ÇANKAYA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
COMPUTER ENGINEERING

MASTER THESIS

CONVERSION OF TWO-DIMENSIONAL VIDEOS TO THREE-DIMENSIONAL

MEHMET UMUT GÖKBULUT

JUNE 2013

Title of the Thesis : **Conversion of two-dimensional videos to three-dimensional**

Submitted by : **Mehmet Umut GÖKBULUT**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University

Prof. Dr. Taner ALTUNOK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Assist. Prof. Dr Murat SARAN
Deputy Chairman of the Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Reza HASSANPOUR
Supervisor

Examination Date: 17/06/2013

Examining Committee Members:

Asst. Prof. Dr. Reza HASSANPOUR (Çankaya Univ.) ------------------------

Assist. Prof. Dr. Kasım ÖZTOPRAK (Karatay Univ.) ------------------------

Assist. Prof. Dr. Abdül Kadir GÖRÜR (Çankaya Univ.) ------------------------

STATEMENT OF NON-PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Surname : Mehmet Umut GÖKBU LUT

Signiture :

Date: 26/07/2013

ABSTRACT

CONVERSION OF TWO-DIMENSIONAL VIDEOS TO THREE-DIMENSIONAL

GÖKBULUT, Mehmet Umut

M.S.c., Department of Computer Engineering

Supervisor : Asst. Prof. Dr. Reza HASSANPOUR

Co-Supervisor: Asst. Prof. Dr. Kasım ÖZTOPRAK

June 2013, 76 Pages

Through existing technology, depth map is provided by means of general purpose video converting programs. However forming the depth map estimation by these programs need not only much more labour but time also. The main contribution of this study is to provide depth map estimation by a faster and easier way than the current methods.

After performing depth map estimation, a new video is formed by integrating this depth map and existing video. This new video can be played side by side on the 3D devices.

Key Words : Image based depth conversion, Depth Map Estimation.

ÖZ

İKİ BOYUTLU GÖRÜNTÜLERİN ÜÇ BOYUTA DÖNÜŞTÜRÜLMESİ

GÖKBULUT, Mehmet Umut

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Supervisor : Asst. Prof. Dr. Reza HASSANPOUR

Co-Supervisor: Asst Prof. Dr. Kasım ÖZTOPRAK

Haziran 2013, 76 Sayfa

Bu tezde, filmlerin iki boyuttan üç boyuta çevrilebilmesi ve sonucunda ortaya çıkan video ile orijinal video arasındaki görüntü kalitesi farkları incelenmiştir. Çalışma temel literatür taraması ve uygulama geliştirmeyi kapsamaktadır. Dönüştürme işi için en önemli ihtiyaç iki boyutlu filme paralel akan bir derinlik haritasının oluşturulmasıdır.

Günümüz araçları ile derinlik haritası ihtiyacı genel amaçlı video işleme programları ile giderilmektedir. Fakat bu araçlar ile derinlik haritalarının oluşturulması hem çok fazla iş gücü hem de zaman gerektirmektedir. Bu çalışmanın en temel amacı, derinlik haritasının mevcut yöntemlerden daha hızlı ve daha kolay bir şekilde oluşturulmasını sağlamaktır.

Derinlik haritası oluşturulduktan sonra film derinlik haritası ile birleştirilerek yeni bir video oluşturulmaktadır. Bu film üç boyutlu film oynatabilen görüntüleme cihazlarında direkt olarak yan yana oynatılabilmektedir.

Anahtar Sözcükler : Görüntü Tabanlı Derinlik Dönüşümleri , Derinlik haritası çıkarımı.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my supervisor Asst. Prof. Dr. Reza HASSANPOUR and Asst. Prof. Dr. Kasım ÖZTOPRAK for their continued guidance, also patience and support throughout my research. They provided counsel, and assistance that greatly enhanced my studies and know-how.

This study is performed as a part of the project STZ.855.2011-2 which is supported by The Ministry of Science, Industry and Technology.

TABLE OF CONTENTS

FIGURES

LIST OF ABBREVIATIONS

IMO	Independently Moving Object
2D	Two Dimensional
3D	Three Dimensional
BG	Background
FG	Foreground
FR	Frame
Thr	Threshold
DIBR	Depth Image Based Rendering
CQ	Colour Quantization
UI	User Interface
CCD	Charge Coupled Device

CHAPTER 1

INTRODUCTION

The scientists from different disciplines (physics, space sciences and most of the engineering sciences) are interested in modelling the world in computer platforms. The technological development in video technology and cinema industry resulted in development and use of 3D video in cinema saloons and homes.

Images perceived in the depth and effect of literality occurred in the brain by means of this depth influences and engulfs the audiences into the images. Therefore 3D video is always more literal and preferable than 2D video is. There is no considerable study on 3D video till 2010 because all studies were performed in 2D. However many studies and algorithms for 3D video conversion are available today and the main aim of them is providing 3D video more vivid and more literal by estimating the depth of a 2D video. [1]

Nowadays, 3D video is very popular because they may present the audiences more inclusive and more literal experiences. They are managing this by providing a clearer depth perception than 2D video is.

Vision researchers are working on algorithms to estimate 3D information by using only images or single camera shots for the past 20 years. Currently, the algorithms are evolved and became mature enough to give good representations of the scenes without requiring much human intervention. As a result of this progress, people are interested in 3D video since they enable us creation of more realistic movies and applications. By using the new production and conversion techniques to create materials, a lot of PC games, new science-fiction movies, and even the 3-D interviews on TV channels became popular.

The health sector also prefers 3D in many applications because it is easier to process in three-dimensional images/video.

The most commonly used 3D image method applied todays is embedding the computer animations into the movies. Therefore, the subject of most of the 3D movies is animation for kids or science fiction. TRON and AVATAR are the best examples of 3D animation movies.

TRON is a science fiction movie and most of the scenes are on space, therefore the environment in most of the scenes is prepared in computer. The depth perception in the space is implemented by computer effects However since the movie was aimed to be played in 3D, before it was taken, the dresses of the actors and the borders of the equipment to be used are rolled edges of the light bands, and then, these light bands were given depth by taking the movie in dark environment generally.

AVATAR was also dreamed up in a virtual world and all the life in this world were prepared by animations and depth perception was given by computer animations again. [1] How to obtain stereoscopic visual perception and associating the colours in 3D conversion is shown in Figure 1. This subject will also be dealt in Chapter 2 again.

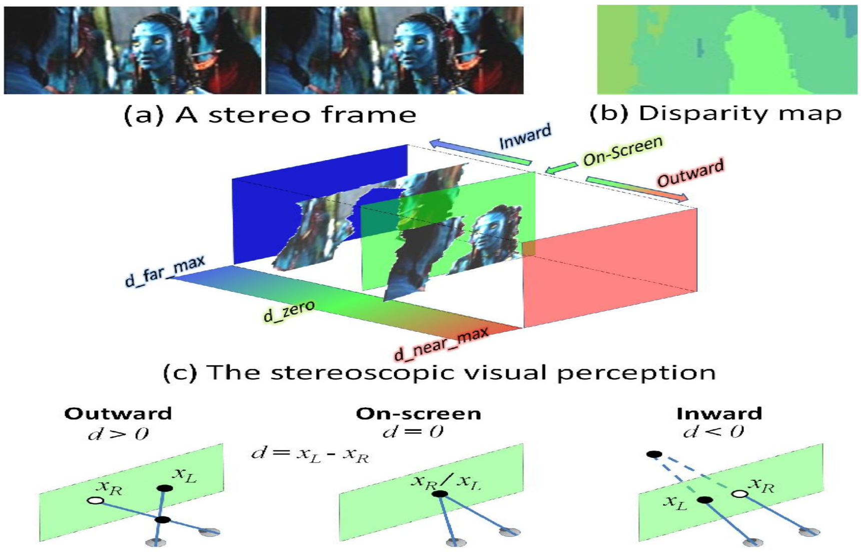


Figure 1 3D stereoscopic screen principle (the best colour display). When looked at a 3D TV, images seen by left eye only and right eye only are different (a), and this difference is due to stereoscopic visual perception. As shown on the disparity map (b), pixel diversities can change depending on the depths in 3D image itself. When looked at the disparity values (d), there exist three different ranges referring three different stereoscopic visual perceptions; outward perception (positive), onscreen perception (zero) and inward perception (negative). As shown on (B), blue, red and green colours are used to show the negative, positive and zero disparities respectively. The farthest pixel is the deepest one and hence, the darkest one [1].

Another method is obtaining images by putting the frames of the same scene and sliding them in a particular angle. This method forms the main technology used in 3D televisions in the market.

In addition to above, binocular cameras heard too much in last days is used in these movies. In contrast to the old cameras, these cameras are telescope-like recording devices, with a distance of about 6.5 cm between them.

All these studies done by nowadays' tools and methods are not good enough to supply 3D images as real and correct as we percept by our eyes.

There is no particular method to convert 2D movies taken before to 3D by adding the depth without using any computer animation. Therefore, we will try to explain the ways to convert these 2D images to 3D by adding them the depth with no computer animation.

1.1 Scope of the Thesis

In this thesis, two different methods were tested and the results were obtained.

Currently used methods to convert 2D images to 3D can be divided into two main groups: Manual and automatic conversion.

Automatic conversion method does not require as much professional labour and long-time as manual conversion does. Furthermore, because the facility of manual intervention is limited in automatic conversion process, it may be difficult to form the desired perception and since this process is new, outputs in the desired quality are not produced yet. Therefore the usage area is also limited in the market.

All the studies done so far shows that manual conversion can provide more depth and the converted images by this method is closer to the real 3D images but it takes longer time. Automatic conversion can do it in shorter time but lots of information is lost and the 3D image obtained is no close to real image because the depth is given automatically in this method.

Hence, we will try a semiautomatic method to obtain more realistic 3D images with less data lost in shorter time.

As the consequence of all our researches, many methods from very simple to very hard have been investigated to convert a video to 3D and a new tool as a combination of all methods has been developed.

1.2 Thesis Outline

Chapter 2 mentions shortly what dimension is. It gives short information about the methods of image formation processes and how an image takes place on the camera and also about the camera types. It also gives fundamental information about how human sees the world in 3D (human vision).

Chapter 3 provides general information about the research performed this subject, algorithms prepared during the course of these studies and about the state of these algorithms in our study.

Chapter 4 explains the method we used in our experimental studies. What is used to form this method and which algorithms and studies directed us are explained in this chapter.

Chapter 5 provides the summary for the overall study as well as some concluding remarks and some open points for possible future studies.

CHAPTER 2

BACKGROUND INFORMATION

2.1 Background

To obtain 3 dimensional images, primarily "DIMENSION" term should be defined:

A geometrical term used to define the quantity measured in axes. Point is accepted to have no dimension. Line has only one dimension in length. Planes have two dimensions (2D) in length and width and volumes have three dimensions (3D) in length, width and height.

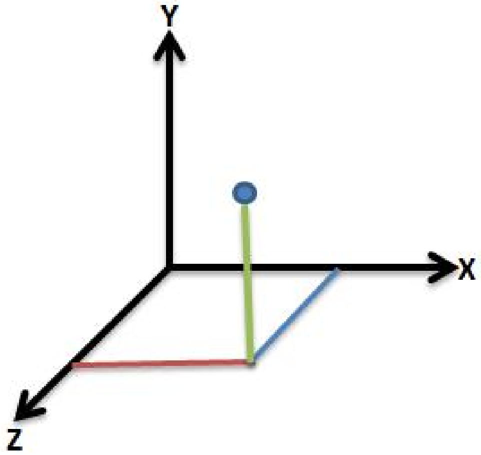


Figure 2 Three dimension in linear plane

2.2 Image Formation

To form an image of an object, main requirements are sufficient illumination and camera on which the image will occur. Reflections and refractions are the main principles of image formation by determining the brightness of the object in connection with the illumination and the surface.

2.2.1 A simple model of image formation

The lights coming from the source of illumination are reflected by the object towards the camera and image is formed on the camera via chemicals of the film inside. This simple model is shown in figure 3 below.

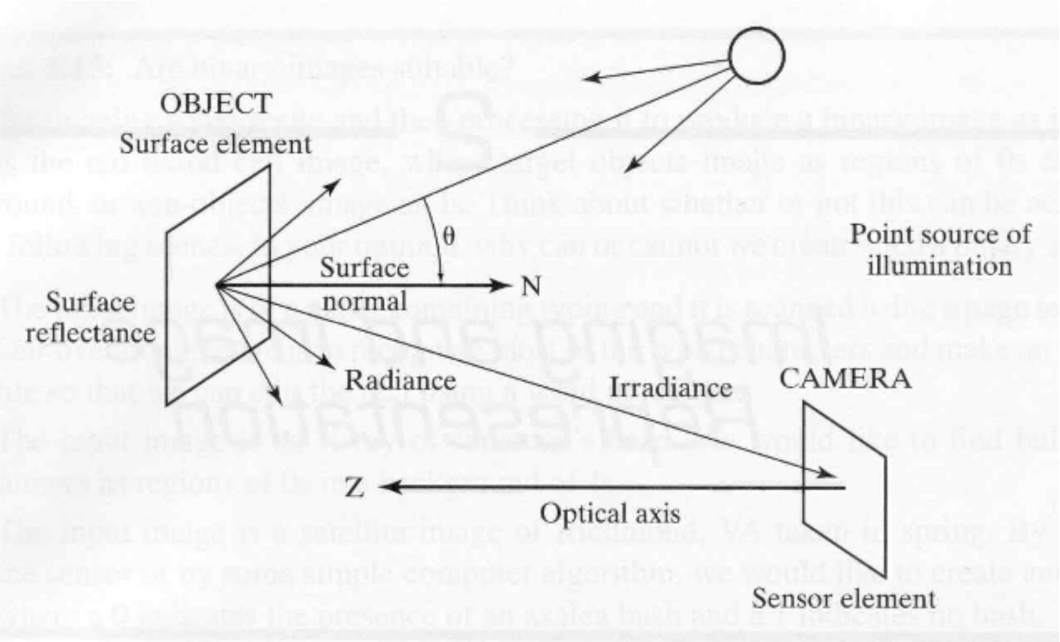


Figure 3 Camera senses system [2]

2.2.2 Principles of a camera

"Pinhole" camera is the simplest device to form an image of a 3D object on a 2D surface. An inverted image is formed on the image plane by the rays of light passing through the hole. Figure 4 shows this principle.

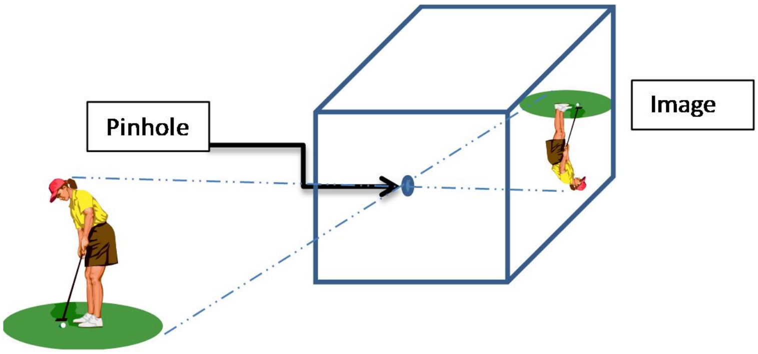


Figure 4 Pinhole Camera

2.2.3 Camera optics

Basic function of lenses is forming a small and clear image on the image plane. Cameras have lenses in different focus distances. How big or small the image will be depends on the lenses only. Lenses are classified as wide angle, narrow angle and variable focus distanced (zoom).

