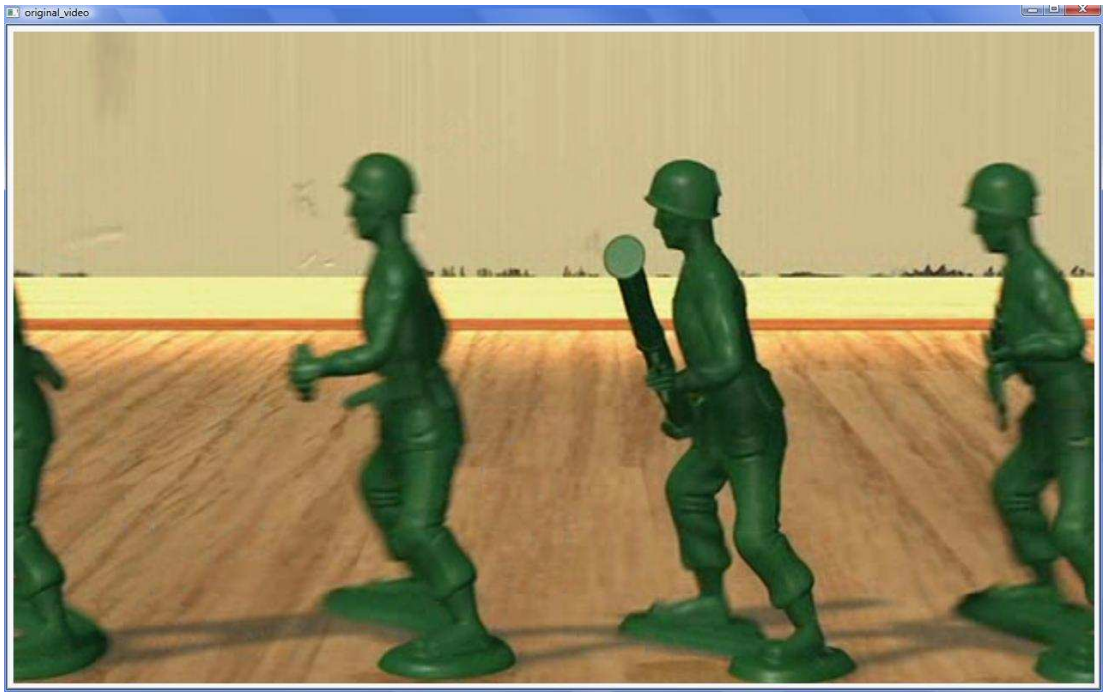

Figure 3.2: The reference frame.

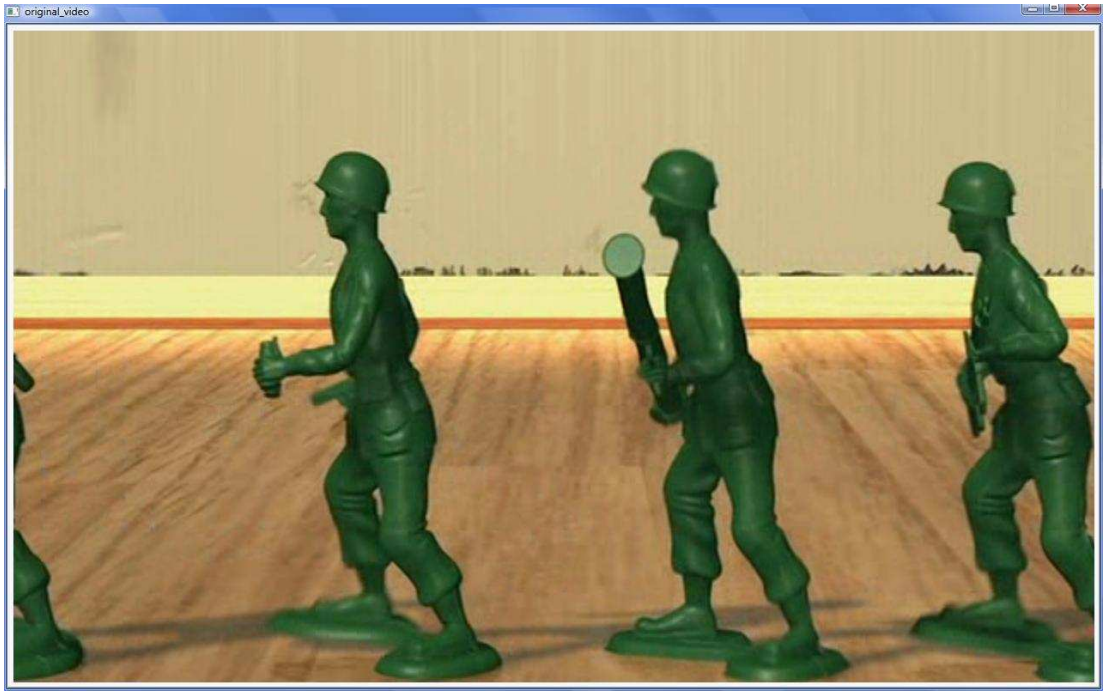

Figure 3.3: The current frame

Figures 3.2 and figure 3.3 show examples of reference (previous) frame and current frame. These frames are from the Toy Story 1 movie. A part of this movie named as the “Soldiers” video sequence and it is converted to its stereo version. In addition, original Toy Story 1 movie is converted to stereo in anaglyph format.

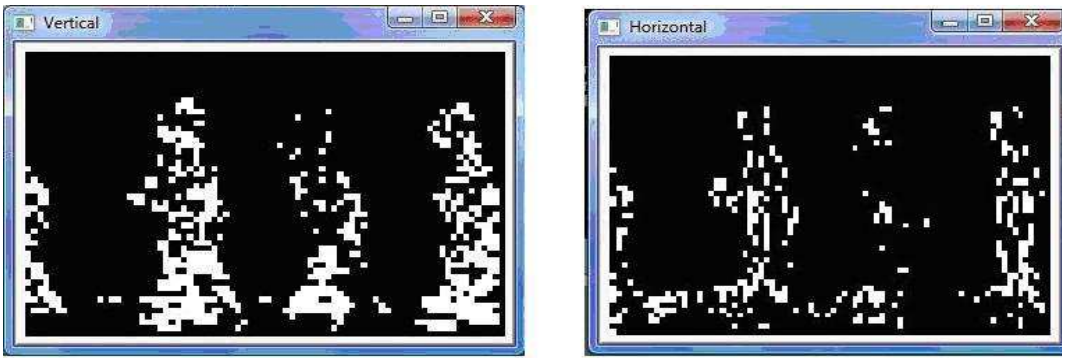

Figure 3.4: Horizontal and vertical velocity vector positions.

Figure 3.4 shows horizontal and vertical velocity images, which are outputs of the used block matching function. These images are in the binary format and they show locations of moving pixels between reference and current frame. Because each motion vector represents a moving block, the resolution of these images is lower than the original frame resolution by the block size, i.e., 1/16 in this work. These images are used only for the visualization of the motion. The obtained location information of motion and the calculated motion vector information are used to synthesize the depth map.

3.2.2 Transformation of Motion Vector Maps to Depth Maps

In this implementation, motion vectors are taken as a unique depth cue. For this reason, horizontal and vertical motion vector values used to calculate the depth information.

$$D(i,j) = 255\frac{\sqrt{MV(i,j)_x^2 + MV(i,j)_y^2}}{\max(\sqrt{MV(i,j)_x^2 + MV(i,j)_y^2})} \quad \textbf{(3.1)}$$

Here, D(i,j) represents the depth map pixel according to the screen. $MV(i,j)_x$ and $MV(i,j)_y$ are horizontal and vertical velocity vector values. As the equation implies, to scale the depth information between 0 and 255 and for representing it via a gray level image, the magnitude of motion vectors is normalized to unity, and the normalized results are multiplied by 255 for 8-bit scale. In this way, the maximum motion determines the maximum depth. Further region of interest is presented with a “0” gray value, and closer region of interest is presented with a “255” gray value; 128 corresponds to the depth position corresponding to the screen.

```
FOR all frames on the video
        ......
        take current frame
        take previous frame
        convert RGB frame to YCbCr frame (current&previous)
        take Y channel (current&previous)
        (velocity_x, velocity_y)=calculate horizontal and vertical motion vectors
                                  from the difference between current and previous frame
        resize velocity_x and velocity_y arrays to the size of original frame
        // transformation motion vectors to depth map
        FOR all columns on the frame
                FOR all rows on the frame
                        magnitude of velocity = sqrt (velocity_x*velocity_x+velocity_y*velocity_y)
                ENDFOR
        ENDFOR
        FOR all columns on the frame
                FOR all rows on the frame
                        normalized magnitude of velocity = magnitude of velocity / maximum magnitude of velocity
                        depth_map = normalized magnitude of velocity*255;
                ENDFOR
        ENDFOR
        filter the depth map with a low pass filter for reducing the blocking artifacts
        ....
ENDFOR
```

Figure 3.5: Pseudo code of depth map via motion information.

Moreover, in order to reduce the blocking artifacts, a smoothing operation, which takes a simple mean of all the depth pixels in a (11x11) window around the corresponding pixel in the input, is applied to the depth map image.