Figure 5 shows basic camera image formation through a simple lense.

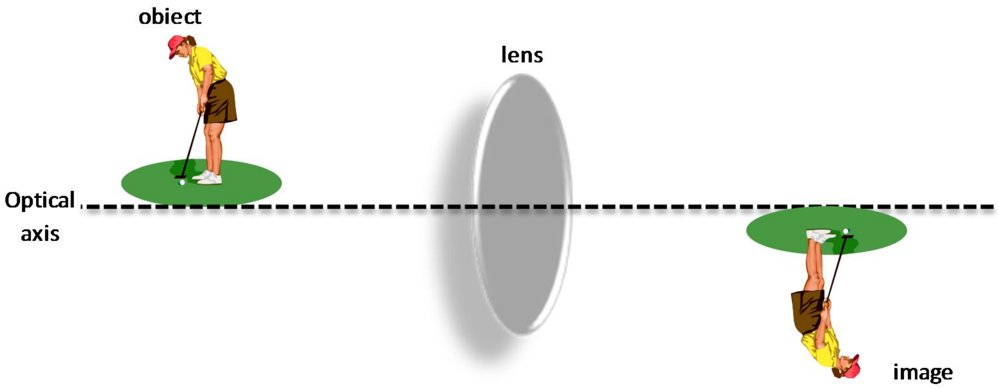
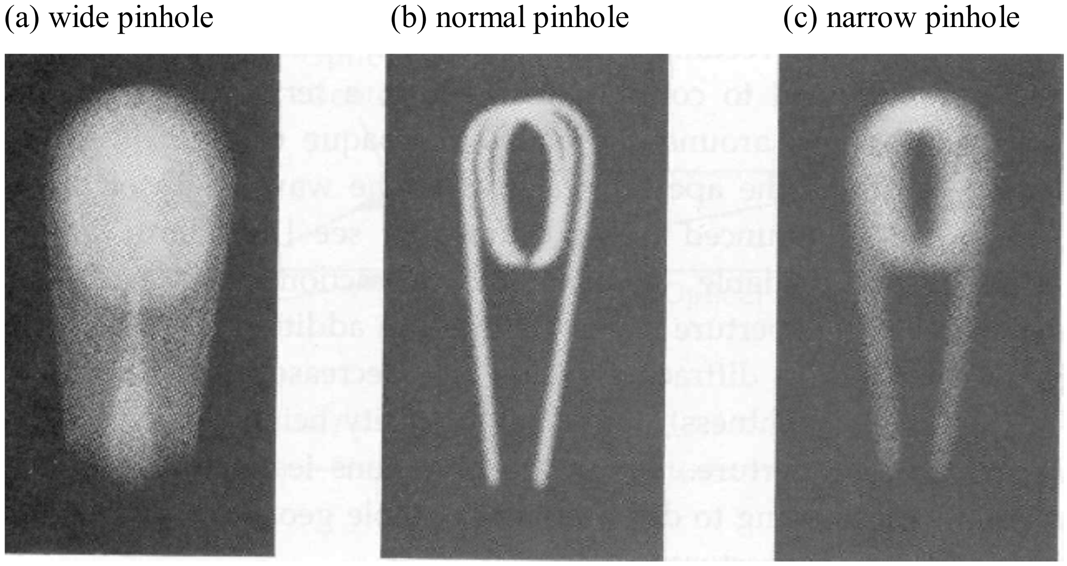


Figure 5 Camera Optics

2.2.4 Diffraction and pinhole optics



(a) wide pinhole (b) normal pinhole (c) narrow pinhole

Figure 6 Diffraction and Pinhole Optics [2]

Pinhole camera is the simplest tool used for generating the images of 3D object on 2D screens.

- If a wide lens is used, the light coming from the source diffuses through the lens as blurred (i.e. not focused).

Figure (6/a) shows the effects of wide lenses.

- If the lens is narrow, very little light can pass through it.

Figure (6/c) shows the effects of narrow lenses:

- Picture clarity is limited by the diffraction of the light.
- Light passes through a small diaphragm but does not go straight, it diffuses.
- This is a quantum effect and it diffuses in many directions.

In general, the aim of using lens is to duplicate the pinhole geometry without resorting to undesirable small apertures.

2.2.5 Human vision

Camera is a good imitation of human eye. In human eye, pupil is small when the light level is high and blurring is due to diffraction; pupil is open when the light level is low and blurring is due to lens imperfections.

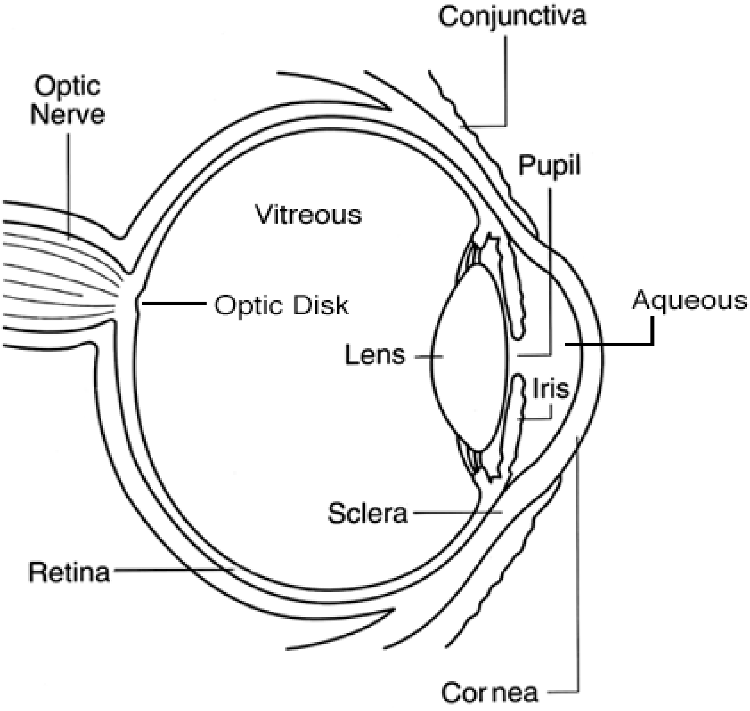


Figure 7 Human eye [36]

2.2.6 CCD cameras

CCD cameras are light sensitive cameras. Principles of CCD cameras are converting the light energy into electrical current through the whole accumulated diode sensors. Voltage collected depending on the light is sent to vertical transfer register and signals received in line by line from vertical transfer register are sent to horizontal shift register, and then to output register point by point from horizontal shift register. Output register is the part providing the video output by raising the electrical signals received from horizontal shift register. In CCD cameras, clear images of moving objects are provided by varying the exposure time. 25 full and 50 half frames are taken in a second. Exposure time of a frame is 1/50 second.

Figure 8 CCD cameras [37]

2.2.7 Frame grabber

When the operation method of the cameras is investigated, it is seen that all the cameras put the images into a frame while they record them. Cameras send all the video signals to an electronic device called frame grabber while they record images. Frame grabber digitizes these signals into a rectangular integer values array.

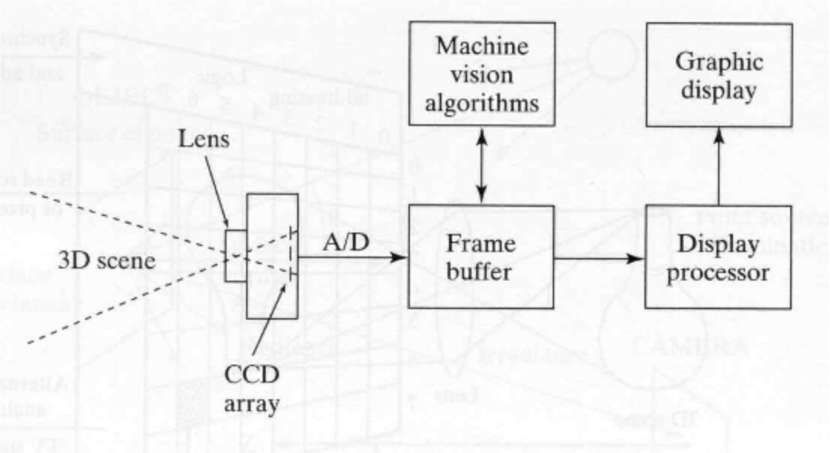


Figure 9 Frame grabber [2]

Depth perception is a perception formed by the human brain. Our brain momently receives two different images taken in different views by our left and right eyes due to the distance between our eyes. The images coming through our eyes are integrated in our brain and are perceived as a single image only. Our brain forms the depth perception by means of the different images in different views.

It forms the 3 dimension concept the distance in about 6.5 cm between two eyes on human anatomy and perception of depth provided by the images taken in two different views due to this distance. If there had been only one eye in human anatomy, he would see the images in 2 dimensions only.

The basic concept in forming 3 dimensional images is the perception of depth. The distinguishing characteristics are not only the depth but also the distance perception occurred for the objects seen by the human eye.

On 3D concept, the main difference between the objects closer to and far from the lenses is that the closer objects move faster whereas the distant ones move slower in perception. Therefore the movement speed and the replacement coordinates of the closer objects are also important in the study of 3D conversion [8][9].

Following is the example of one of the most important details that surprised human when noticed and this fact has not been changed for centuries:

To understand better how our eyes perceives the depth, put your forefinger in the mid of your both eyes in arm distance and close one of your eyes while the other is open and close the open one while opening the close one and do it in serial times by changing your closed eye in order. You will see that your finger is in different places changing according to the view of your open eye. This is the perception forming the basis of the 3D concept. In connection with this fact, if there had been only one eye in human anatomy, he would see the world in 2 dimensions only.

To form the perception of depth, left and right eyes should receive different images taken in different views. For that purpose, some special equipment is used for staging 3D movies in todays. i.e. Glasses, stereo projectors, and auto stereoscopic televisions. Some of this equipment needs different images prepared for left and right eyes whereas some of them need depth map.

Most of the 3D videos that we started to watch too much are produced by animations prepared by special effects by computer or by binocular cameras that has two lenses approximately in 6.5 cm distance between.

The discovery of binocular cameras was modelled on human anatomy. Two lenses in 6.5 cm distance are connected to a single recorder. Two different images are recorded in two different angles, so is the depth information.

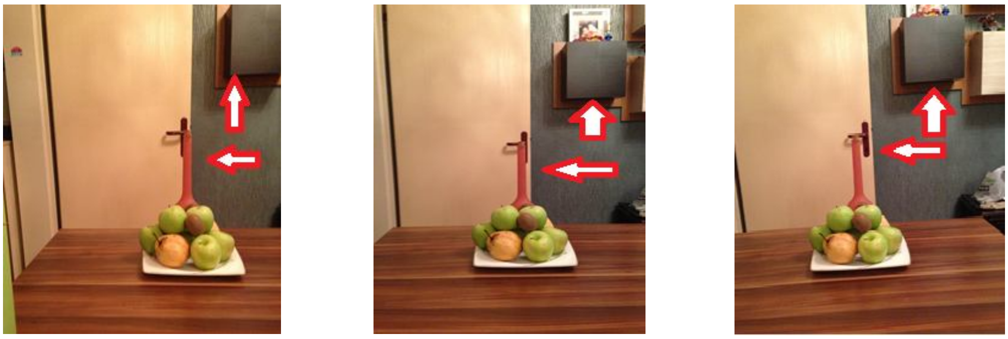
2.3. What is 3D images and how to obtain?

Charles Wheatstone invented the first stereoscopic viewer of the world based on the Renaissance perspective in 1838. This instrument having various mirrors in different perspectives was containing 2 separate images for left and right eyes. When looked at these images at the same time; using Wheatstone's invention a stereo image was obtained. A new age in the area of moving and stationary images has started by this invention.

I will use a plate of fruits on the table as the example. For better perceiving, I will put the plate at my eye level and look at it while my one eye is closed.

Figure 10 Eye closing method

In this way after looking at the plate for a while, when I look at the plate with my eyes open, I may perceive the parts of the fruits closer to me. The objects at the background: In left eye looking, the boxes at the rightmost are not seen. Similarly in right eye looking, the leftmost part of the door is not seen.



(1) (2) (3)

Figure 11 That left and right eyes see the images in different angles forms the basis of 3D images. When we look at the object by our left eye first (1) and then right eye (3), we may see that the object moves slightly. When we look at the same object by our left

Let us take another example. Take a sewing needle and thread and try to have the thread passed through the eye of the needle while your one eye is closed.

As you notice, the thread cannot pass through the eye of the needle.

Now, please try the same with two eyes open and you will notice that you can have the thread passed through the eye of the needle much more easily.

Interesting but this is the truth. Our eyes cannot perceive the depth, can see only 2 dimensional. Actually we can never see the third dimension. Having two eyes ensures us to perceive an object in two different perspectives. Since the distance between the eyes is slight more than 5 cm, the images on the two retinas are different from each other. The images taken from two different perspectives of an object is combined in the visual centre of the brain. The perception of the third dimension is activated in the brain and thus we may see the image of an object in three dimensions. Contrary to common belief, third dimension is obtained in the brain, not in the eyes. Third dimension is a perception and all the perceptive processes are realized in the brain. Depth and distance between the objects are perceived in this way. Therefore, people having only one eye cannot see three-dimensional objects/scenes.

The images on the television screen or on any paper are two-dimensional. Todays, although, there exists different kinds of 3D glasses are available; the aims of all are same. The glasses are designed for separating two images taken in different perspectives and transferring them into each eye respectively. Therefore, eyes can take two different images from one image on the television or paper and again the brain combines both images and perceives its depth and distance.

2.3.1 3D image techniques

Two types of techniques will be investigated in this study.

Anagram

SBS (Side by side frames)

2.3.1.1 Anagram image technique:

Anagram image is one of the oldest 3D techniques using red-cyan glasses. This technique can obtain depth in cinema, television and even better on paper.

Figure 12 Anagram glasses [38]

Why are the colours red-cyan used on these glasses?

Because, when looked though the red glasses, red coloured images cannot be seen since both are in same colour.

Similarly, you cannot see the blue coloured images through the cyan glasses.

Figure 13 Above are two photographs seems same to each other but both were taken in two slightly different angles, similar to the difference of angles due to the distance between two eyes.

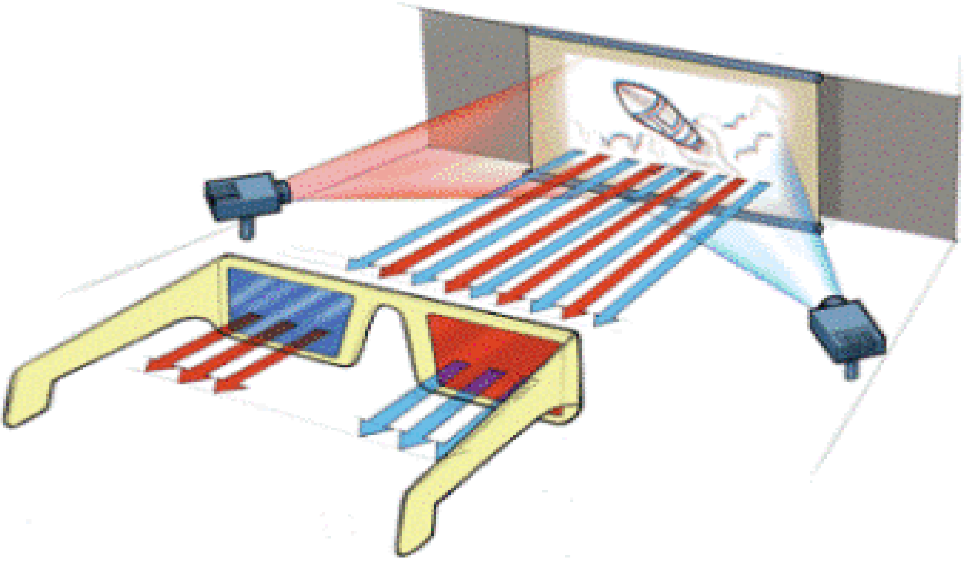


Figure 14 Anagram glasses working method [38]

Now both images should be placed on the same frame however the one on the left side should be seen only by our left eye and the one on the right side should be seen only by our right eye. For that, the colour tonnes in the image are separated. Considering that we have CYAN glasses on our left eye and RED glasses on our right eye, the image is separated in two colours on our left and right eyes as RED and CYAN respectively. We may have two different images in two different angles on our two eyes in this method and then it's the brain turn to combine both images in his perception.

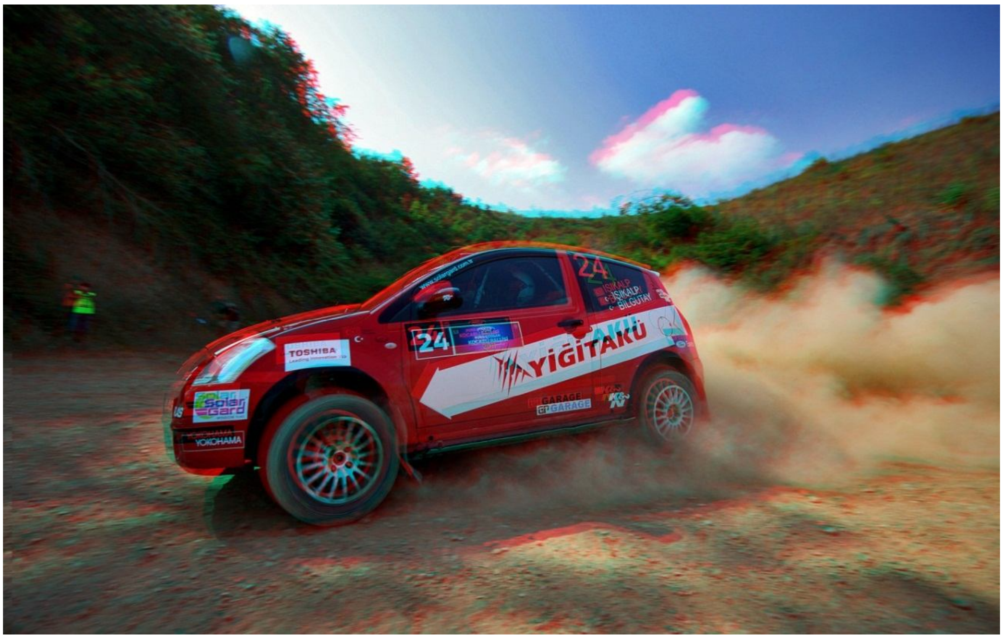


Figure 15 When you look at this derived photograph by means of red-cyan glasses, you may recognize the depth and distances very well.

2.3.1.2 SBS (Side by Side Frames) technique:

Similar to the Anagram technique, the aim of SBS (Side By Side) is to reflect two different images on two eyes separately, but this system runs only in digital media, so it cannot be done on paper like anagram technique.

This technique has two application methods for the moment; one is **XPAND** (MASKING) and other is **POLARIZING**.

XPAND:

Figure 16 Xpand glasses [39]

if we separate the pixels of 2 different images and call them as 0 and 1 for the first and second images respectively, as stated in its definition (SBS - side by side frames), when both images are put side by side to each other, first half will be pixel 0 and second half will be pixel 1 and the appearance of the image will be as follows:

```
00000000001111111111   =>   01010101010101010101

00000000001111111111   =>   01010101010101010101

00000000001111111111   =>   01010101010101010101

00000000001111111111   =>   01010101010101010101

00000000001111111111   =>   01010101010101010101

00000000001111111111   =>   01010101010101010101
```

And both images will take place in one frame as nested.

This nesting process is done by the image source. At the same time, source gives the 0 to pixels in vertical polarized light whereas it gives the 1 to pixels in horizontal polarized lights.

Below practice shows the logical explanation of this 3D polarized process:

Figure 17 3D polarized image

For a Full HD image in 1920x1080 pixels, we divide it into two images in 960x1080 pixels for left and right frames.

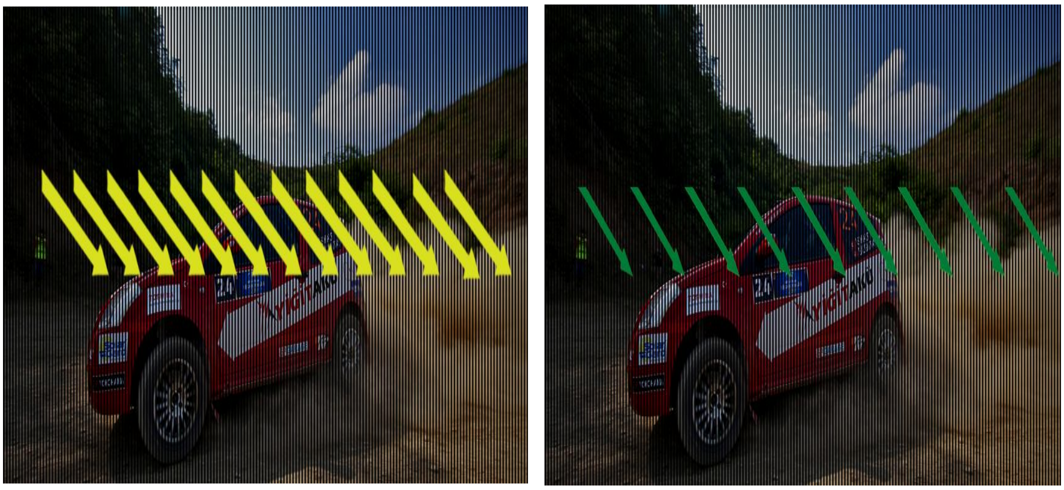


Figure 18 The left image is given with X vertical polarized light and the right image is given with Y horizontal polarized light.

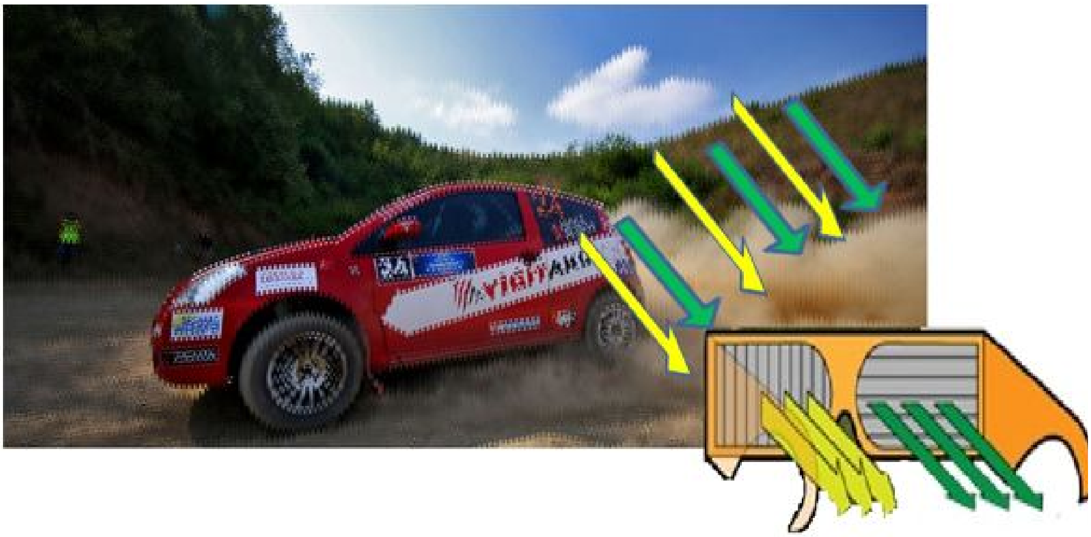


Figure 19 Xpand Glasses working method

As the result, image source is nesting both images into each other in the same frame and reflecting it in two polarized lights as vertical and horizontal and you will perceive the vertical lights on your left eye and horizontal lights on your right eye through the vertical left and horizontal right polarized glasses and consequently two different images reaches to two different eyes. And again it's the brain's turn to combine both images in his perception. Depth and distances occurred in this perception.

CHAPTER 3

RELATED WORKS

Many studies performed nowadays are usually based on dimensioning on computer games, 3D movies or videos prepared by computer animations are easier and more real compared to the ones taken by binocular cameras.

According to Dr. Lai-Man Po's studies, much attention has been paid to depth research. Because, the closer to reality the depth map of a video is, the more quality image can be generated. This thesis is based on Dr. Lai Man Po's studies because it is the most result-oriented study done so far. Stereoscopic video is relied on the illusion effect of the human eye. Thus, the main purpose of the 2D-to-3D conversion system is generating an additional view to the videos taken by mono-lens cameras. The structure of the automatic 2D-to-3D video conversion system is based on depth estimation from move by using the block matching and from colour based regional segmentation [3].

Synthesis view selection: In this method, the original video taken by monoscopic camera is threated as right eye view and left eye view is generated by mesas of depth map estimation by DIBR and the original video. In this selection, the dominant perception in the human brain for which eye will be used is the main tool. The common approach for use of left or right eye in human perception is 70% right eye, 20% left eye and 10% no preference. Hence, it is important to study on the 3D images taking the right eye perception as basis. Right eye should be based on all the studies on DIBR applications, especially (1) Hole-filling and (2) Depth Map Pre-processing. The quality of converted 3D image could be increased by this method [3].

According to Lai-Man, there are two major processes to convert 2D videos to 3D videos; if they are correctly applied, the real three-dimensional image can be achieved.

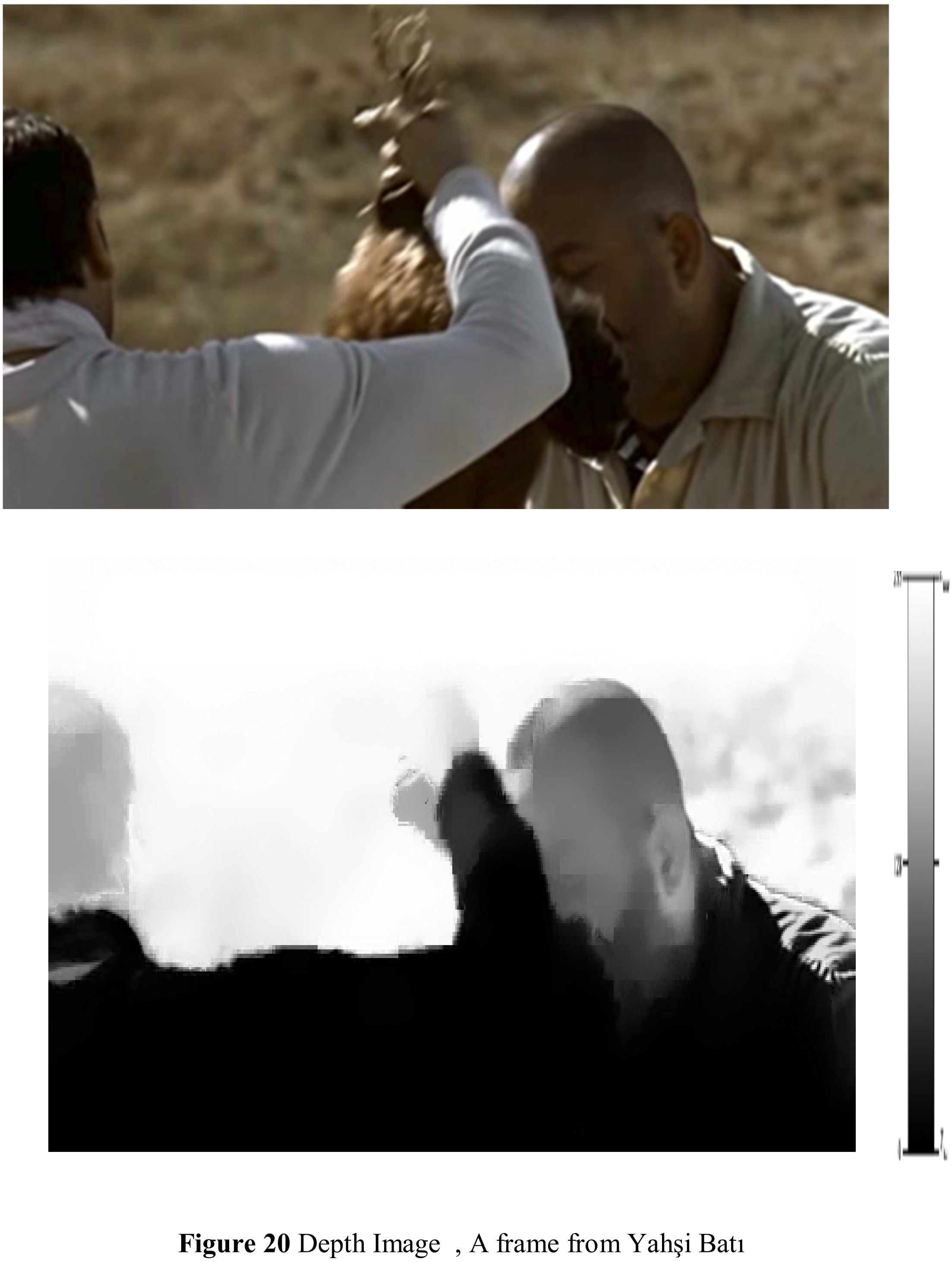


Figure 20 Depth Image , A frame from Yahşi Batı

3.1 Depth Map Generation

3.2 Depth Image Based Rendering (DIBR)

What is Depth Map or Depth Image?

Each depth image stores depth information in 8-bit grey values with the grey level 0 for the furthest value and the grey level 255 for the closest value [3].

Changing the grey level values is the most commonly used method in most of the studies done for depth map information and same was done in this study again.

3.1.1 Depth map estimation

For the estimation of the depth map of an environment (If there is no information about lighting and objects) an image recorded in at least two different views is required. It is the main aim of the depth map estimation to match the equivalent of each pixel in both images (stereo images).The basic method of matching is defined in literature as Block Matching Method using the brightness density difference in a specific area (3X3, 5X5 frames) for each pixel in two images[3].

Followings are the usual three methods of depth map estimation:

3.1.1.1 Depth from blur

Depth estimation by changing the blur of the objects, by increasing or decreasing the blurring levels.

First step is taking the negative of the image by blurring on which the depth estimation will be performed.

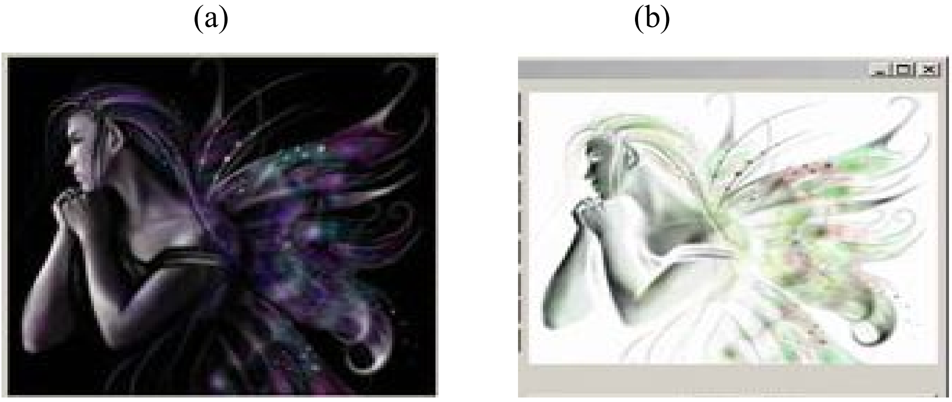


Figure 21 Depth From Blur: First study for the image's depth is blurring the colours seen on the first image and taking its negatives and drawing its main lines.