In figure 3.6, the moving pixels represent close objects and static pixels are counted as distant objects of the image, and the magnitude of motion vectors determines the closeness to the camera of the object. Moreover, to create an artificial 3D monocular depth cue, distant regions, i.e., image pixels corresponding to dark areas in figure 3.6, will be blurred; this blur is different than the depth map low pass filtering.

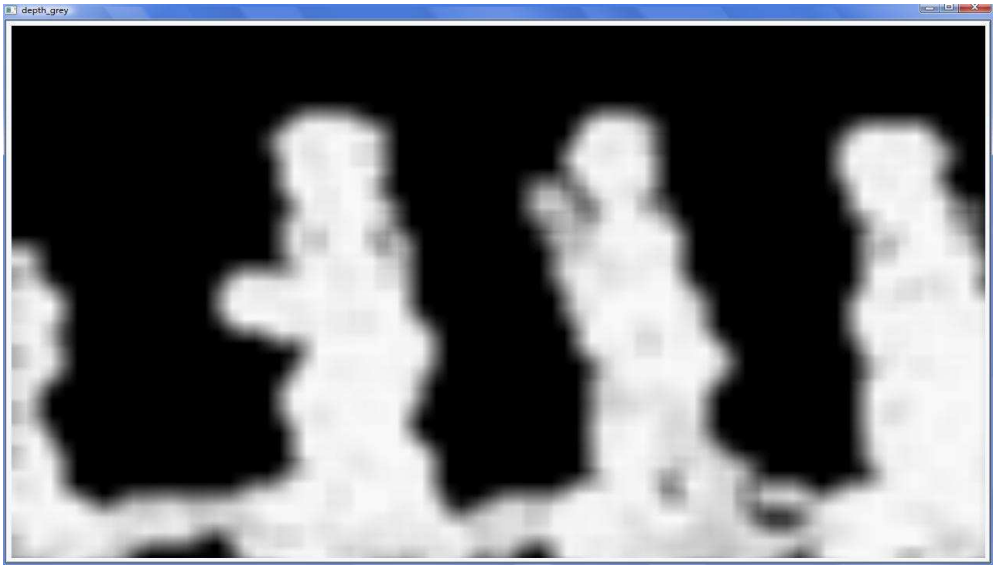

Figure 3.6: The depth map of the reference frame

3.3 The Depth Map Generation via using Edge Information

3.3.1 Edge information

The second way to obtain the depth map from a 2D image is based on the use of the edge information.

An edge is the boundary between two regions with relatively distinct gray-level properties. For the extraction of edges, the Canny edge detection algorithm is used and this method is known to many as the optimal edge detector [36]. The Intel OpenCV V.1.0 library is used for the Canny edge detection algorithm. The implementation uses two thresholds. If a pixel has a gradient larger than the upper threshold, then it is accepted as an edge pixel; if a pixel is below the lower threshold, it is rejected. If the pixel's gradient is between the thresholds, then it will be accepted only if it is connected to a pixel that is above the high threshold. Lower threshold value is set to 10 and the upper threshold value is set to 100; the applicable range is 1 to 255 for the threshold.

Figure 3.7: The edge map of the original frame.

Figure 3.7 shows the edge map of the current frame of figure 3.2 after Canny edge detection.

3.3.2 Transformation of edge information to depth maps

First, the found edges are dilated to preserve edges. Nonzero pixels of the obtained binary image are counted as focused regions and zero valued pixels are counted as unfocused regions.

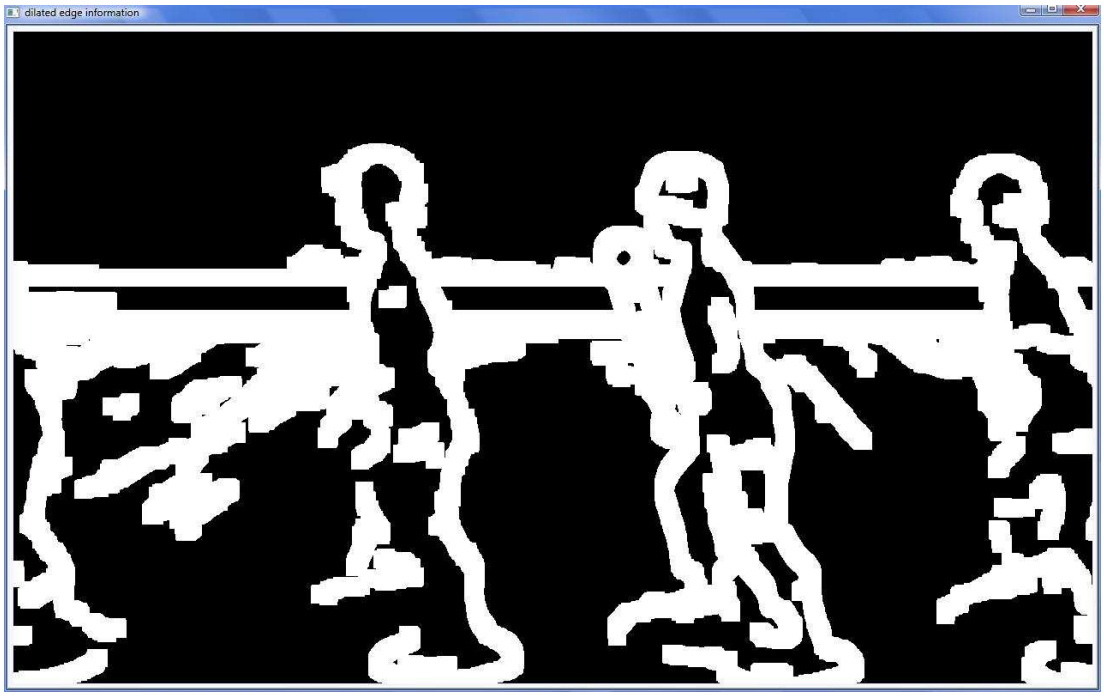

Figure 3.8: The dilated edge map.

In addition, the magnitude of the gradients, which is used as “edge strength”, defines the depth information. To find the edge strength, it is required to calculate the gradient of the image. The Canny Edge Detection function of OpenCV V.1.0 used in the implementation does not return the gradient information. For this reason, the gradient operation is done separately using the Sobel operator. The Sobel operator uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows), as shown in figure 3.9.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 3.9: Horizontal and vertical filters.

With the help of convolution with these filters, horizontal and vertical gradients obtained. The edge strength of the gradient is approximated using equation 3.2 where G_x and G_y denote the horizontal and vertical gradients, respectively.

$$|G| = |G_x| + |G_y| \quad \textbf{(3.2)}$$

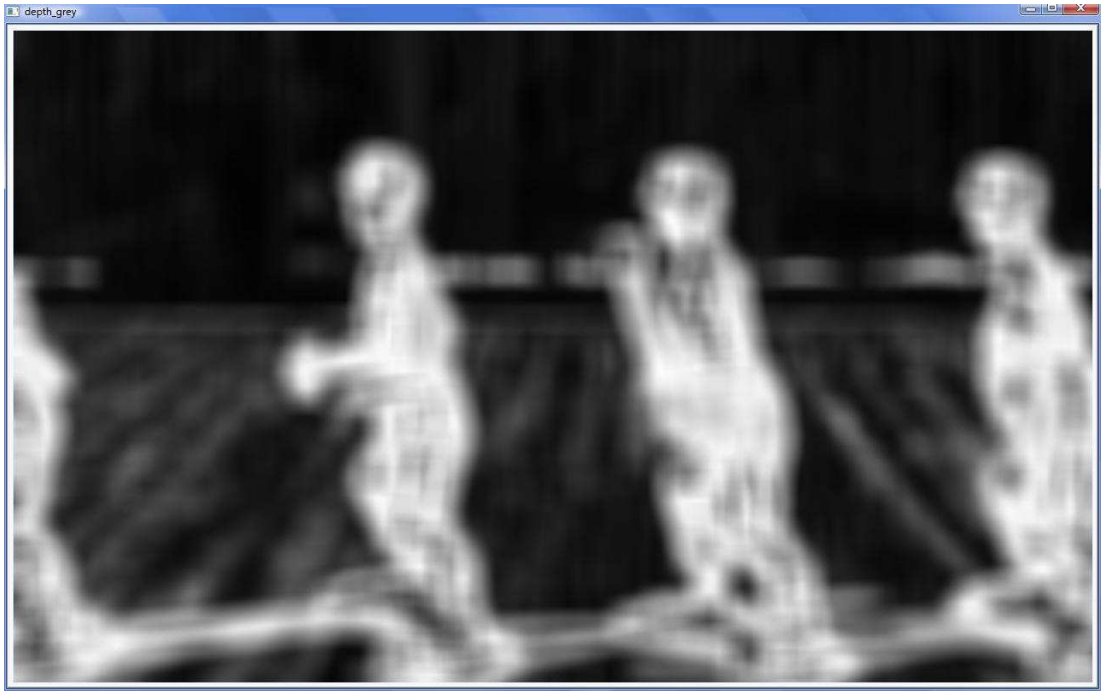

Figure 3.10: The depth map.