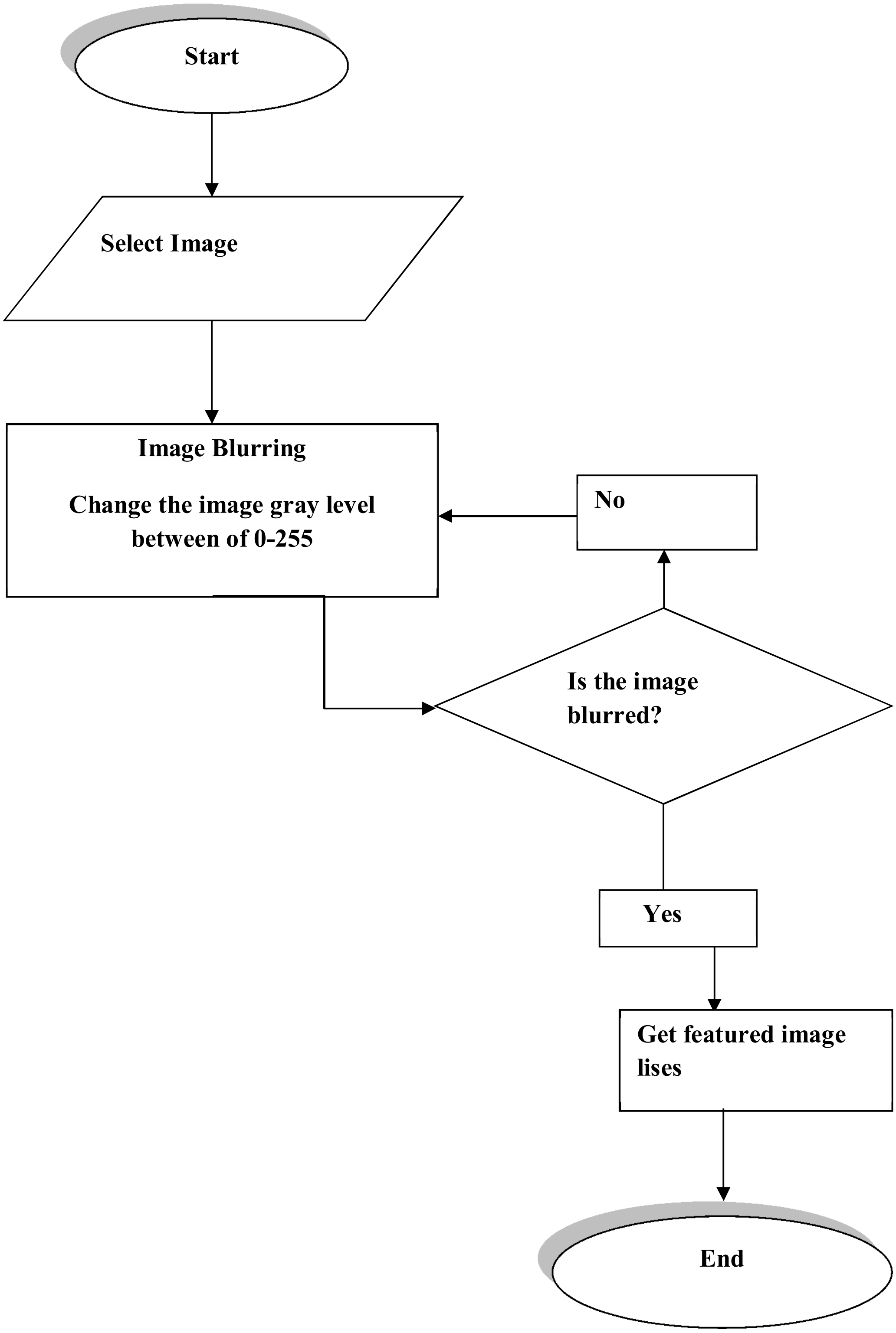


Figure 22 Depth From Blur Algorithm - Grey level of the pixels should be changed till the image is blurred.

3.1.1.2. Vanishing point based depth estimation

Depth estimation by the furthest point of a picture according to the vanishing point.

Changing the colour of the image will bring the part to be changed to the fore.

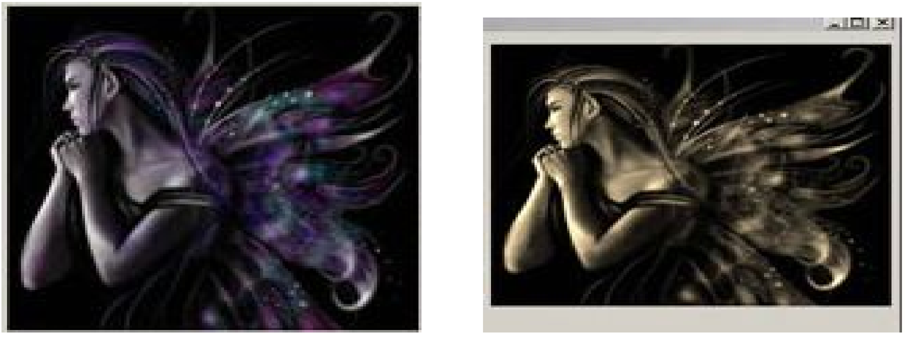


Figure 23 Vanishing Point: Changing the colours in the image, the furthest point in the image could be selected and image is highlighted with its main lines.

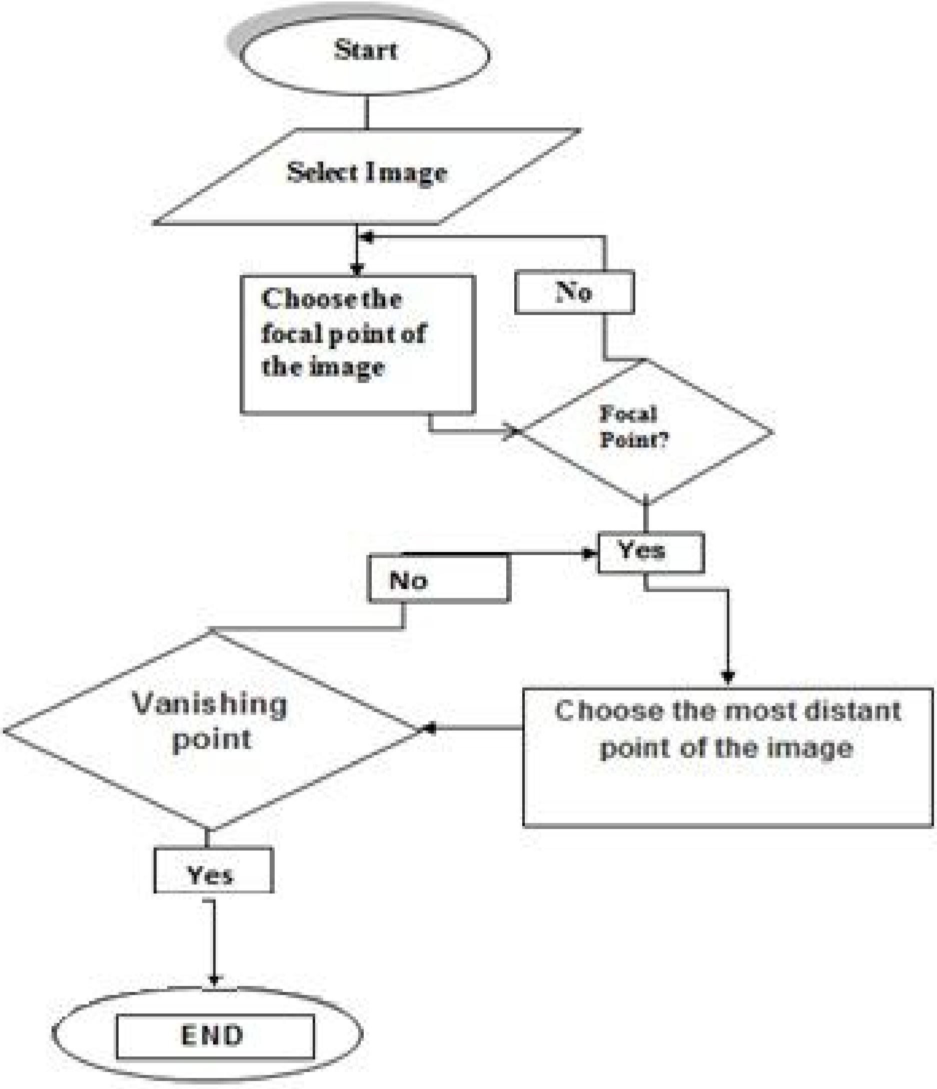

Start
Select Image
Choose the focal point of the image
No
Focal Point?
Yes
No
Vanishing point
Choose the most distant point of the image
Yes
END

Figure 24 Vanishing Point Based Depth Estimation Algorithm - The focal point should be the farthest point in the image. Selecting the farthest point on the image, the distance between vanishing point and this farthest point is calculated and depth estimation.

3.1.1.3. Depth from motion parallax

Depth estimation according to the movement and distance of the objects.

We may change colours by dividing the picture into 1000 parts.

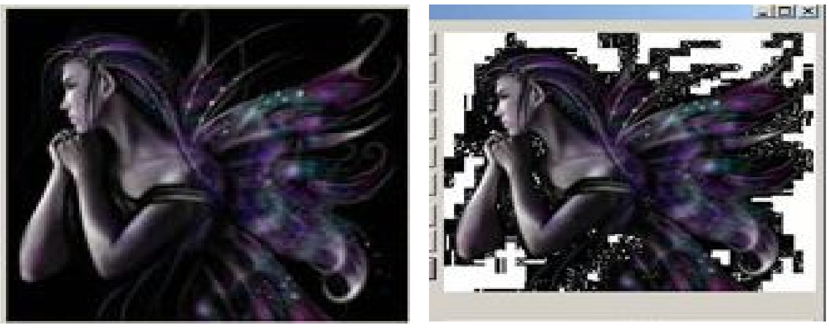


Figure 25 Depth from Motion Parallax

Separate depth calculation of each pixel is done after dividing the picture into 1000 pixels and which pixel is going to be brought to the fore is determined according to that depth.

3.1.2 Block-Matching based depth map estimation

Above mentioned three methods are usually used while estimating the depth on a stereo video.

The most common used method of movement depth calculation is Block Matching Method. Block matching method calculates the replacements of the objects in time in more than one frame and thereby calculates the movement depth.

As seen in below figure, to estimate the replacement of an object in a frame in the following scene, location analysis calculation by movement vectors is used. More accurate results can be obtained by matching the pixels in the basic of this method. That the matching of each pixel placing in the same line is a characteristics providing the search of matching problem in a single (horizontal) dimension. Matching is done by shifting the blocks in a specific area and defining the match where the difference of brightness density is minimal. The most important disadvantage of block matching method is that the regional specifications are lost as the block gets bigger and the match found as giving the minimal difference can be false sometimes.

Block matching methods can give wrong results especially in the areas where the brightness density is continuous. At the same time, the borders of the objects in the image can be lost [3].

The most accurate operation to eliminate such disadvantages is increasing the number of divisions in the picture's own structure; the colour specifications and the general regional borders do not get lost in this way and most accurate estimation may be done. Basic assumption used for depth map estimation by using over-segmented stereo images is that depth of pixels in each segmentation is same. Thus, matching the pixels which is the basic of block matching can be said as the matching of segmentations in lower resolution by segmenting.

For matching the segmentations, each segmentation is slide in horizontally in a particular width and the brightness density difference between each pixel and its matching in the second image is found. To calculate the penalty of the segment depending on the breach value, the calculated absolute values of the differences are summed.

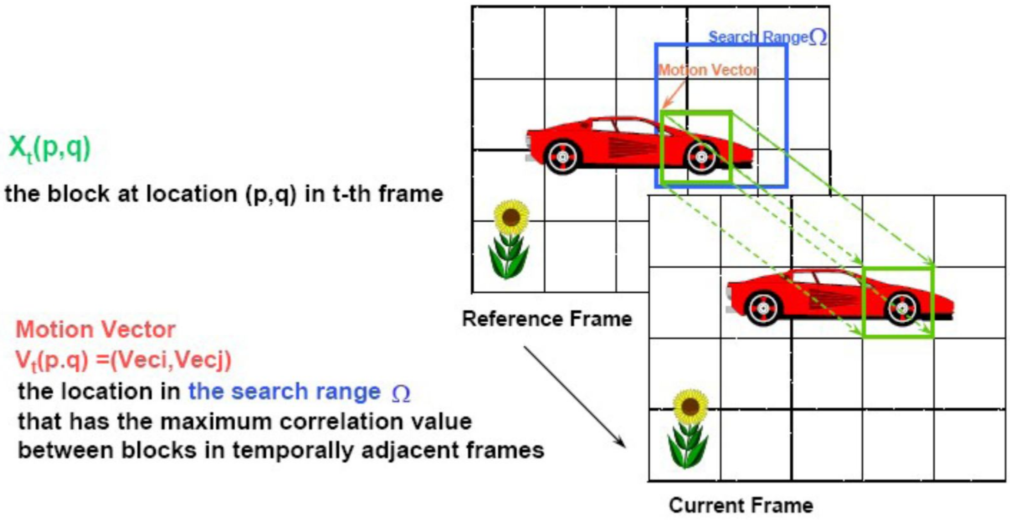

$X_t(p,q)$
the block at location (p,q) in t-th frame
Motion Vector
$V_t(p.q)$ =(Veci,Vecj)
the location in the search range Ω
that has the maximum correlation value
between blocks in temporally adjacent frames
Search RangeΩ
Motion Vector
Reference Frame
Current Frame

Figure 26 Block - Matching [3]

The relative depth information is calculated by [3] ;

$$D(i,j) = \sqrt[c]{MV(i,j)_x^2 + MV(i,j)_y^2}$$

Figure 26, the most practical way of the method is seen that 2D images are divided into blocks in 4X4 and then movement estimation by using the first block as reference to match the same of the second frame. Depth values D (i,j) are estimated as per the magnitudes of the movement vectors. MV (i,j) is a vector having MV (i,j) x and MV (i,j) y values as axis and ordinates in horizontal and vertical directions and c is the pre-defined constant. The disadvantages of this method is that there are many surfaces to be traced and prevailed for calculation. In block matching method, the movement is investigated in terms of the status in the previous and next frames and calculation for “how many frames and which direction the movement happens in” is done. It is very important to apply this method for better depth map estimation [3].

3.1.3 Colour segmentation

As the proposed system, regional colour segmentation is used. Colour segmentation helps the depth regions not realized by the methods applied before to be discovered better. This method is applied as follows; each pixels of the image to be depth estimated are coloured and the texture obtained by this process forms a colour map on the image, this map can give an estimation on the depth of the objects [3].

3.1.4 Fusion

The aim of the fusion is to be subsidiary to the block-based depth map estimation by using the border data of the colour-segmented image. In addition, fusion is used for better depth estimation in each region by means of average of the depth values in that region. This operation does better estimation if it finds small or large depth values in a part of the region in which fusion is applied [3].

3.2. Depth Image Based Rendering (DIBR)

DIBR is used to generate the stereoscopic 3D video, by forming the left eye image from the depth estimation and monoscopic video input.

The DIBR algorithm consists of two processes:

(1) 3D Image Warping
(2) Hole-filling.

3.2.1 3D Image warping

However if more than one locations of an image are used as reference in this method, correct results cannot be obtained.

As the viewpoint is changed, flashings can occur in micro-triangles and WIBR textures do not match and therefore WIBR should be applied by taking only one viewpoint reference.

One of the solutions is that the data is processed initially for forming a single mesh. But this is difficult and possibly prohibitive for the data of real time to be warped [21].

The process includes two steps:

Original image points (e.g. m'(x',y')) from the real view are re-projected into the 3D world

The 3D space points (e.g. M(X,Y,Z)) are projected into the image plane of the "virtual" view (e.g. m(x,y)).

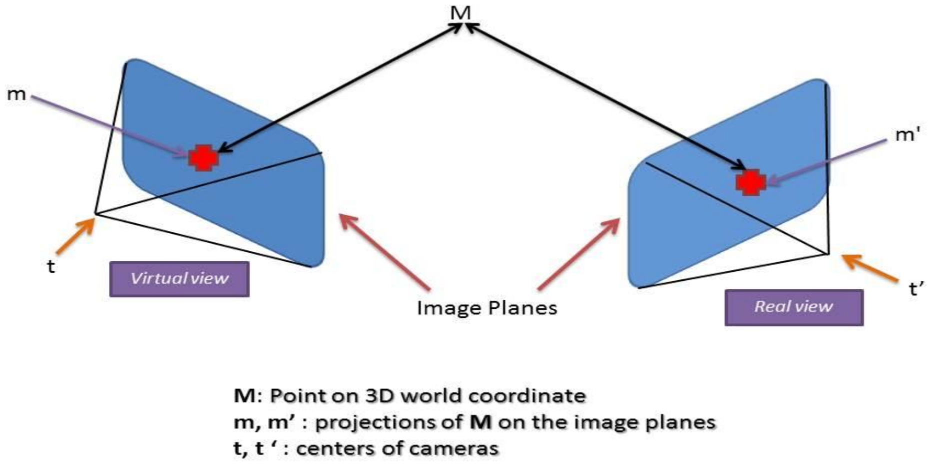

M
m
m'
t
t'
Virtual view
Real view
Image Planes
M: Point on 3D world coordinate
m, m' : projections of M on the image planes
t, t ' : centers of cameras

Figure 27 3D Image Warping

3.2.2 Hole filling

Especially after 3D Warping, there may occur gaps on the image on pixel basis.

There might be some gaps because of images overlapping or lengthening the distance between two frames.

These gaps has to be filled by the nearest pixels to increase the smoothness of imagine.

The method of this process is hole filling.

There are three items of hole filling method.

- All of the holes will be detected.
- The filling of the holes according to the nearest pixels by looking to the average of texture.
- Linear interpolation technology

3.3 Other Studies

Regarding to the surveys done by Berlin University, there are similar methods used like as Dr Lai-Man's. The most important difference is that they examined the back ground variation in the flow chart and calculation of movement depth of each frame in more details.

The summary of the methods used in Berlin University is shown as in the below flow chart.

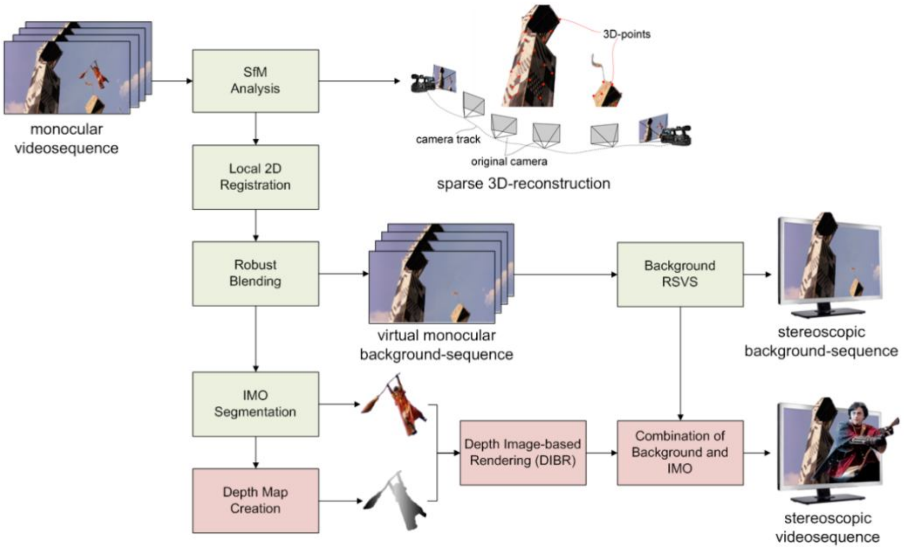


Figure 28 Berlin University Technique of 3D conversion [4]

This study presented a new approach for the segmentation of independently moving foreground objects in sequences recorded with a moving camera, and is thus, utilizable for structure-from-motion-based 2D/3D conversion. This approach extends the applicability of the previous work, which was designed to handle static scenes only. Additionally, a system for the overall monoscopic to stereoscopic view conversion was drafted. It consists of an automatic background conversion and a manual object-based foreground conversion.

The main idea of the segmentation approach is the robust image background estimation using 2D image transformation and blending techniques. The pixel classification is performed applying change detection on the anisotropic smoothed difference image. Experimental results show that in cases where the IMO's motion is significant, high quality segmentation results can be obtained. Simultaneously, the image background estimation ensures photo consistency for uncovered regions while converting the foreground objects. [5]

The limitations of this approach are twofold. Since SFM is used, a translation of the camera is necessary. Hence, conversion of fixed camera shots, shots from panning camera only, or shots from forward moving cameras is beyond the scope . Furthermore, the type of objects and their motion are of main importance. Multiple occluding objects or objects with only little motion are difficult to detect.
Regarding to the surveys done by Yonsei University, calculation of movement depth can be done by estimation of the camera way during the stereoscopic video [6].

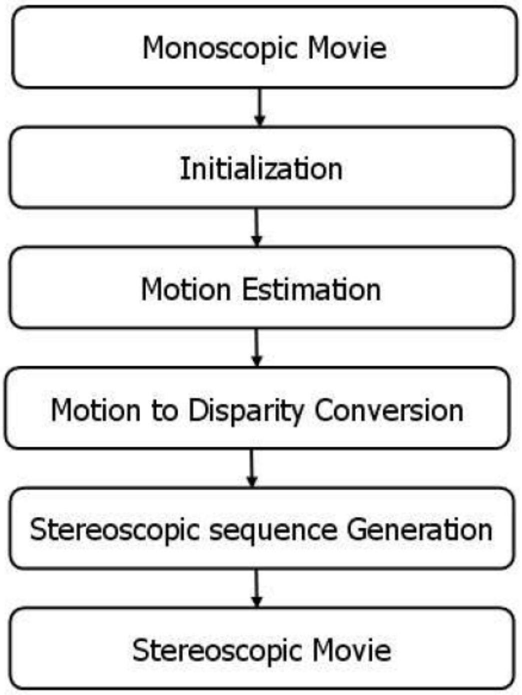


Figure 29 Block diagram of overall system [6]

In the most of these surveys, the estimation of the depths is done by handling the frames one by one.
Similarly, our study is investigating depth estimation by keeping the background stable and changing the foreground to assign the depth maps according to the changes on the movement on the foreground objects. Paint select tool is used for this process.

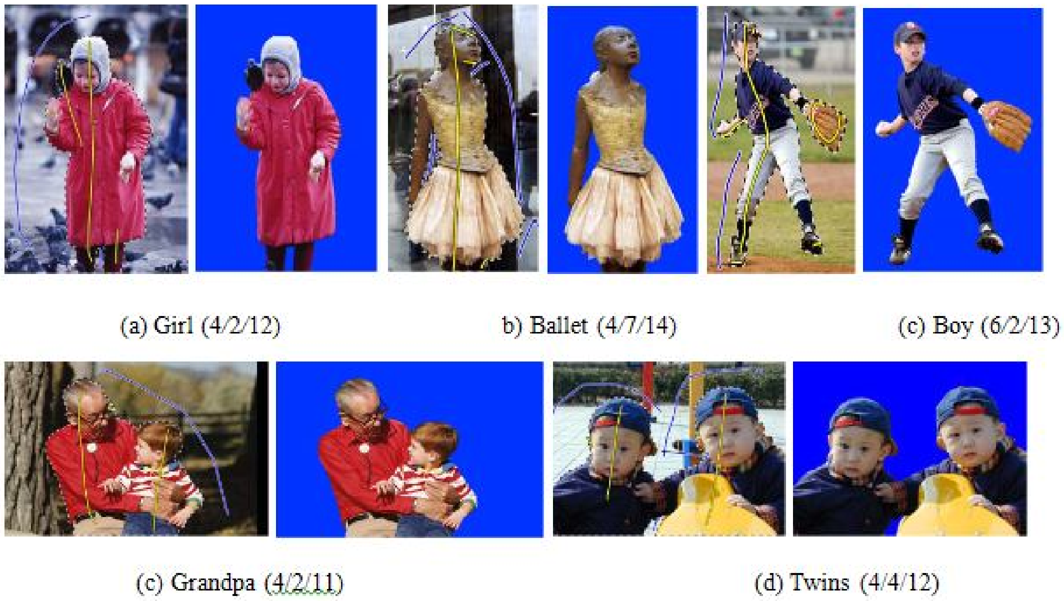
The basic of this survey is also to estimate the depth of frame. If we assume a video with 24 frame per seconds, this means a video will be 1440 frame for one minute. Hence, it will take long time to convert a video with 90 minutes to 3D.
For this long period process, the aim is to find a survey that will make automatic depth calculation of frames and a faster 3D conversion method [7].

The method proposed in this study is making the stereoscopic video conversions according to the differences on the motions. Image diversities can be identified according to the maximum and minimum motions on any stereoscopic video. Subsequently, motion vectors can be defined by using colour segmentation method and KLT tracker features. Then, motion of the camera can be calculated by means of scaling the motion estimation. In another words, in which views the camera took which image and which direction the camera was moving to during the stereoscopic recording those images can be found [10].

Regarding to the surveys done by Hong Kong University of Science and Technology, Lazy snap tool is an interactive graph cut tool. It separates the selected object from the background and from other objects neigh boor to itself by cutting it from the picture. Even for the images having unclear edges and objects very close to each other, it catches the object in the image by selecting it easily.

And because it is an interactive tool, it sets the selected object off in the selection stage and gives visual feedbacks in case of selection faults. It provides us to do faster and correct selections by means of its easy user interface (UI). It is better than

Magnetic Lasso and Adobe Photoshop if considered with its kind of specifications. Magnetic Lasso in Adobe Photoshop[11].



(a) Girl (4/2/12)

b) Ballet (4/7/14)

(c) Boy (6/2/13)

(c) Grandpa (4/2/11)

(d) Twins (4/4/12)

Figure 30 Experiments obtained by Lazy Snap method: (a) The girl and her bird is marked as foreground and others in the picture is marked as background. (b) Ballet is separated from the picture by marking by Lazy Snap tool to define the foreground object. (c) Boy is separated by marking in yellow and blue marked places are defined as background. (d) Similarly, grandpa and the boy is marked to separate from the picture. (d) Twins are marked by lazy snap tool to separate from the picture. Remaining part of the picture is background and no important after selection.

This study mentions about a system which is easy to learn compare to other cut tools and easy to do more interactive operations in shorter time. Furthermore, when lazy snap is used on the image, more quality cut graphs are produced compare to the other cut tools. Lazy Snap does the object selection and boundary separation more professionally. Two different user interfaces (UI) are designed for lazy snap. First is designed for selection the object on the image easily. The other one is designed for drawing the boundaries on the object of image in a simpler and faster way. Both interfaces are suitable for controlling or extending the markings on selection area by using pen-computing devices [14].

CHAPTER 4

OUR METHOD

Conversion of a 2D video to 3D video actually means attaching the third dimension information, depth map, to 2D video record. Depth map should be defined for each frame of the video and each pixel of the frames. The depth information flowing in synchronization with video is called depth map. So the main aim of the system is to form this depth map.

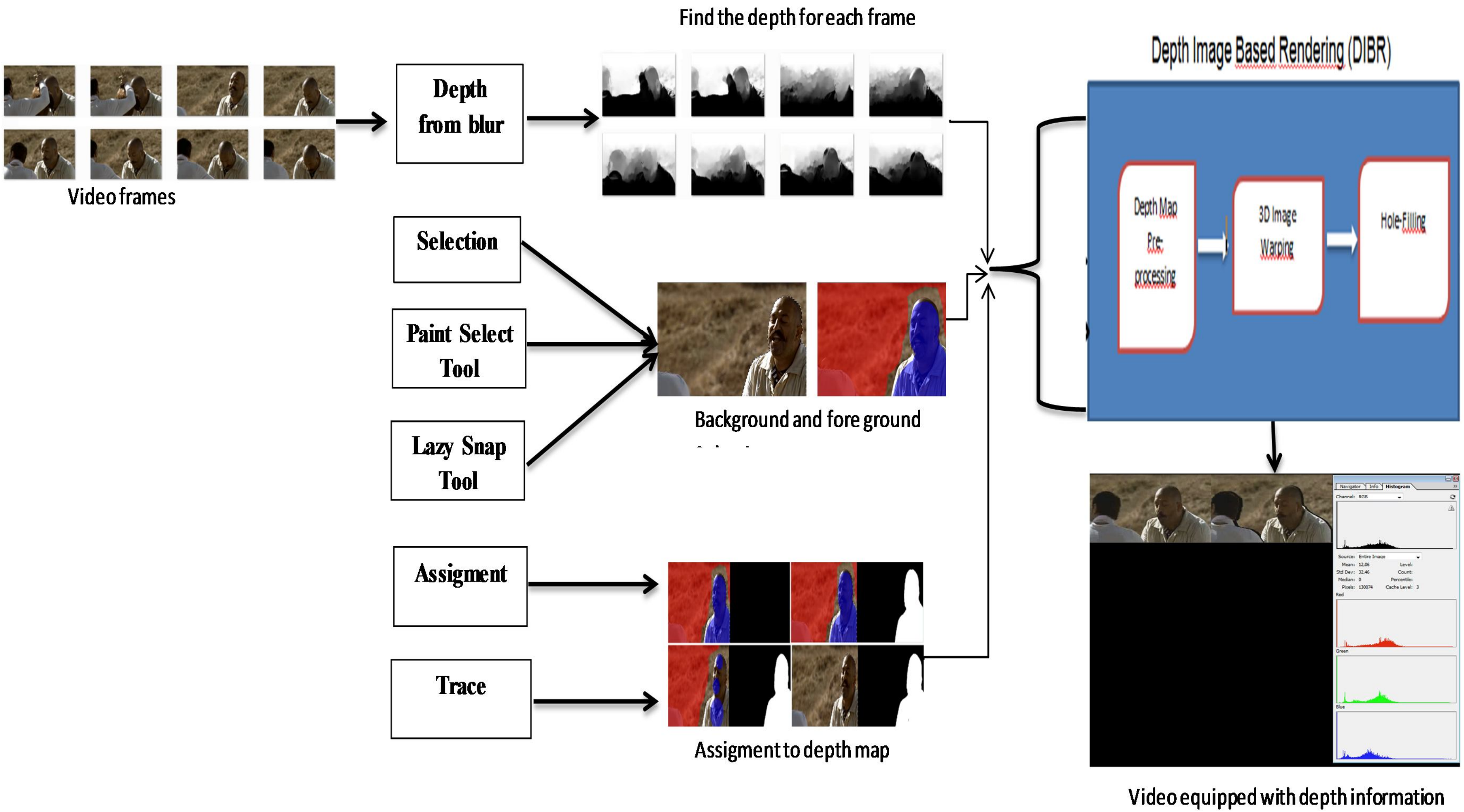


Figure 30 Our 3D conversion method algorithm.