Figure 3.10 shows the depth map obtained for the original frame given in Figure 3.2. The depth information in the pixels of Fig. 3.10 is the magnitude of gradients of the image. Distant region of interest is presented with “0” gray value, and nearby region of interest is presented with “255” gray value.

```
....
horizontal filter = {-1,0,1;-2,0,2;-1,0,1}
vertical filter = {1,2,1;0,0;-1,-2,-1}
FOR all frames on the video
        ....
        take current frame
        convert RGB frame to YCbCr frame
        take Y channel
        //depth map
        gradient_x = convolve gray image with horizontal filter
        gradient_y = convolve gray image with vertical filter
        FOR all columns on the frame
                FOR all rows on the frame
                        depth_map = sqrt (gradient_x*gradient_x+gradient_y*gradient_y)
                ENDFOR
        ENDFOR
        ....
        ....
ENDFOR
```

Figure 3.11: Pseudo code of depth map via edge strength information.

We can also interpret the situation as follows: High frequencies of the image associated with high gradients are accepted as near regions of interest, and low frequencies associated with low gradients are accepted as distant regions. To create an artificial 3D monocular depth cue, distant regions are blurred, and with the help of the obvious dilated edge information, the edges are preserved.

4. SYNTHESIS OF ARTIFICIAL STEREO VIDEO FRAMES

4.1 Image Enhancement for 3D Effect Visualization

In typical photographs, the near-positioned objects have higher focus than the far-positioned objects and the background image. Therefore, these focus values are inversely proportional to the distance from the camera to the objects. Therefore, blurriness may also take as a depth cue. Distant objects are expected more blurred because the far positioned objects have a higher blur than the focused region of the image. Of course, this assumption is only valid when the camera is focused on a relatively closer object in front of a distant background.

According to the results of Canny edge detection and dilation, nearby region of interest and distant region of interest are segmented. In addition, in the depth map implementation via motion information the moving pixels represent nearby objects and other pixels counted as distant objects of the image. Due to reasons of human depth perception mentioned earlier, the distant region of interest will be smoothed with a simple (9x9) blur filter.

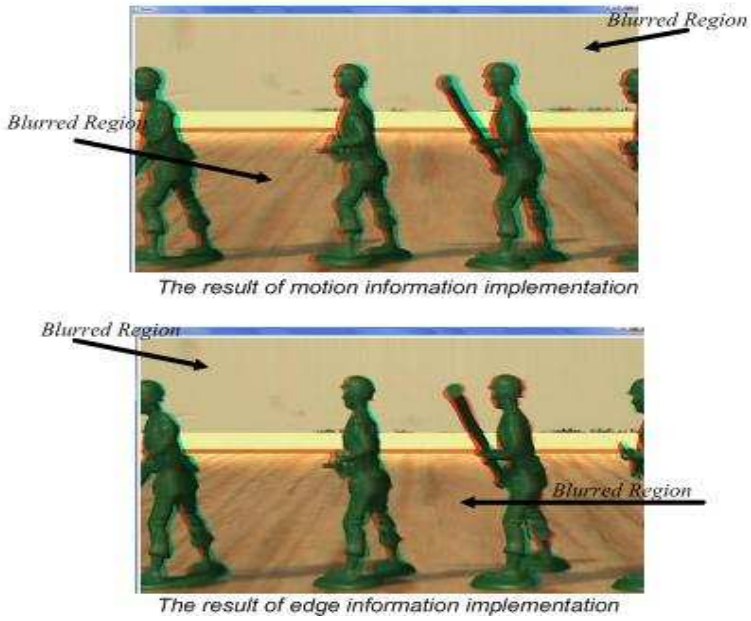

Figure 4.1: Blurred distant region of interest.

Figure 4.1 shows the visual degradation of the image quality. In the implementations of the thesis, only the distant regions are blurred.

```
FOR all frames on the video
.....
take current frame
convert RGB frame to YCbCr frame
take Y channel
create edge map   or  create horizontal and vertical motion vectors
dilate edge map     or calculate the magnitude of motion vectors
.......
blur_frame = smooth all of the original frame
// artificially blurring
        FOR all columns on the frame
                FOR all rows on the frame
                        IF dilated_edge_map pixel is equal to zero THEN
                                it is distant pixel (or unfocused pixel)
                                original_red_channel_pixel = blurred_red_channel_pixel
                                original_green_channel_pixel = blurred_green_channel_pixel
                                original_blue_channel_pixel = blurred_blue_channel_pixel
                        ENDIF
                    or  IF mag_vel_map pixel is equal to zero THEN
                                it is distant pixel (unfocused pixel)
                                original_red_channel_pixel = blurred_red_channel_pixel
                                original_green_channel_pixel = blurred_green_channel_pixel
                                original_blue_channel_pixel = blurred_blue_channel_pixel
                        ENDIF
                ENDFOR
        ENDFOR
.....
ENDFOR
```

Figure 4.2: Pseudo code of decision for blur.

4.2 The Formation of the Stereo Image Pair

For the formation of a stereo image pair, the depth map of the image and the hardware parameters of the 3D display are required. A stereoscopic display is one that differs from a planar display in only one respect: It is able to display parallax values of the image points. Parallax produces disparity in the eyes, thus providing the stereoscopic cue.

Objects with negative parallax appear to be closer than the plane of the screen, or between the viewer and the screen and objects with positive parallax appear behind the screen to the viewer.

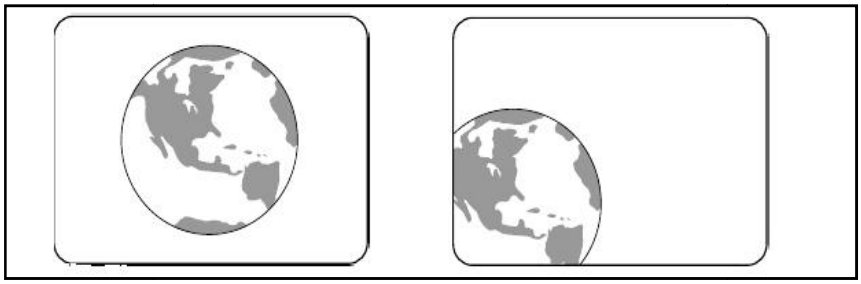

Figure 4.3: The disadvantage of the negative parallax.

The first drawing shows the globe, in viewer space with negative parallax, centered within the screen surround. There will be no conflict of cues. However, the second drawing shows the globe cut off by the surround. In this case, the interposition cue tells the eye-brain the globe is behind the surround, but the stereopsis cue tells the eye-brain that the globe is in front of the surround. It is a conflict of depth due to the object being behind the window hence negative parallax is risky.

The maximum amount of parallax value must also be determined. This is the number of pixels difference between the nearest and the furthest distant point. It depends on the screen width in pixels, the distance between the screen and the viewer, the image resolution, and the parallax angle. As a rule, parallax angles cannot be more than 1.5°; it tends to make viewing uncomfortable. The distance between the viewer and the screen must be at least approximately ~18 inch, ~46 cm [38]. Parallax angle is defined in equation 2.4; this formula is used to calculate maximum screen disparity.
In this thesis, 3D video will be displayed at full screen; the frame to be displayed should have the same dimension and resolution of the monitor.

```
calculate_shift_max  input:width&height
BEGIN
        screen_width=39.116;//cm
        pi = 3.14159265;
        pixel_width = screen_width/width;  // cm/pix
        parallax_angle = 0.75;//degree
        distance = 50;//cm
        disparity = 2*distance*tan((parallax_angle/2)*(pi/180));
        shift_max = disparity/pixel_width;
        return shift_max;  //output: shiftmax
ENDBEGIN
```

Figure 4.4: The pseudo code of maximum shift calculation.

After the maximum value of the shift is decided, then the amount of shift in pixels is direct proportion to the depth value. Also, according to the depth information, which object is nearer or further can be determined. With this information, the shift algorithm creates two images: left eye and right eye views. The depth values change between 0 and 255, and for the formation of stereo image pair, the following pixel shift equation is applied.

$$p(i, j) = p_{\max}\left(\frac{d(i, j)}{255} - \frac{1}{2}\right) \quad \textbf{(5.1)}$$

d(i,j): depth value

p_{max} = maximum pixel shift

p(i,j) = calculated pixel shift value for every pixel in the image.

In Eq. 5.1, zero pixel shifts corresponds that object is on the screen, and both types of parallax allowed.

```
FOR all frames on the video
        ......
        take current frame
        shift_max = calculate maximum shift
        FOR all columns on the frame
                FOR all rows on the frame
                shift = (shift_max*(depth_map/255)-0.5)
                shift R,G,B channels according to the amount of shift
                ENDFOR
        ENDFOR
        ....
ENDFOR
```

Figure 4.5: The pseudo code of shift method.

4.3 The Anaglyph Projection

Anaglyph is a method to view stereoscopic images using colored filters. The anaglyph method is used since 1853 but it was patented in 1891 by Louis Ducos du Hauron [32].

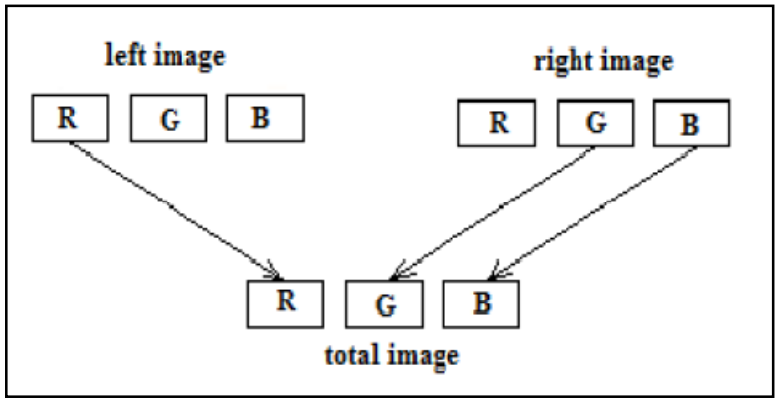

Figure 4.6: The synthesis of anaglyphs [33].