We may examine generating the depth map in three steps:

1. Selection: Selecting the exterior contours of the object to be assigned depth map
2. Assignment: Assigning depth map into that selected area.
3. Tracing: Ensuring the depth map to be spread among the frames properly by tracing the selected object and its depth map in the video to the extent possible.

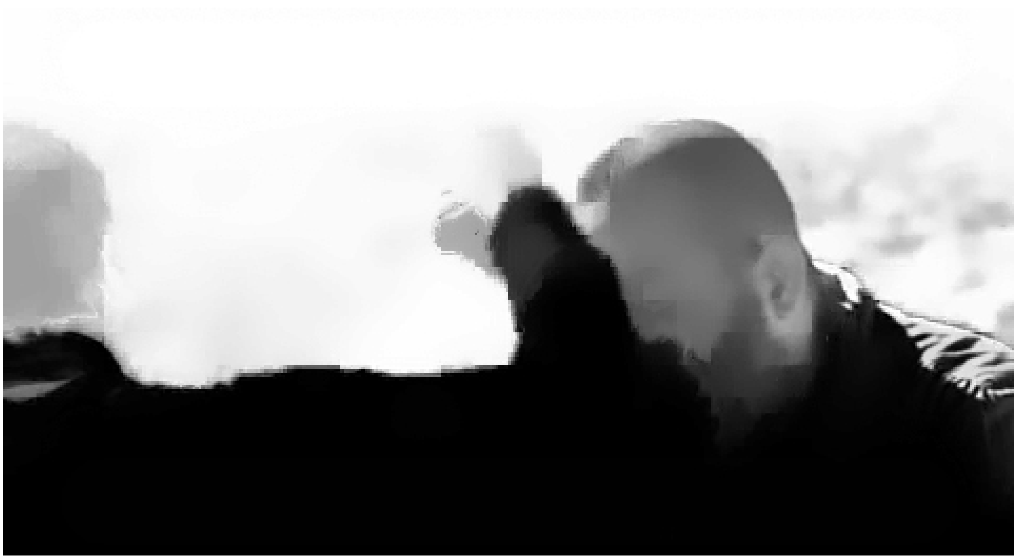


Figure 31 An example of depth map – A frame from "Yahsi Bati"

2D videos have no third dimension, depth as easily understood from its definition. It is necessary to produce the depth information, i.e. depth map for each pixel of each frame in the video to convert these videos into 3D. After producing the depth map, different images for left and right eyes can be produced by using DIBR (depth Image Based Rendering) methods. Therefore, we gave priority to the creation of the depth map in this study.

Currently used methods to convert 2D images to 3D can be divided into two main groups: Manual and automatic conversion.

4.1. Manual Conversion

This method is based on fundamental of being performed of conversion process frame by frame by using general purpose video processing software by a professional.

User chooses the object manually in this method, then copies it to a different frame in parallel and furnishes this are with the necessary depth information. After creation of depth map of a frame, next frame is processed. It usually takes minutes the conversion of a single frame. Considering that a film has averagely 24-25 frames in 1 sec and a film takes about 1.5-2 hours, months are needed to convert it to 3D.

This method is suitable for conversion in required quality as per the customer demand. So it is the most commonly used method in the market. However it requires a very high cost due to the needs of long time and professional labour.

4.2. Automatic Conversion

This method is based on idea of creation of depth information by using some data placed in the video or by exploiting the relationships between the sequent frames.

Some of the commonly used methods to obtain the automatic depth estimation are: Block matching, depth from motion estimation, vanishing point based estimation, depth from motion parallax, blur and binocular parallax.

This method does not require as much professional labour and long-time as manual conversion does. Furthermore, because the facility of manual intervention is limited in automatic conversion process, it may be difficult to form the desired perception and since this process is new, outputs in the desired quality are not produced yet. Therefore the usage area is also limited in the market.

After all the researches, all the methods from very simple to the most difficult ones have been investigated to convert a video to 3D and a new tool as the combination of all these methods have been developed.

Below, the methods to be used to convert 2D images to 3D are described in the examples of the software developed.

4.2.1 Selection

The objects in the videos desired to be converted may have different structures. For example, some of them may have indented shapes whereas some of them may have regular geometrical shapes. In some of the scenes, the objects desired to be selected are separated easily from the background as their colour contrasts are separable; in some others, the contrast ratio is very little. Unfortunately it is not possible to develop a unique selection tool covering all the scenarios in such different types. Therefore, tools to be used for selection need to have same varieties [15].

Six different selection tools have been developed so far for use in the system. While some of them are processed in basic level, some others allow the selection of objects in a smarter way.

Sometimes, even in the course of selection of an object only, usage of more than one tool may be needed. Therefore, the selections done by these selection tools need to be used together. Processes such as to add, subtract and intersect in addition to the existing selections in the systems and forming more complex selections from the simple ones are possible.

All picture tools used in any picture compiling programme are rectangular selection tool, oval selection tool, lariat tool, polygonal lariat tool [16]. However there are two main selection methods setting light to us in this project and they are as follows:

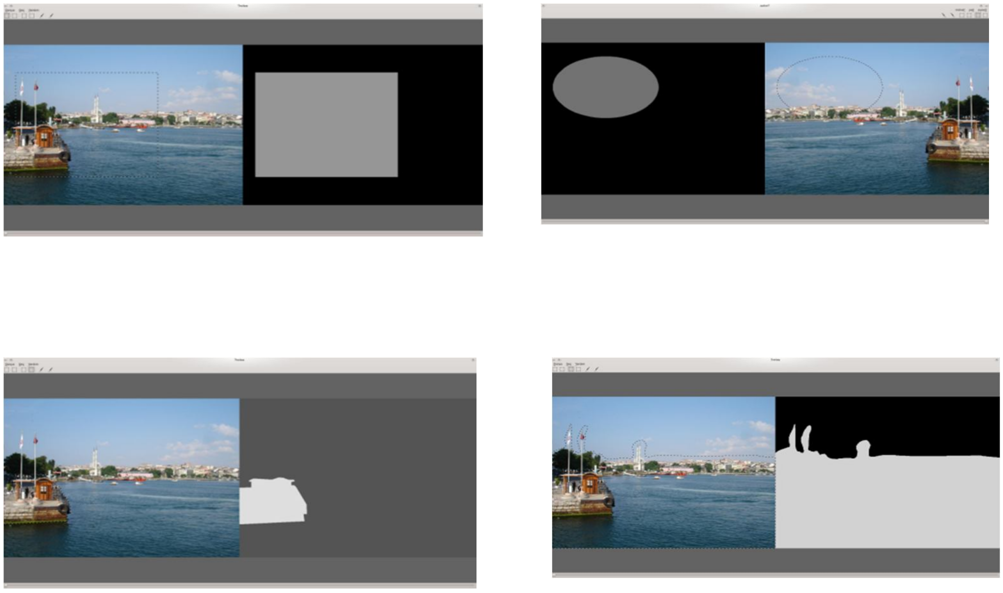


Figure 32 Samples selected by selection tools.

4.2.1.1 Paint select tool

This tool furnish us to click firstly by the mouse to any point of the picture like paint by brush and then to select by dragging the mouse to the required direction. [17]

As the mouse is dragged, it can be placed automatically to the best border line same as in the lazy snap tool.

The working principle of paint select tool is very similar to the quick select tool at Adobe Photoshop. Hence, paint select tool is inspired by the development of quick select tool at Adobe Photoshop.

The paint select tool has a real timed structure which responding immediately as the mouse is dragged. This is the basic difference from the lazy snap tool.

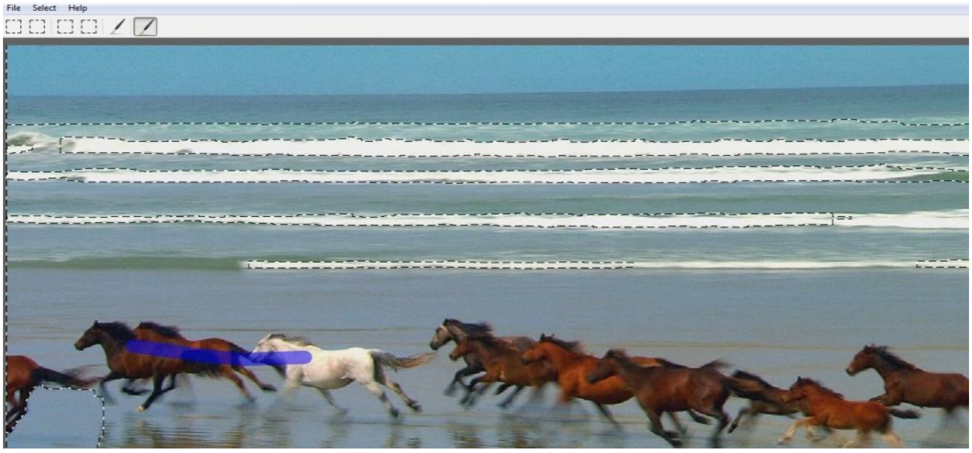


Figure 33 A picture selected by paint select tool.

4.2.1.2 Lazy snap tool

The lazy snap tool is the second smart selected tools.

This tool furnishes us the selection of mixed coloured and complicated shaped objects in the frame easier.

The lazy snap tool's steps are as follows;

- Roughly determination of the object's itself and back ground by using mouse.
- This tool determines the border line of the object automatically by using the position and colour information of signed forefront and back ground.

A sample of lazy snap tool is mentioned as below.

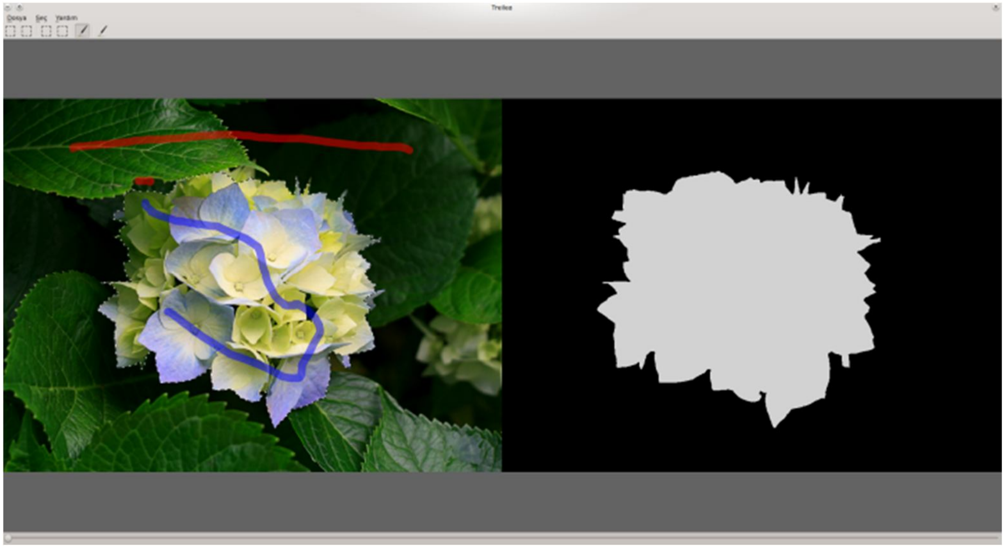


Figure 34 Lazy snap tool

This tool provides us great eases especially for the indented border lined objects whose selection is more difficult by manual and other basic selection tools. This property of lazy snap tool affects the working time automatically. This means, the working time will be lowered to seconds from minutes [18][19].

This tool is the application of the detection algorithm for minimum section splitting a graph into two parts (max-flow min-cut) on to the pictures [20].

The aim of this algorithm is to find the section that split a graph into two parts in condition with the intersection minimum quantity of edge section.

If the density of the edges is different from each other, then the aim is to minimize the total destiny of cut edges.

The full picture assumed as a graph. Each pixel of the picture assumed as a node.

After this assumption, edges are inserted between each node’s and their neighbours located at left, right, bottom, up.

The densities of each edge are determined by the colour difference of two pixels and the distance between two pixels.

After these, two more nodes which are defined as forefront and background will be added to the graph.

One each edge will be added from each nodes of pixels to all the nodes which representative as forefront and background.

The density of these edges will be determined according to the similarities of colour for the pixels signed as forefront and background on the picture.

In the last step of this tool, minimum section detection algorithm is running over the graph in order to find the section. The determination of which pixels will be signed as forefront and which pixels will be signed as back ground will be done regarding to this section.

If the picture is huge sized, the minimum section calculation will take a long time period.

In such cases, in order to get a high performance, the picture will be splitted into small parts by watershed [21] method. After the application of this method, the graph is generated by using these small parts.

In this period; small parts' average colour will be used for the calculation of the pixels difference, centre points of each area will be used for the calculation of distance.

Lazy Snap tool is the most proper tool that is used in our surveys. However there are some deficiencies of lazy snap tool as follows [22];

1. The basic usage of Lazy Snap tool is watershed and this watershed consist of separated and independent frames from each other.
2. Frames independent from each other have difficulties to associate with the pixels level.
3. In the lazy snap tool, there are edges between each node's and their neighbours located at left, right, bottom, up. Graph cut; hence segmentation is occurred by usage of these edges. Because of this reason, the suitable graph cannot prepared in order to get 3D image.
4. In Lazy Snap tool, on-going video flow may have breaks between frames because the preceding and succeeding frames are not matched with each other.

In our study, a connection is established between the frames previously selected by referring to the lazy snap tool. By means of this connection, matching in the pixel level between the preceding and succeeding frames selected.

Converting the used graph into 3D. In lazy snap, each pixel has edges between itself and its upper, bottom, left and right pixels each. So graph cut and segmentation by using these edges [23].

On the video, a graph similar to the lazy snap is formed for each frame. However, in addition to this, each pixel is connected to the matching of itself located in the same coordinates in preceding and succeeding frames.

Then we are running the graph cut algorithm on this 3D graph generated. Thus we mean to obtain a selection providing continuity in the dimensions of x, y, and time. Another change is that we may make signing on any of the selected frames in our version similar to that the foreground and background are marked on the picture in original lazy snap. These markings also become a part of a three-dimensional graph and affect the selection on the frames coming before or after itself [24].

4.2.2. Assignment

Depth information can be assigned to selected objects quickly. It is enough to keep the mouse cursor on the selected area. And the depth information valid for the selected area is increased or decreased quickly by using the scroll of the mouse [25].