This old method is easily accessible and is the cheapest way to obtain a 3D visualization effect. Synthesis of anaglyphs is a simple process in which the red channel in one image replaces the red channel of the second image of the stereo pair in constructing the RGB anaglyph image. As shown in figure 4.6, the red component corresponds to the red in the left image and the blue and green component correspond the right image. In this way, a new stereoscopic image can be obtained and projected. In addition, for viewing these anaglyphs, special red cyan glasses are needed.

5. EXPERIMENTAL RESULTS

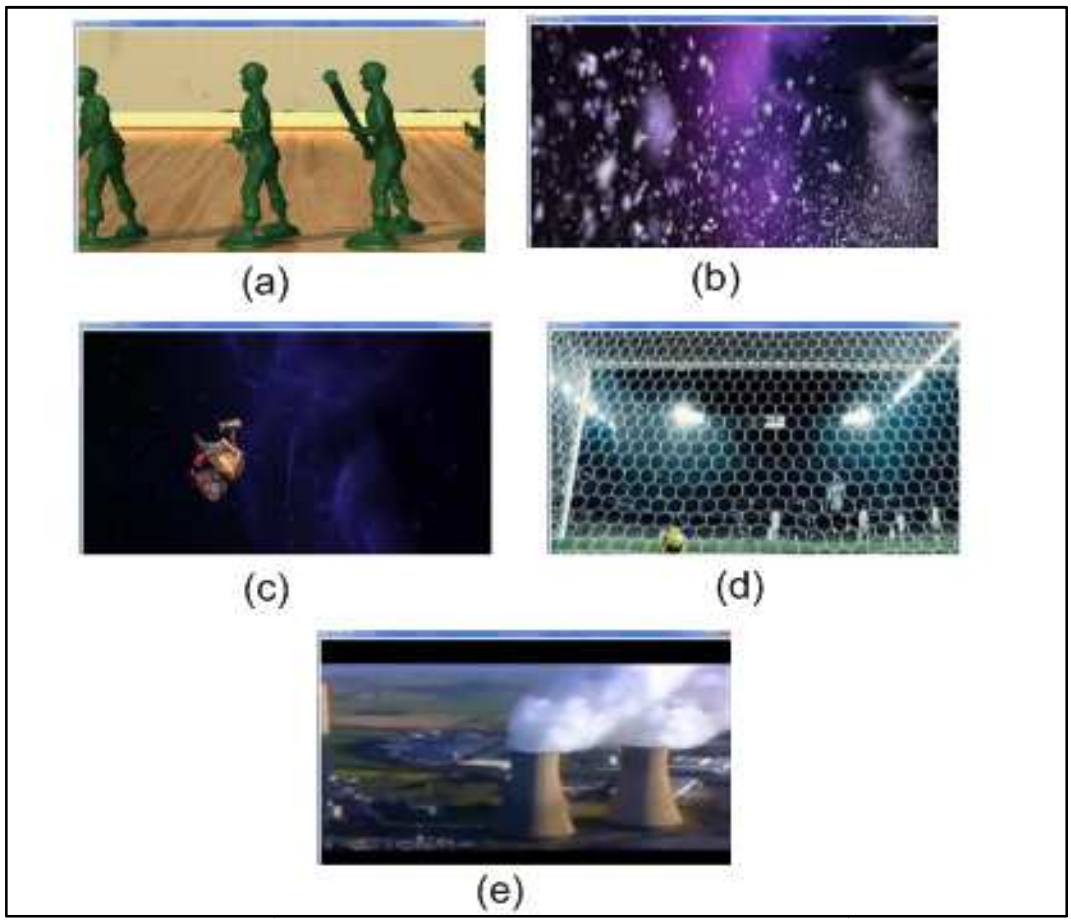

Figure 5.1: Test sequence sets.

In order to evaluate the proposed algorithms, five sequences were used and tested.

a- Soldiers video sequence.

b- Scattering video sequence.

c- The Robot video sequence.

d- Football video sequence.

e- The Real Simpsons video sequence.

All videos resized to 1280x800 resolution. “Soldiers” video contains 52 frames, its original resolution is 640x352, and its frame rate is 25 fps. “Soldiers” video has motion in a right to left direction, and there is camera movement. “Scattering” video contains 20 frames, its original resolution is 688x284 and its frame rate is 23.98 fps. “Football” video contains 20 frames and its frame rate is 25 fps. “The Robot” video contains 122 frames, its original resolution is 688x284, and its frame rate is 23 fps. “Football” and “Scattering” videos have wavy motion and scattered movements, respectively. “The Robot” video has motion in a left to right direction, and there is camera movement. “The Real Simpsons” video contains 3000 frames, its original resolution is 320x240 and its frame rate is 54 fps. In addition, “The Real Simpsons” video contains panning and zooming.

First, depth maps are calculated from frames of the original sequences via motion estimation and edge detection algorithms. Then, their parallax values are calculated according to the resolution and the depth map. In this way, artificial left and right stereo image pairs are generated. An Intel Pentium Dual Core T4200 PC running at 2.00GHz with a 15.4 inch 1280x800 LCD screen was used with red-cyan glasses. Algorithms are implemented using the C/C++ language in a Visual Studio 2005 environment, with the help of OpenCV V.1.0 library.

The distance between the viewer and the screen is taken as 50 cm, and the screen size is 15.4”, in other words 39.116cm. The display resolution and the video resolution are the same i.e., 1280 x 800. The constant pixel width is ~0.031 cm. and the maximum positive parallax angle is taken as 0.75°. From the parallax angle equation, the disparity value is 0.65 cm. and the maximum amount of pixel shift is 10 pixels from the pixel shift equation.

Following figures include an original scene and its stereoscopic scene version. The 3D effect from anaglyph images can be seen through red-cyan anaglyph glasses; the color print quality may affect the quality, however, the test sequences are also included in the attached CDROM.

Figure 5.2: Original frame (a).

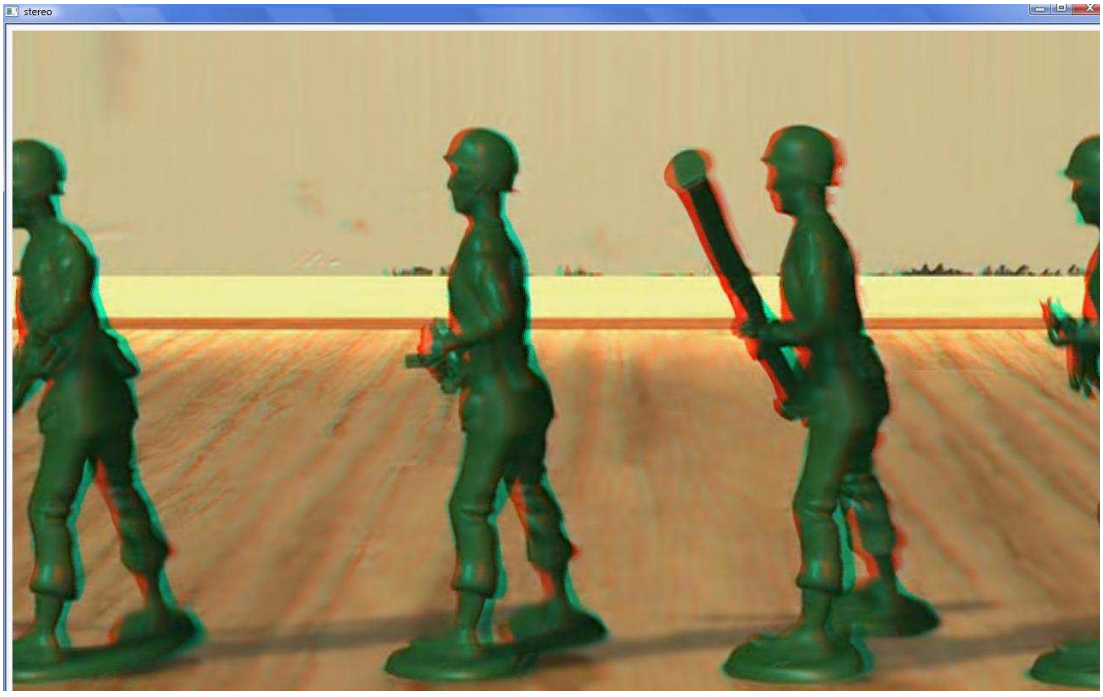

Figure 5.3: Stereoscopic video frame via edge information (a).

Figure 5.2 is the original version of figures 5.3 and 5.4. These frames belong to the "Soldiers" video sequence. "Soldiers" video sequence is from Toy Story 1 movie.

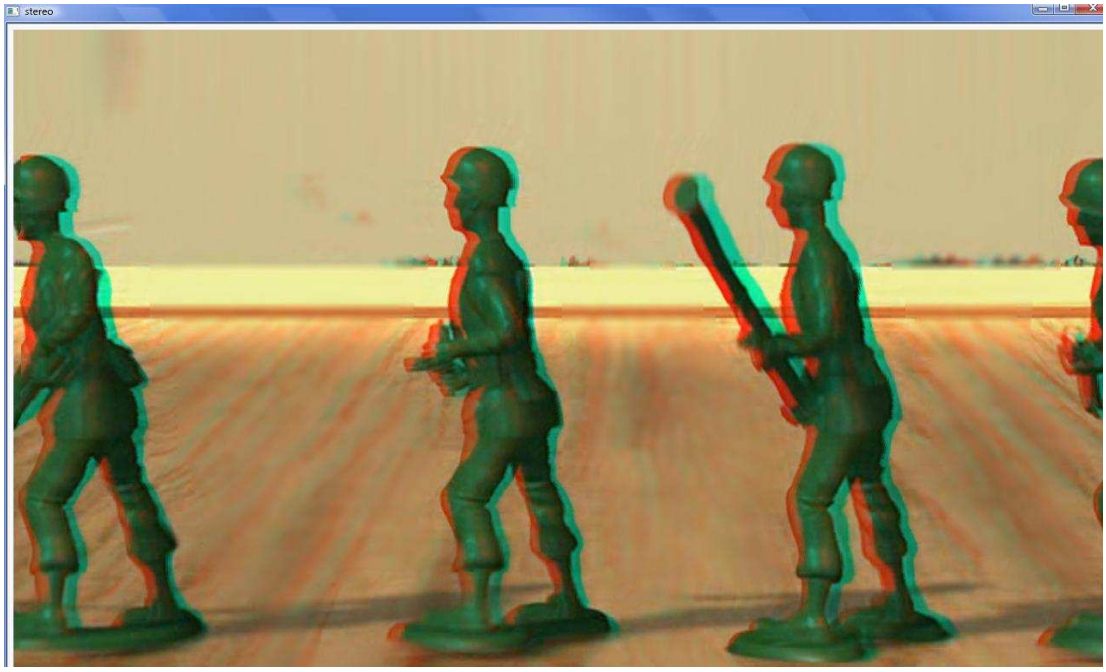

Figure 5.4 : Stereoscopic video frame using motion information (a).

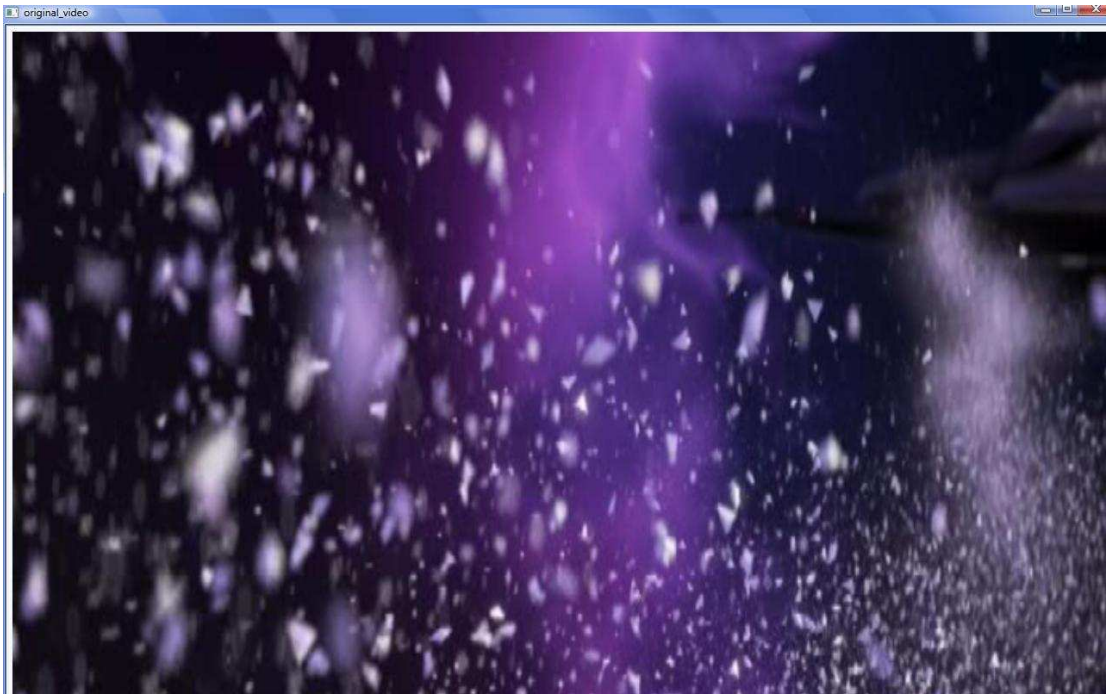

Figure 5.5 : Original frame (b).

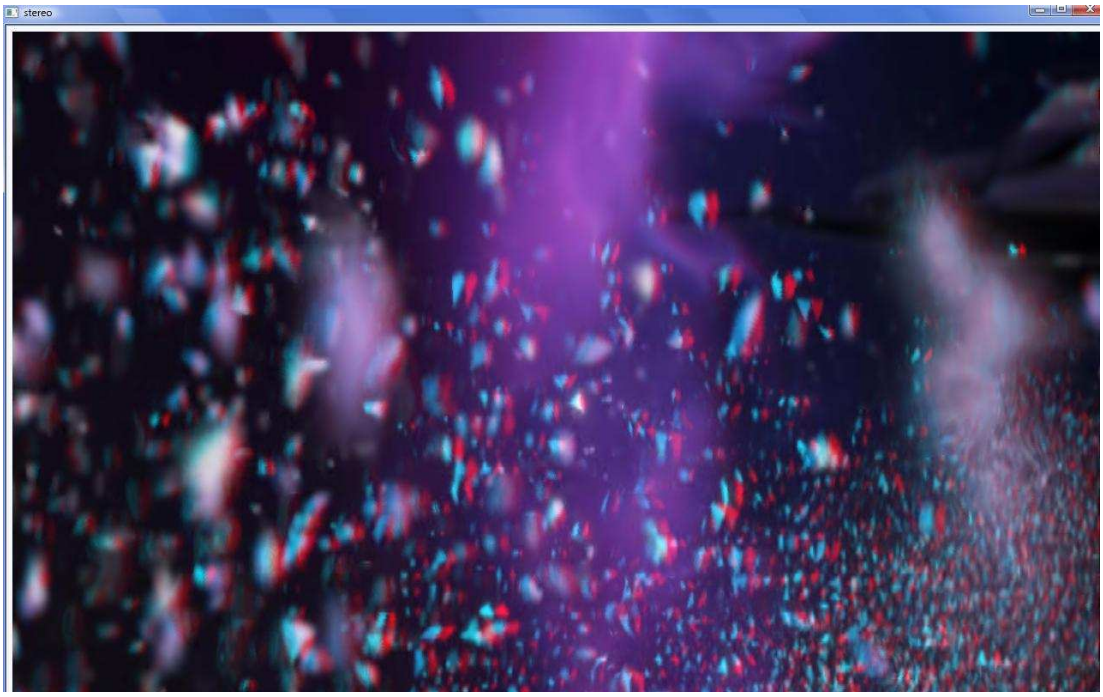

Figure 5.6 : Stereoscopic video frame via edge information (b).

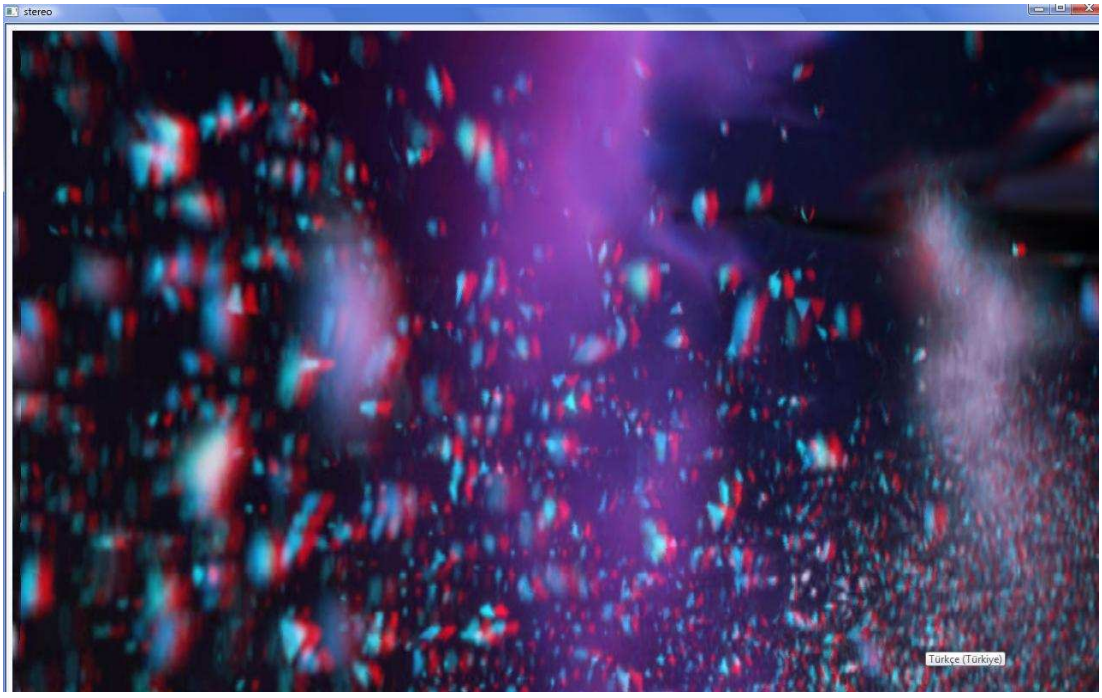

Figure 5.7 : Stereoscopic video frame via motion information (b).