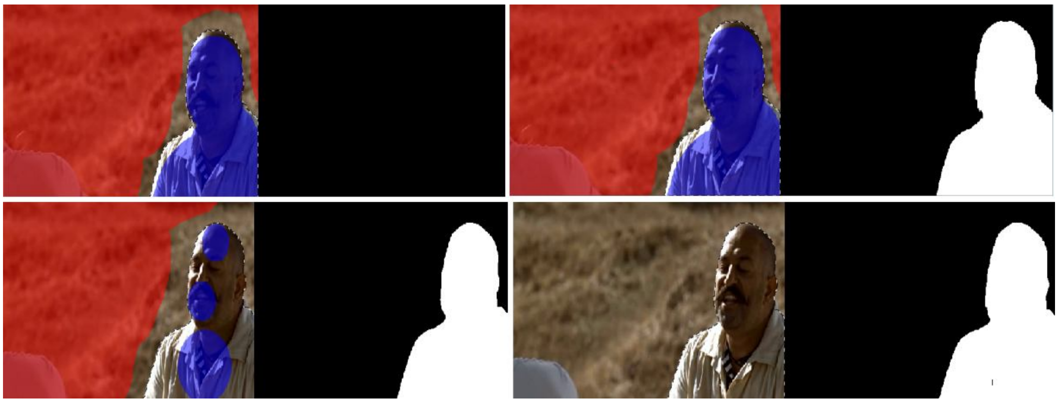


Figure 35 Depth information assignment

4.2.3. Trace

It is seen that some methods such as Active Contours, Mean shift etc. are used for trace. Several tools allowing for selection on more than one frame in the same time have been developed. At the result of the researches, methods called “Video Snap Cut” and “Video Brush” has been seen to allow for this selection [26].

That the selected object is not selected properly by Video Snap Cut need more detailed retouches but the objects selected by Video Brush is selected in more finely detail and with their own real lines[27]. Therefore, it is found more convenient to use Video Brush method basically in the developed programme. By integrating these methods into the system, firstly selection, masquerading, depth assignment and trace are done automatically on any video. [28]. After the object to be traced is selected by Video Brush method from the first frame on, the selected object means to have depth map on the flowing video record through its selected lines. Image can easily be converted to 3D by assigning this map onto the video.

4.3 Proposed Method

The method proposed in this study may be seen as a hybrid method containing some characteristics of each of manual and automatic conversions’ mentioned below. It decreases the time required in a considerable range by enabling the depth map to be transferred between the frames forming the video while enabling the outputs to be produced in the required quality and form.

The first and most important one of the innovations in this study is improving a selection tool running simultaneously on multiple frames to provide the user to do the same. The method developed can perform a detailed selection on tens of frames simultaneously. This selection can be used to assign the depth estimation later.

Selection can be considered as a dual labelling. As the consequence of this labelling, some of the pixels are to be marked as foreground whereas some others are to be marked as background. Graph cut method which is quite popular in recent years and frequently used in image fragmentation is used for such a dual labelling [29][30].

A two dimensional image can be expressed as a graph in a form G=<V, E>. Let us assume here that V shows all the nodes and E shows all the edges connecting the nodes. Generally nodes in an image symbolize the pixels forming that image and edges symbolize the correlations connecting the each pixel to their neighbours. When considered a video; further to these, edges connecting the reciprocal pixels to each other are involved. Labelling problem can be solved by grouping the nodes on the graph as foreground and background [31].

One of the methods to be used for fragmentation of a graph is graph cut algorithm which is quite popular in recent years. Graph cut algorithm is a method which is dividing a graph into two and used to find the fragment of which the sum of weight is minimum.

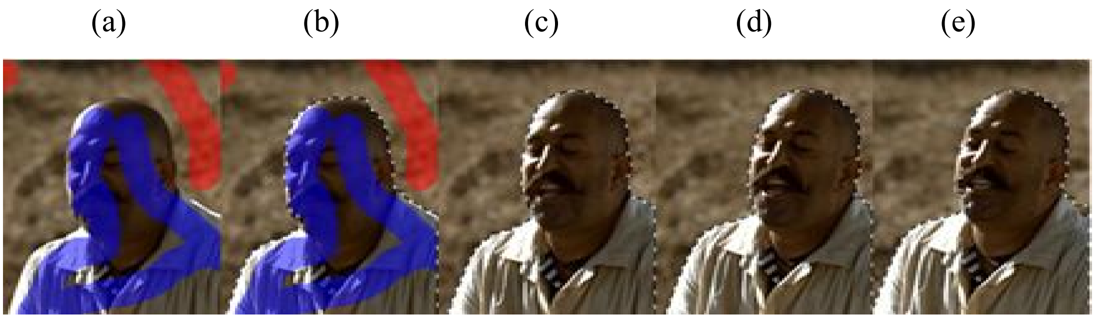


Figure 36 Areas marked as foreground (blue) and background (red) – Pictures (a) and (b) The tracking of the selection found on the basis of user mark-ups through the frames. A frame from Yahsi Batı

The weights of the edges between the nodes play a major role in the graph-cut algorithm. The colour differences between the pixels are used in calculation of the weight of these edges in the proposed method. By assigning higher weights to the edges between the pixels close to each other in colour, selection is provided to be placed automatically onto where colour changes are high. In addition to the nodes expressing the pixels, two special nodes are involved in graph-cut algorithm. These nodes are called as “source” and “sink” respectively. Also special edges symbolizing the correlations of each pixel to the “source” and “sink” nodes are involved in the algorithm. The weights of these edges play a major role to decide for each pixel whether they will be marked as foreground or background. User mark-ups are used to define the weights of these edges as in the Lazy Snap algorithm [19].

When the graph-cut algorithm is applied by using these inputs, continuous and high quality selections are obtained on the video [32].

Another convenience provided in the user interface is facilitating the establishment of the depth map by providing the assignment of depth map in a simple manner onto the selection spread on more than one frames.

5. EXPERIMENTAL RESULT

The output figures obtained by 2D original video, 3D converted by the software and 3D converted manually were compared to each other for the movie “Yahsi Batı” in software Photoshop which defines the outputs as follows:

Below the graph is a full range gradient bar with gray values from white to black. When looked at the result histogram, each line corresponds to one of these values changing between 0 (black) and 255 (white). The height of the lines on the histogram refers to the number of pixels for each value.

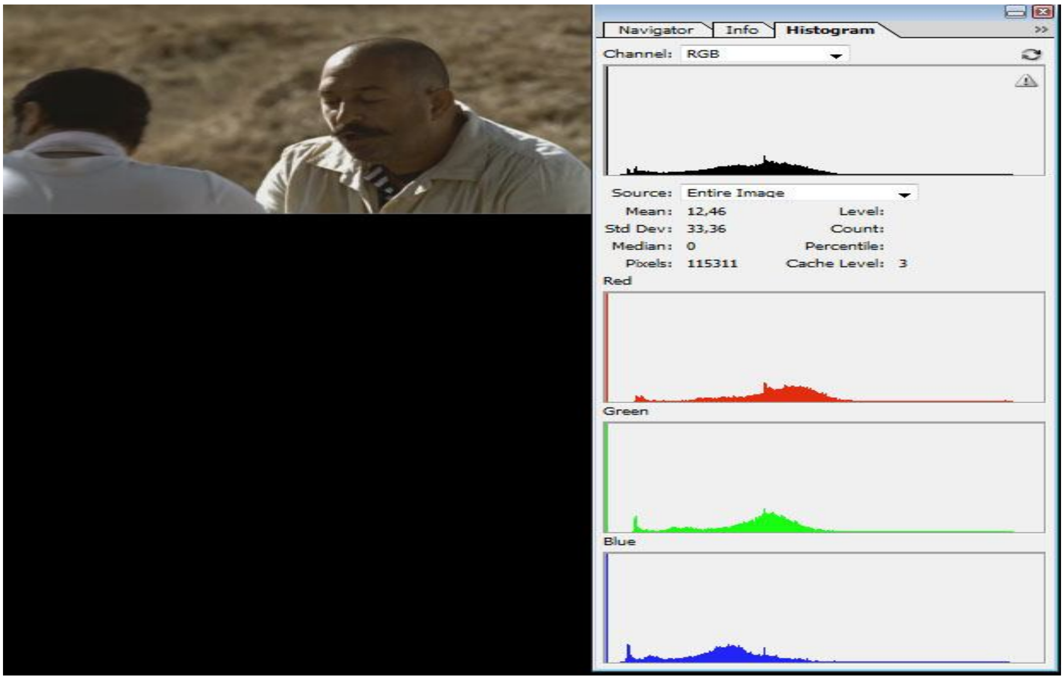
Mean: Average of the whole pixel values being formed between black and white is called MEAN value. This output corresponds to average pixel density.

Median: It corresponds to the pixel value in the mid of the pixels being formed between black and white. (the point where half the pixels are darker and half are lighter).

Standard deviation: The main objective of this statistical term is standardizing the deviations to be occurred during the figuring out the density between two extreme values and calculating the standard deviation of the found values. It corresponds to the standard deviation of the densities of pixels found between black and white in our histogram.

Pixels: Total number of pixels used to represent (calculate) the histogram.

Program Outputs;

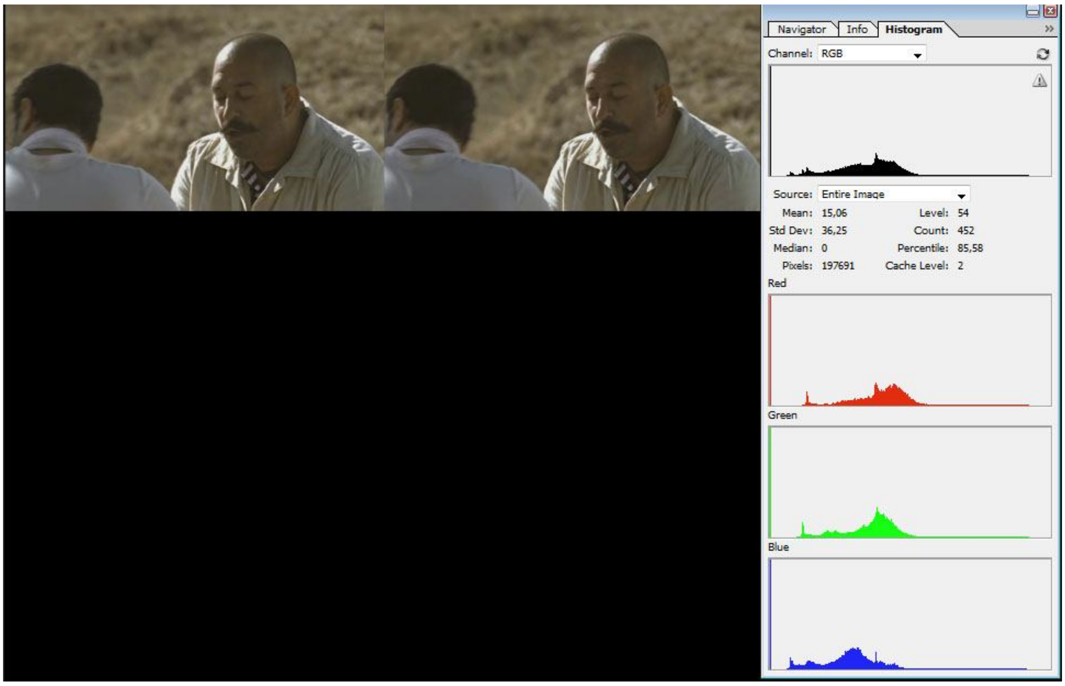


Channel: RGB

Source: Entire Image

```
   Mean: 12,46          Level:
Std Dev: 33,36          Count:
 Median: 0         Percentile:
 Pixels: 115311   Cache Level: 3
```

Figure 37 2D Video (Source Video) Outputs for Yahsi Batı

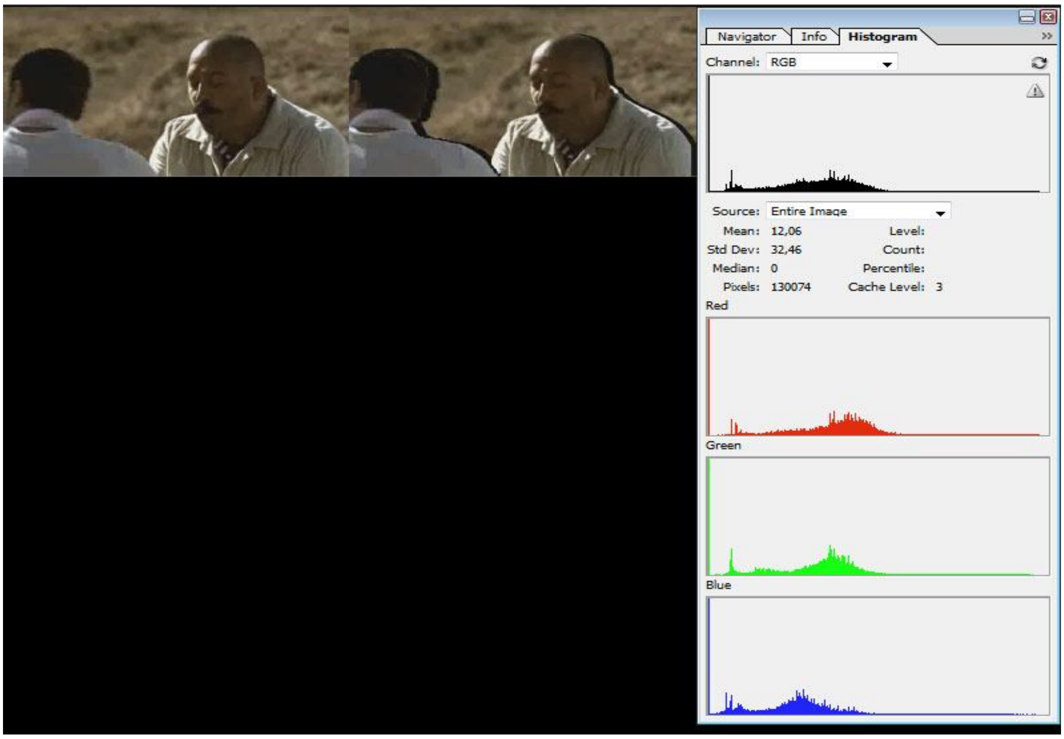


Channel: RGB

Source: Entire Image

```
   Mean: 15,06          Level: 54
Std Dev: 36,25          Count: 452
 Median: 0         Percentile: 85,58
 Pixels: 197691   Cache Level: 2
```

Figure 38 Manuel Conversion Outputs for Yahsi Batı



Channel: RGB

Source: Entire Image

Mean: 12,06
Std Dev: 32,46
Median: 0
Pixels: 130074
Level:
Count:
Percentile:
Cache Level: 3

Figure 39 Outputs for Yahsi Batı converted to 3D by the software

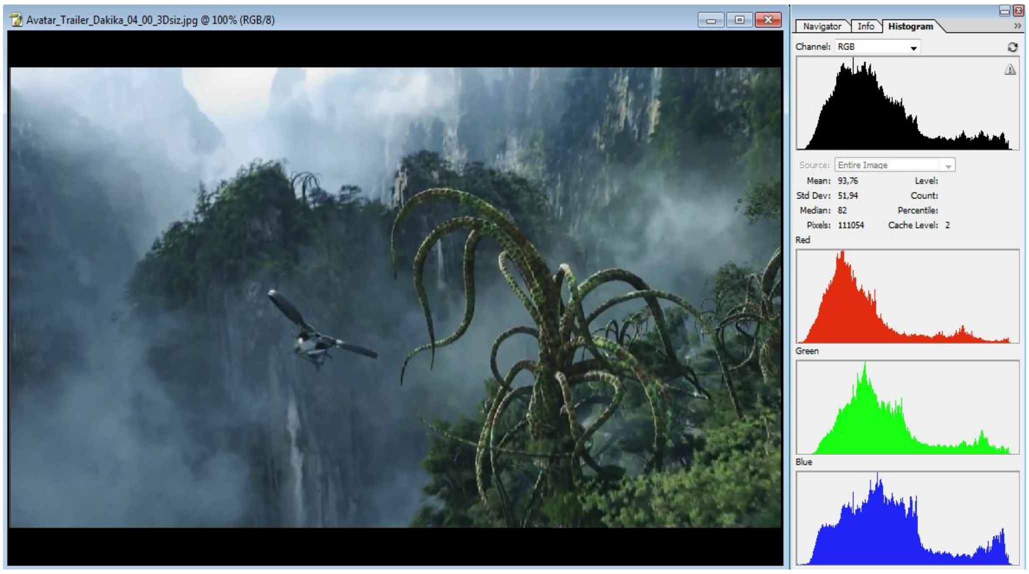
Comparison for the outputs on a table:

	Yahşi Batı				
	2D Real	**3D converted by the software**	**3D Manual**	**3D Manuel / 2D Real**	**3D software/ 2D Real**
Mean	12,46	12,06	15,06	121%	97%
Std Dev	33,36	32,46	36,25	109%	97%
Pixels	115.311	130.074	197.691	171%	113%

There are two types of Avatar videos in the market; 2D and 3D. Since 3D video is already available, no need to make any conversion from 2D to 3D manually. The results in the column 3D Manual/2D Real and 3D software/3D Real of "Yahsi Bati" table are taken to find out the theoretical figures of a 3D Avatar derived by manual and automatic techniques to compare them with the real 3D values:

	Avatar			
	2D Real	**3D Real**	**If manual converted**	**If software Converted**
Mean	93,76	87,36	113,32	90,75
Std Dev	51,94	56,26	56,44	50,54
Pixels	111.054	110.805	190.392,73	125.271,99

2D Real and 3D Real outputs are obtained through the following pictures:

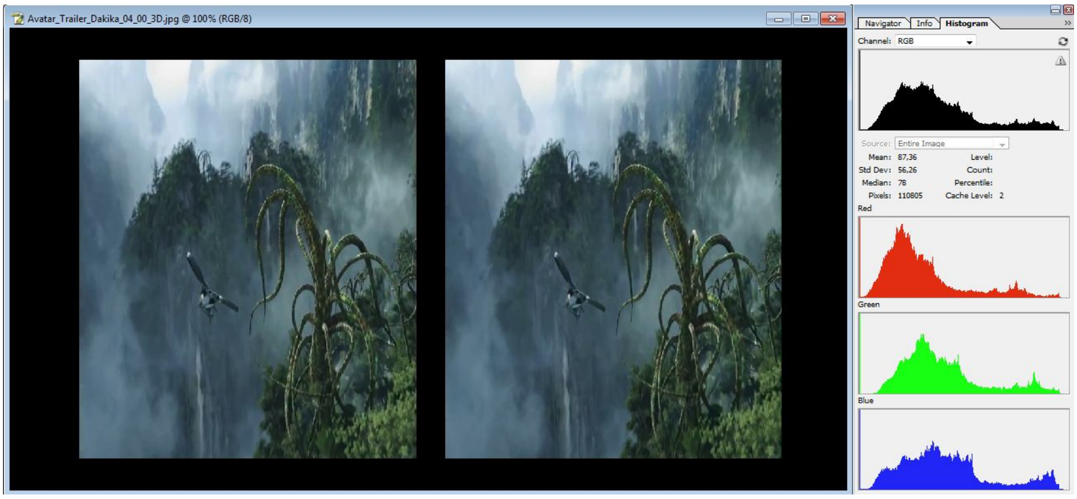

Avatar_Trailer_Dakika_04_00_3Dsiz.jpg @ 100% (RGB/8)

Channel: RGB

Source: Entire Image

Mean: 93,76
Std Dev: 51,94
Median: 82
Pixels: 111054
Level:
Count:
Percentile:
Cache Level: 2

Figure 40 2D Video (Source Video) Output for Avatar



Avatar_Trailer_Dakika_04_00_3D.jpg @ 100% (RGB/8)

Channel: RGB

Source: Entire Image

Mean: 87,36

Std Dev: 56,26

Median: 78

Pixels: 110805

Level:

Count:

Percentile:

Cache Level: 2

Figure 41 3D video Output for Avatar (by the binocular camera)

Here, it is easily seen that manual conversion is the best resolution one in the methods. It is even better than the binocular camera shots. However, bearing the fact that the binocular camera is the nearest one to our real perception and anatomy, we may say that manual conversion and software conversion can be funnier in watching but real 3D movies can be done by binocular cameras only. That's why the other two methods are usually used for animations only for todays and we have not seen a successfully converted 3D video of an old 2D movie yet.

5. CONCLUSION and FUTURE WORK

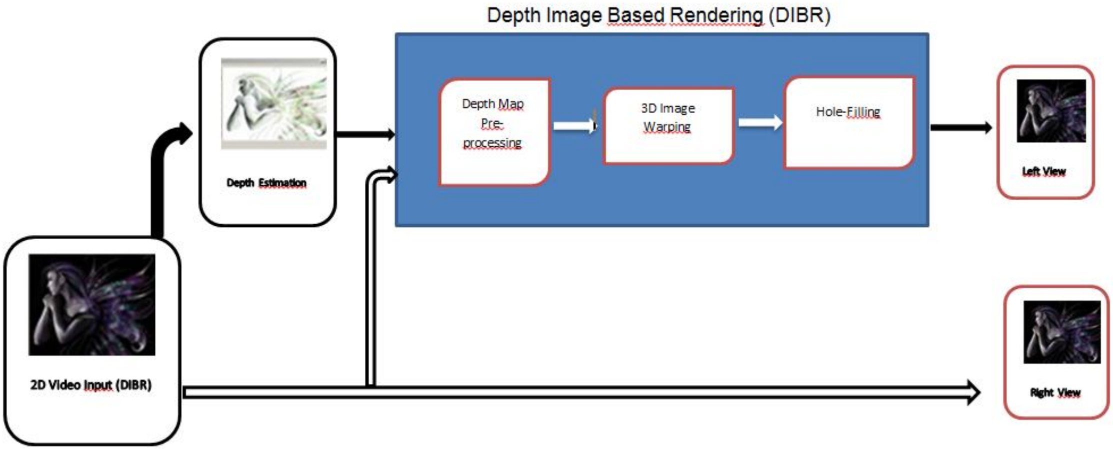


Figure 42 Figure 28- Depth image based rendering

In stereo videos, by above mentioned methods, all images can be converted to 3D by dividing into frames by means of the depth assigned to all moving and stationary objects and applying all processes in the flow diagram.

By present medias, depth map estimation is done by general purpose video processing software. However since these medias are not specific, they are slower than required for depth assignment process. Our aim is to assure the depth estimation to be performed in maximum speed by developing special software.

Stereo images, several conditions must be provided when creating the depth map. The most important are:

- All the pixels in each image can be match with only one pixel.

- Ranking of pixels are valid for also their equals, for image which has fixed object. (A pixel at more left side should be match with a pixel which is located at left side instead of right side.
- A depth map should be changed slowly in the spatial dimension.
- Second image should be defined close to its original using first image which realized with depth map.

Especially the first two conditions which are valid for pixels are provided by Dynamic Programming (DP) [33].
Disadvantage of DP methods is not using regional and neighbourly properties.

Applying condition of pixel to segments will bring on solutions which protect regional properties.
However, providing 4 conditions for segments is a little bit complicated according to pixel based methods.

Basically, providing last conditions for segments means providing the first two conditions of segments. Since shapes of segments are more complex than pixels; it's not very easy to examine ranking and identical conditions. For this purpose; the most important condition is having similar image as its original when shifting provision segments according to depth map.
There are 4 main points during having new image using segments: [34]:

- Brightness value of Area of Moving segments and its area on original image should be low.
- Segments should not be overlapped into their areas.
- There hasn't to be blanks into areas where segments are placed.
- The resulting depth map should be slowly varying at spatial terms and should be smooth.

If that four conditions is satisfied, generated image will be close to fact.

Final depth map can be obtained by reducing a penalty function which has
four conditions of placement of segments by evolving out of depth map [35].

REFERENCES

[1] Stereoscopic Learning for Disparity Estimation **ZHEBIN ZHANG YIZHOU WANG ,TINGTING JIANG, WEN GAO FELLOW**, IEEE Key Lab. of Intelligent Information Processing ,Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China Graduate School, Chinese Academy of Sciences, Beijing, 100039, China National Engineering Lab. for Video Technology, Key Lab. of Machine Perception (MoE), School of EECS, Peking University, Beijing, 100871, China

[2] Retrieved from: 2-3
http://www.cse.unr.edu/~bebis/CS791E/Notes/ImageFormation.pdf (last visit 05.11.2012)

[3] **LAI-MAN PO, XUYUAN XU, YUESHENG ZHU, SHIHANG ZHANG , KWOK-WAI CHEUNG, AND CHI-WANG TING** department of Electronic Engineering, City University of Hong Kong, Kowloon.

[4] Unsupervised object segmentation for 2D to 3D **CONVERSION MATTHIAS KUNTER, SEBASTIAN KNORR, ANDREAS KRUTZ, AND THOMAS SİKORA, İMCUBE MEDIA,** Technische Universität Berlin ,Einsteinufer 17, 10587 Berlin, Germany

[5] **KNORR, S. AND SİKORA, T**, "An Image-based Rendering (IBR) Approach for Realistic Stereo View Synthesis of TV Broadcast Based on Structure From Motion", IEEE Int. Conf. on Image Processing (ICIP), San Antonio, Texas, USA, Sept. 16-19, 2007.

[6] **DONGHYUN KİM, DONGBO MİN, AND KWANGHOON SOHN** Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea

[7] Pacific Graphics 2011 ,**BING-YU CHEN, JAN KUTZ, TONG-YEE LEE, AND MING C. LİN**

[8] **Y. BOYKOV AND V. KOLMOGOROV**. An experimental comparisonof min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(9):1124–1137,Sept. 2004.

[9] **Y. BOYKOV AND M.P. JOLLY.** Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In International Conference on Computer Vision, volume 1, pages 105–112, July 2001.

[10] **LIU J., SUN J., SHUM H.-Y**. : Paint selection. ACM Trans. Graph. 28, 3 (2009), 69:1–7.

[11] Adobe Cooperation :. http://www.adobe.com, 2010.

[12] Lazy Snapping **YIN LI JİAN, SUN CHI, KEUNG TANG HEUNG ,** Hong Kong University of Science and Technology Microsoft Research Asia

[13] An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In Energy Minimization Methods in Computer Vision and Pattern Recognition, 2001.

[14] **VINCENT, L., AND SOILLE** , P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-13, 6 (June), 583–598.

[15] **BLAKE, A., ROTHER, C., BROWN, M., PEREZ, P., AND TORR**, P. 2004. Interactive image s egmentation using an adaptive gmmrf model. In Proceedings of ECCV

[16] **GRADY, L.** 2006. Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 28, 11, 1768–1783.

[17] **WANG, J., AGRAWALA, M., AND COHEN, M. F**. 2007. Soft scissors: an interactive tool for realtime high quality matting. ACM Trans. Graph. 27, 3, 9.

[18] **M. T. M. LAMBOOIJ, W. A. IJSSELSTEIJN AND I. HEYDENRIKX**, “Visual discomfort in stereoscopic displays: a review,” in Stereoscopic Displays and Virtual Reality Systems XIV, 6490 of Proceedings of SPIE, pp. 1–13, San Jose, Calif, USA, (2007).

[19] Lazy **SNAPPING YIN LI, JIAN SUN,CHI-KEUNG TANG ,HEUNG-YEUNG SHUM** Hong Kong University of Science and Technology Microsoft Research Asia http://www.business-sites.philips.com/3dsolutions/

[20] **CHUANG, Y.-Y., CURLESS, B., SALESIN, D. H., AND SZELISKI, R.** 2001. Abayesian approach to digital matting. In Proceedings of CVPR 2001.

[21] **FALCAO, A. X., LOTUFO, R., AND ARAUJO, G.** 2000. The image foresting transformation.In Relatorio Tecnico IC-00-12, 2000.

[22] **AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M**. 2004. Interactive digital photomontage. In Proceedings of ACM SIGGRAPH 2004.

[23] A Scalable Graph-Cut Algorithm for N-D Grids Andrew Delong University of Western Ontario **YURİ BOYKOV** University of Western Ontario

[24] **D. DE SILVA, W.A.C. FERNANDO, H. KODIKARA ARACHCHI**, “A New Mode Selection Technique for Coding Depth Maps of 3D Video”, IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), (2010)

[25] Video SnapCut: Robust Video Object Cutout Using Localized Classifiers **XUE BAI, JUE WANG DAVID SIMONS, GUILLERMO SAPIRO** University of Minnesota 2Adobe Systems

[27] Video Brush: A Novel Interface for Efficient Video Cutout **RUO-FENG TONG, YUN ZHANGY, MENG DING** Institute of Artificial Intelligence, State Key Lab of CAD&CG, Zhejiang University, China

[28] **CHEN T., CHENG M.-M., TAN P., SHAMIR A., HU S.-M**.: Sketch2photo: Internet image montage. ACM Transactions on Graphics 28, 5 (2009), 124:1–10.

[29] Multi-View Reconstruction using Narrow-Band Graph-Cuts and Surface Normal Optimization Alexander Ladikos Selim Benhimane Nassir Navab Chair for Computer Aided Medical Procedures Department of Informatics Technische Universit¨at M¨unchen Boltzmannstr. 3, 85748 Garching, Germany

[30] **B. GOLDLUCKE AND M. MAGNOR**. Spacetime-coherent geometry reconstruction from multiple video streams. In IEEE CVPR, 2004.

[31] **V. KOLMOGOROV AND R. ZABIH**, Computing Visual Correspondence with Occlusions via Graph Cuts, Proc. Int'l Conf. ComputerVision, vol. II, pp. 508-515, 2001.

[32] **ALI KEMAL SINOP AND LEO GRADY** Siemens Corporate Research, Princeton USA Department of Imaging and Visualization

[33] **ZHANG,Y. AND KAMBHAMETTU,C**., "Stereo Matching with Segmentation-Based Cooperation," In Proc. of ECCV 2002

[34] **TAO,H AND SAWHNEY,S.H.,** "Global Matching Criterion and Color Segmentation Based Stereo," Workshop on the Applications of Computer Vision, 2000

[35] http://bj.middlebury.edu/~schar/stereo/data/tsukuba

[36] http://www.ski.org/Vision/Basics/

[37] http://www.canon.com

[38] http://aylincsknn.blogspot.com/2012_08_01_archive.html

[39] http://www.teknoblog.com/

EK-1

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Soyisim, İsim	**: GÖKBULUT, Mehmet Umut**
Uyruğu	**: TC**
Doğum Tarihi ve Yeri	**: 12/11/1982 ANKARA**
Tel	**: 05062939365**
E-Posta	**: umutgokbulut@gmail.com**

EĞİTİM

DERECE	KURUM	MEZUNİYET TARİHİ
Lisans	Anadolu Üniversitesi	2005
Lise	Polatlı Anadolu Ticaret Meslek Lisesi	2001

İŞ DENEYİMİ

YIL	YER	POZİSYON
2005- Halen	YİĞİT AKÜ A.Ş	BilgiTeknolojileri Müdürü

YABANCI DİL

İngilizce – İyi Seviyede

HOBİLER

Ney üflemek

Spor yapmak