Figure 5.5 is a scene from the “Scattering” test sequence which is obtained from the “Wall-E” movie. It is the original frame of figures 5.6 and 5.7. During watching the two results of this test sequence, the stereoscopic video obtained via edge information gives better result subjectively.

Figure 5.8 : Original frame (c).

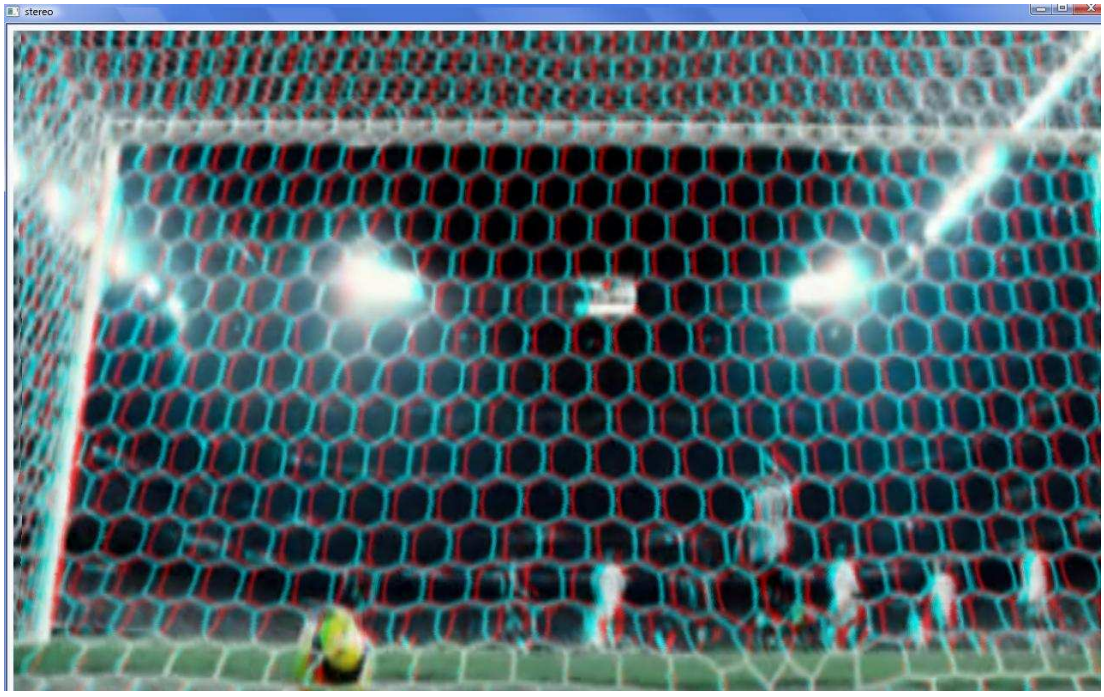

Figure 5.9 : Stereoscopic video frame via edge information (c).

Figure 5.8 is the original scene from the “Football” test sequence. Figures 5.9 and 5.10 are its converted stereoscopic frames. “Football” video test sequence belongs to the Goal movie, and it contains real life scenes only.

.

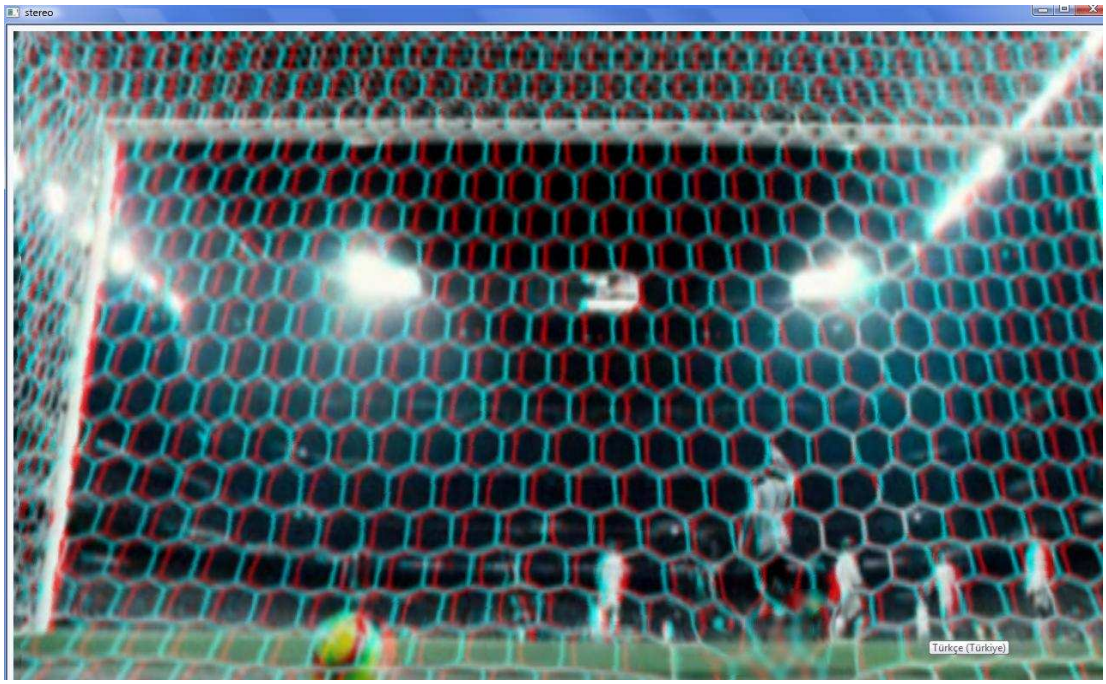

Figure 5.10 : Stereoscopic video frame via motion information (c).

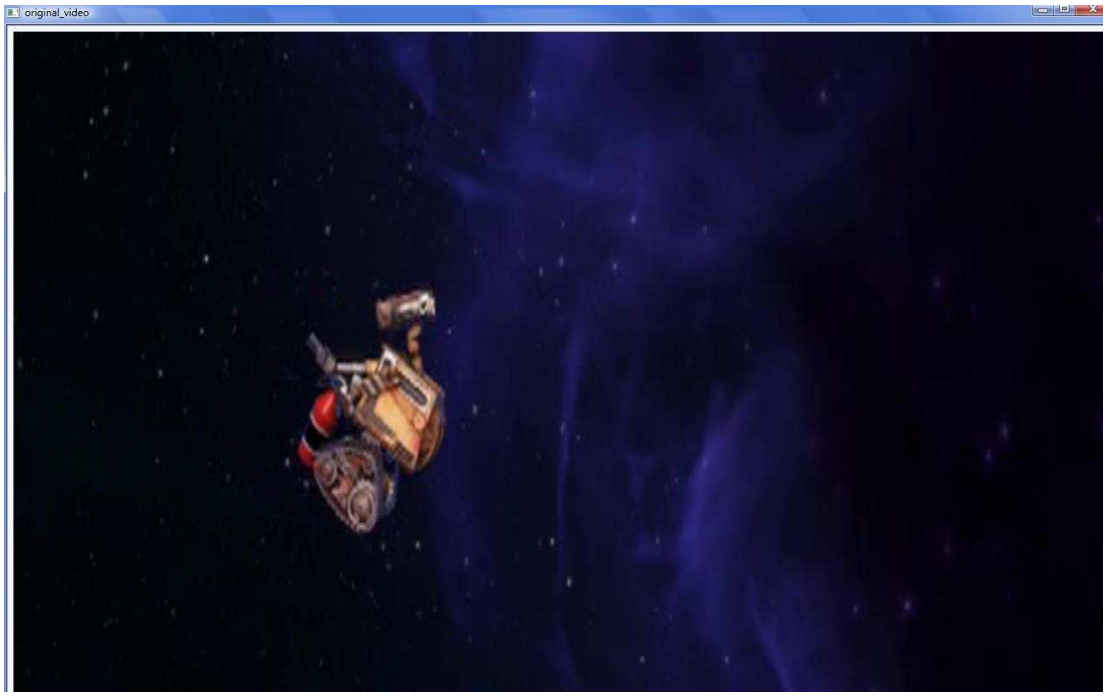

Figure 5.11 : Original frame (d).

Figure 5.11 is the original scene of figures 5.12 and 5.13. This frame is taken from the “The Robot” video test sequence, and it belongs to the “Wall-E” movie.

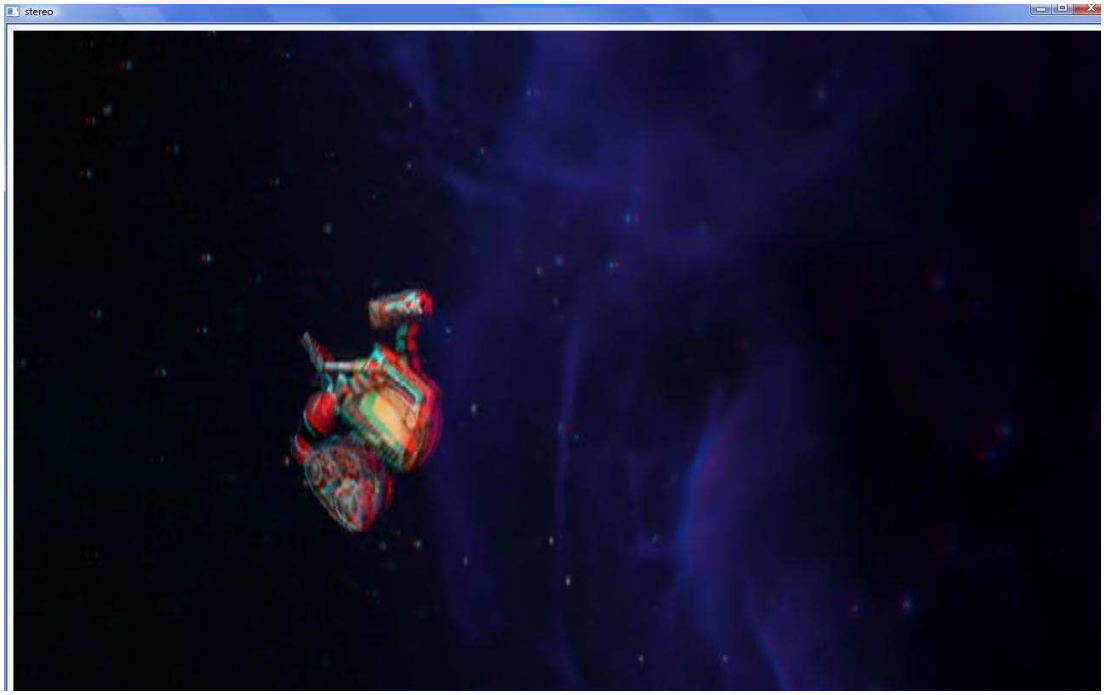

Figure 5.12 : Stereoscopic video frame via edge information (d).

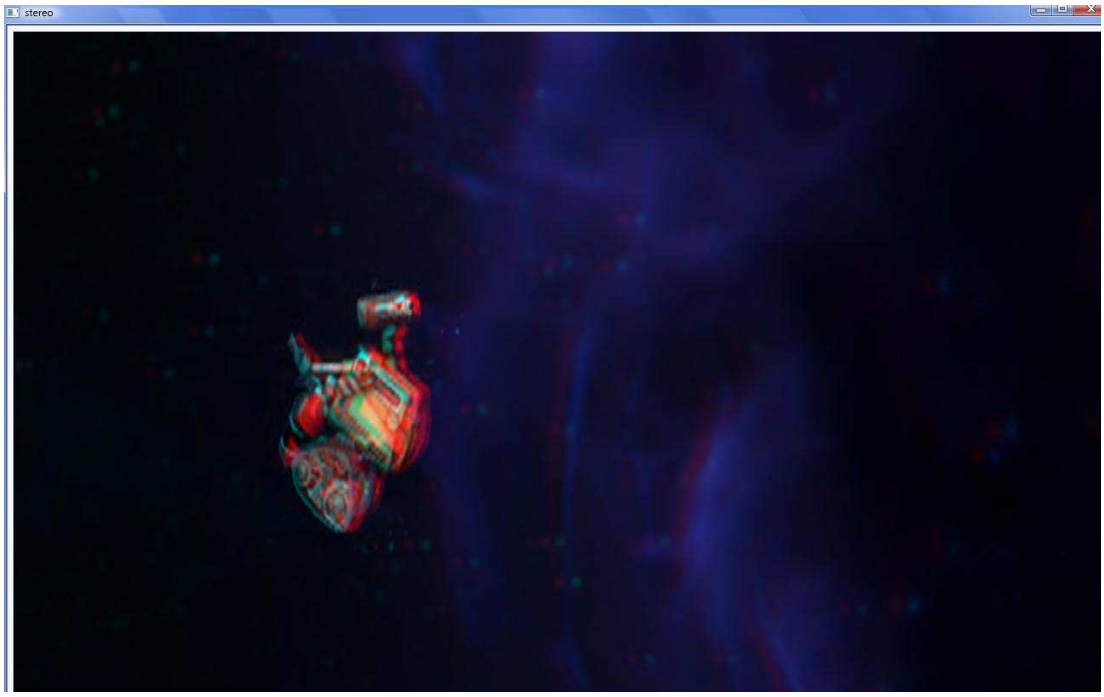

Figure 5.13 : Stereoscopic video frame via motion information (d).

Figure 5.14 : Original frame (e).

Figure 5.15 : Stereoscopic video frame via motion information (e).

Figure 5.14 is the original scene of figures 5.15 and 5.16. This frame is from the Real Simpsons video test sequence; this test sequence belongs to the “Real Simpsons” miniseries shown on FOX TV.

Figure 5.16 : Stereoscopic video frame obtained via edge information (e).

When there are no motion vectors, the depth value is assigned to zero and because of this reason, depth maps, which are obtained via motion vector information, do not give the approximate disparity results for the background of the image. In addition, it can be seen from the above figure that the stereoscopic video frame obtained via motion information has more blur in the background than the edge algorithm's test result, and the depth estimation via motion estimation cannot give the correct result every time. For example, The Real Simpsons sequence contains zooming and panning motion. Stereoscopic image pair consists of left and right images; artificially generated left and right images should have a horizontal parallax difference. The result of the zooming and panning motion do not give the correct horizontal parallax values. Panning gives identical motion vectors, and zooming creates a radial motion vector distribution which grows out from the center uniformly. These results cannot generate the correct depth map. In addition, in the case of a static scene, no depth map can be recovered using motion vectors, and 3D perception would be impossible. Alternately, the video broadcasting or the shot scene may have a high degree of blur; in this situation, the edge detection algorithm would be unsatisfactory for 3D perception.

Figure 5.17 : Original frame (e).

Figure 5.18 : Stereoscopic video frame via motion information (e).

It can be seen from figure 5.18 that there is no 3D effect generation. The reason for this situation is the particular conversion between the scan formats used to generate this test video. With a high probability, this sequence captured via progressive scan and converted to the format of interlaced scan for displaying on TV. During the conversion to interlaced format, the odd and even fields are extracted from a progressive frame, and then interpolated to full frames. Because of this reason, the motion algorithm cannot find motion vectors between adjacent frames corresponding to the odd and even fields extracted from the progressive frame; hence, it cannot assign a depth value. This explanation is further supported by the fact that its frame rate is 54 fps. The edge algorithm is not affected from such a situation.

Figure 5.19 : Stereoscopic video frame obtained via edge information (e).

Figure 5.19 shows a result frame of the Real Simpsons sequence. As can be seen from this image, when two such similar frames are compared to each other, the edge algorithm does not fail like the motion algorithm.

Average run time results of both algorithms for the various test sequences are compared to each other, and the results are given in figure 5.20.

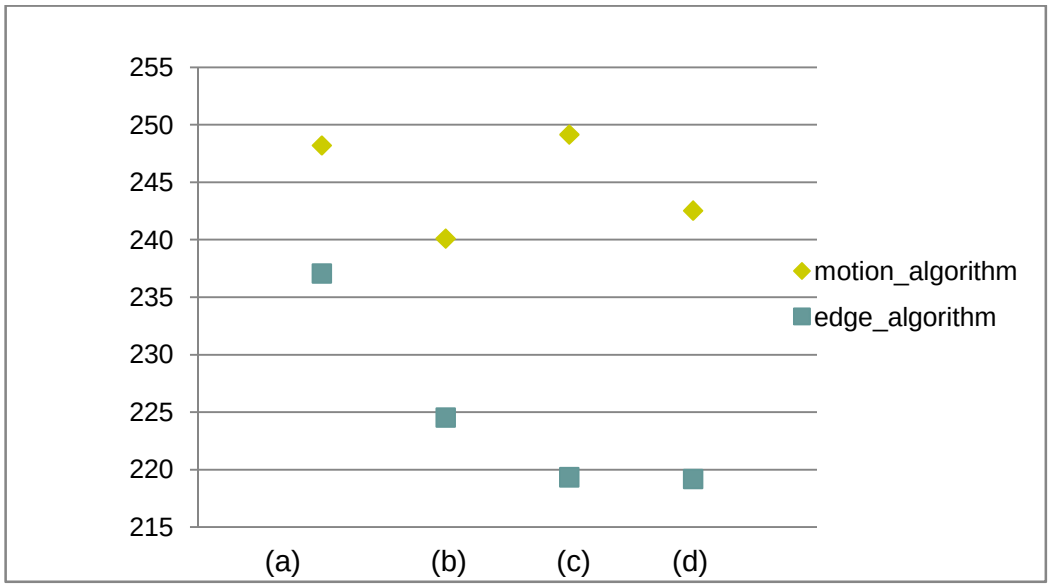

Figure 5.20 : Average computational run time results.

(a)Football Sequence

(b) Scattering Sequence

(c) Soldier Sequence

(d)The Robot Sequence

We see that the motion algorithm's average computational run time results are larger than the edge algorithm results.

6. CONCLUSION AND FUTURE WORK

In this thesis, the fundamentals of stereoscopy, 2D video to 3D effect generation methods are investigated, and automatic stereoscopic conversion algorithms are implemented.

Two methods are used to generate stereo pair video frames starting from a single video frame source and its related depth map. One of them uses motion vector calculation via the block matching algorithm. This method combines motion vector information with depth extraction. Depending on the magnitude of motion vectors, the corresponding block is assumed to be either near the camera or far from the camera. This implementation does not work in static scenes and it does not give reliable results for complex motions such as zooming, panning, and fast motion. The other method uses edge map calculation via the Canny edge detection algorithm. The magnitude of gradients, which is the edge strength, defines the depth information. The weaker edge strength is assumed far from the camera and vice versa. Stereoscopic video generation via edge information cannot handle blurred videos. The weaker edge strengths and regions, which contain no motion, assumed unfocused pixels, and for generating more reliable 3D effect, they are artificially blurred.

The relation between the source scene and the depth map allows reconstructing artificially formed stereo image pairs producing a 3D effect depending on screen dimension and video resolution. The implemented algorithms simply assign a shift to the source points to give a 3D entertainment effect to the viewer, and do not attempt to acquire actual true depth values. Using anaglyphs, visualization of the 3D videos is achieved using any 3D display device available. However, the anaglyph stage may be replaced with any other available 3D display technique. By applying the obtained left and right images to a LCD shutter glass scheme, higher quality 3D images may be viewed without the color problems of the anaglyph.

Subjective quality results of edge algorithm are better than other results of implemented motion algorithms. Computational run time results show that stereoscopic conversion implementation with edge algorithm is faster than motion algorithm.

For future work, to find better parallax value for more convincing 3D effect generation, image segmentations via classification algorithms and the perspective cue of the scene can be combined and used.

REFERENCES

[1]**Onural, L., Sikora, T., Ostermann**, **J., Smolic, A., Civanlar, R., Watson, J.,** An assessment of 3DTV Technologies, 2006, NAB Broadcast Engineering Conference, Las Vegas, USA, April 22-27.

[2]**Fehn, C.,** 2005, 3D TV Broadcasting, 3D Videocommunicaton Algorithms, Concepts and real time systems on human centered communication, Chapter 2, Schreer O, Kauff P. and Sikora T., Wiley, West Sussex, England.

[3]<*http://www.nyu.edu/its/pubs/connect/archives/98fall/hannastereo.html*>, accessed at 26.02.2011.

[4]**Goldstein, E.B.,** 2002, Sensation and Perception, Wadsworth, Belmont, USA.

[5]**Holliman, N.**, 2005, 3D Display Systems, Department of Computer Science, University of Durham Science Laboratories, Durham, England.

[6]**Minoli, D.,** 2011, 3D Television (3DTV) Technology, Systems, and Deployment, CRC Press, NewYork, USA.

[7]<*http://depthbeyond.com/blog/wp-content/uploads/2010/10/parallax.jpg*>, accessed at 27.02.2011.

[8]**Zilly, F., Kluger, J., Kauff, P.,** 2011, Production for Stereo Acquistion, Proceedings of the IEEE, IEEE, **Vol. 99**, No.4.

[9]<*www.eruptingmind.com/depth-perception-cues*>, accessed at 28.02.2011.

[10]<*http://en.wikipedia.org/wiki/3D_modeling*>, accessed at 28.02.2011.

[11]<*http://www.ccrs.nrcan.gc.ca/resource/tutor/stereo/chap3/chapter3_2_e.php*>, accessed at 28.02.2011.

[12]<*http://en.wikipedia.org/wiki/Anaglyph_image*>, accessed at 28.02.2011.

[13]**Onural, L.,** 2010, 3D Video Technologies An Overview of Research Trends, SPIE, Washington, USA.

[14]**Choi, C., Kwon, B., Choi, C.,** 2004, A Real-Time Field-Sequential Stereoscopic Image Converter, IEEE Transactions on Consumer Electronics, IEEE, **Vol. 50**, No. 3.

[15]**Coll B., Ishtiaq F., O'Connell K.,** 2010, 3DTV At Home: Status Challenges, and Solutions for Delivering a High Quality Experience, Motorola - Applied Research Center, Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Arizona, USA, January 13-15.

[16]**Kim D., Min D., Sohn K.,** 2008, A Stereoscopic Video Generation Method Using Stereoscopic Display Characterization and Motion Analysis, IEEE Transactions On Broadcasting, **Vol. 54**, No. 2.

[17]**Okino, T.,** 1995, "New Television with 2D/3D Image Conversion Technologies," SPIE Photonic West, SPIE, **Vol. 2653**, pp. 96-103.

[18]**Lin, C., Chin, C., Fan, K., Lin, C.,** 2005, A novel architecture for converting single 2D image into 3D effect image, Cellular Neural Networks and Their Applications, 2005 9th International Workshop on, Hsinchu, Taiwan, May 28-30.

[19]**Feng, Y., Jayaseelan, J., Jiang, J.,** 2006, Cue based Disparity Estimation For Possible 2D to 3D Video Conversion, Visual Information Engineering IET International Conference on, Bangolore, India, September 26-28.

[20]**Chang, Y., Fang, C., Ding, L., Chen, S., Chen, L.**,2007, Depth Map Generation for 2D-to-3D Conversion by Short-Term Motion Assisted Color Segmentation, Multimedia and Expo, IEEE International Conference on, Beijing, China, July 2-5.

[21]**Tam, W., Zhang, L.,** 2006, 3D-TV Content Generation: 2D-to-3D Conversion, Multimedia and Expo, IEEE International Conference on, Toronto, Canada, July 9-12.

[22]**Po, L., Xu, X., Zhu, Y., Zhang, S., Cheung, K., Ting, C.**, 2010, Automatic 2D-to-3D video conversion technique based on depth-from-motion and color segmentation, Signal Processing (ICSP), 2010 IEEE 10th International Conference on, Beijing, China, October 24-28.

[23]**Ideses, I., Yaroslavsky, L., Fishbain, B.**, 2007, Real-time 2D to 3D video conversion, Journal of Real-Time Image Proceedings, Springer, **Vol. 2** pp.3–9.

[24]**Ideses, I., Yaroslavsky, L., Fishbain, B.**, **Vistuch, R.**, 2007, 3D from Compressed 2D Video, Stereoscopic Displays and Virtual Reality Systems XIV., Proceedings of the SPIE, **Vol. 6490**.

[25]**Wu, Y., An, P., Wang, P., Zhang, Z.,** 2010, Stereoscopic Video Conversion Based on Depth Tracking, Signal Processing (ICSP) 10th International Conference on, Beijing China, October 24-28.

[26]**Xu, F., Er, G., Xie, X., Dai, Q.,** 2008, 2D-to-3D Conversion Based on Motion and Color Mergence, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, Istanbul, Turkey, May 28-30.

[27]**Chang, Y., Chen, W., Chang, J., Tsai, Y., Lee, C., Chen, L.,** 2008, Priority Depth Fusion for the 2D-to-3D Conversion System, Proceedings of. SPIE 6805, pp.13.

[28]**Chiang, T., Tsai, T., Lin, Y., Hsiao, M.,** 2010, Fast 2D to 3D Conversion Based on Wavelet Analysis, Systems Man and Cybernetics (SMC) 2010 IEEE International Conference on, Istanbul, Turkey, October 10-13.

[29]**Cheng, C., Li, C., Chen, L.,** 2010, A Novel 2D-to-3D Conversion System Using Edge Information, IEEE Transactions on Consumer Electronics, IEEE, **Vol. 56**, No. 3. [30] **Fehn C.,** 2003, A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR), Visualization, Imaging and Image Processing, Benalmadena, Spain, September 8-10.

[31]**Fehn, C.,** 2003, A 3D-TV System Based On Video Plus Depth Information (DIBR), Signals, Systems and Computers Conference Record of the Thirty-Seventh Asilomar Conference on, California, USA, November 9-12.

[32]**Dubois, E.,** 2001, A projection method to generate anaglyph stereo images, International. Conference. on Acoustics Speech Signal Processing, Utah, USA, May 7-11.

[33]**Omeltshenko, S.,** 2010, 3-D-display, TCSET'2010, Lviv-Slavske, Ukraine, February 23-27.

[34]**Tekalp, M.,** 1995, Digital Video Processing, Prentice Hall, NJ, USA.

[35]**Redert, A., Berretty, R. P., Varekamp, C., Willemsen, O., Driessen, H.,** 2006, Philips 3D Solutions From Content Creation to Visualization, The 3rd Int. Symposium on 3D Data Processing, Visualization, and Transmission, Chapel Hill, USA, June 14-16.

[36]**Canny, J.,** 1986, A computational approach to edge detection, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **Vol. PAMI-8** No.6.

[37]< *http://www.pages.drexel.edu/~weg22/can_tut.html*>, accessed at 02.05.2011.

[38]**Tam, J., Wa, S. F., Yano, S., Shimono, K., Ono, H.,** 2010, Stereoscopic 3DTV Visual Comfort, IEE Transactions On Broadcasting, IEEE, **Vol. 57**, No.2.

[39]**Richardson I.,** 2002, Video Codec Design, John Wiley & Sons, West Sussex, England.

CURRICULUM VITAE

Candidate's full name:	Özlem AYDOĞMUŞ
Place and date of birth:	Üsküdar/10.10.1985
Permanent Address:	Selimiye Park Üstü Sok. No:12/6
	34668 Üsküdar ISTANBUL
Universities and Colleges attended:	ISIK University, Physics, Major
	ISIK University, Electronic Engineering, Minor