A PROJECT REPORT

ON

"IMAGE DETECTION USING STEREOSCOPY"

SUBMITTED TO

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

In Partial Fulfilment of the Requirement of departmental seminar

BY

AKASH PATHAK 12000315005

ARNAB DUTTA 12000315021

UNDER THE GUIDANCE OF

PROF. TAPAS MANDAL

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DR. B.C. ROY ENGINEERING COLLEGE,

Durgapur

2017-2018

Affiliated to

MAKAUT

DR. B.C. ROY ENGINEERING COLLEGE

Department of Electronics and Communication Engineering,

DURGAPUR

CERTIFICATE

This is certify that the project entitled

"IMAGE DETECTION USING STEREOSCOPY"

submitted by

AKASH PATHAK	**12000315005**
ARNAB DUTTA	**12000315021**

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Electronics and Communication Engineering) at *Dr. B.C Roy Engineering College*, Durgapur under the **MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY**. This work is done during year 2018-2019, under our guidance.

Date: **10 /05 /2018**

(*Prof.* **TAPAS MANDAL**)
PROJECT MENTOR

(*Prof.* **NARENDRA NATH PATHAK**)
HEAD OF DEPARTMENT,ECE

ACKNOWLEDGEMENTS

We are profoundly grateful to ***Prof.*** **TAPAS MANDAL** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

We would like to express deepest appreciation towards ***Prof.*** **Dr. Narendra Nath Pathak**, HEAD OF DEPARTMENT of Electronics and Communication Engineering whose invaluable guidance supported us in completing this project.

Finally, we would like to express thanks to my groupmates for their sincerity and helpfulness throughout, DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING of **Dr. B. C. Roy Engineering College**, Durgapur for providing moral and resourceful support for the completion of this report.

AKASH PATHAK
ARNAB DUTTA

ABSTRACT

A stereoscopic motion or still picture in which the right component of a composite image usually red in color is superposed on the left component in a contrasting color to produce a *three-dimensional effect* when viewed through correspondingly colored filters in the form of spectacles.

The modes of 3D presentation you are most familiar with are the paper glasses with red and blue lenses. The technology behind 3D, or stereoscopic, movies is actually pretty simple. They simply recreate the way humans see normally.

Since your eyes are about two inches apart, they see the same picture from slightly different angles. Your brain then *correlates these two images* in order to gauge distance. This is called *binocular vision*, this process by presenting each eye with a slightly different image.

The binocular vision system relies on the fact that our two eyes are spaced about 2 inches (5 centimeters) apart. Therefore, each eye sees the world from a slightly different perspective, and the binocular vision system in your brain uses the difference to calculate distance. Your brain has the ability to correlate the images it sees in its two eyes even though they are slightly different.

The cross-eyed viewing method swaps the left and right eye images so that they will be correctly seen cross-eyed, the left eye viewing the image on the right and vice versa. The fused three-dimensional image appears to be smaller and closer than the actual images, so that large objects and scenes appear miniaturized. *This method is usually easier for freeviewing novices*. As an aid to fusion, a fingertip can be placed just below the division between the two images, then slowly brought straight toward the viewer's eyes, keeping the eyes directed at the fingertip; at a certain distance, a fused three-dimensional image should seem to be hovering just above the finger.

Keywords: Three-dimensional effect, binocular vision, depth.

Contents

List of Figures

Chapter 1

Introduction

Stereoscopy (also called stereoscopics, or stereo imaging) is a technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision[2]. The word stereoscopy derives from Greek (stereos), meaning 'firm, solid', and (skope), meaning 'to look, to see'.

Any stereoscopic image is called a stereogram. Originally, stereogram referred to a pair of stereo images which could be viewed using a stereoscope.

Most stereoscopic methods present two offset images separately to the left and right eye of the viewer. These two-dimensional images are then combined in the brain to give the perception of 3D depth. This technique is distinguished from 3D displays that display an image in three full dimensions, allowing the observer to increase information about the 3-dimensional objects being displayed by head and eye movements.

Figure 1.1: STEREOSCOPIC IMAGE

Chapter 2

History

In *280 A.D., Euclid* was the first to recognize that depth perception is obtained when each eye simultaneously receives one of two dissimilar images of the same object. In 1584 **Leonado da Vinci** studied the perception of depth and, unlike most of contemporaries, produced paintings and sketches that showed a clear understanding of shading, texture and viewpoint projection.Around the year 1600, **Giovanni Battista della Porta** produced the first artificial 3-D drawing based on Euclids notions on how 3-D perception by humans works.

Queen Victoria visited the World's Fair in London in 1851 and was so entranced by the stereoscopes on display that she precipitated an enthusiasm for three-dimensional photography that soon made it a popular form of entertainment world-wide.

It was **Sir Charles Wheatstone** who in 1833 first came up with the idea of presenting slightly different images to the two eyes using a device he called a reflecting mirror stereoscope. When viewed stereoscopically, he showed that the two images are combined in the brain to produce 3-D depth perception. The invention of the Brewster Stereoscope by the Scottish scientist **Sir David Brewster** in 1849 provided a template for all later stereoscopes. This in turn stimulated the mass production of stereo photography which flourished alongside *mono-photography*. Stereo photography peaked around the turn of the century and went out of fashion as movies increased in popularity.

Chapter 3

Features of an image

3.1 Image and Video

An image is visual representation of a data. While a video can be defined as a collaboration of images. When a particular *frame* is fixed and captured in different time instances, it merges to create a video.

3.2 Pixels

Image consists of a matrix of pixels, a tuple which contains values of **Red, Green and Blue** ranging between (0-255). The combination of values R,G,B determine the colour of the pixel. When video is considered apart from this three, a fourth variable *time* comes in consideration.

3.3 Grayscale and binary

To compress an Image, *to reduce the bandwidth*, Image is converted into grayscale, or the form where it loses its R,G,B combination and turns black and white. But the *light intensity* is present there, depending upon the intensity regions can be separated in image. This is not possible in binary, as it is completely hindering the color properties, it is *pure black or white* there is no space for intensity.

3.4 Object detection

Every object class has its own special features that helps in classifying, having unique color or shape. Using those features the object can be extracted.

Chapter 4

Working

Most human beings use what is known as binocular vision to perceive depth and see the world in 3D. The binocular vision system relies on the fact that we have two eyes, which are approximately 3 in apart. This separation causes each eye to see the world from a slightly different perspective. The brain fuses these two views together. It understands the differences and uses them to calculate distance creating our sense of depth and ability to gauge distance.

If you've ever used a View-Master or a stereoscopic viewer, you have seen your binocular vision system in action. In a View-Master, each eye is presented with an image. Two cameras photograph the same image from slightly different positions to create these images. Your eyes can then correlate these images automatically because each eye sees only one of the images.

Special kind of Stereo cameras or stereo lenses can be used, which helps in implementing the vision more precisely.

(a) VERASCOPE 40

(b) FEDSTEREO

Figure 4.1: STEREO CAMERAS

Chapter 5

Ways to View Stereoscopic Images

There are many ways to view Stereoscopic Images -

5.1 Stereo Pairs

Typical stereo pair images are two separate images of the same object taken a few inches apart. In this method, the two images are not interlaced but rather presented side by side (left eye image on left and right eye image on right). The images are directly viewable using parallel ”free-viewing” glasses which allow each eye to only see its corresponding image.

5.2 LCD Shutter Glass Method

In the LCD shutter glass 3D display, the left and right images are alternated rapidly on the monitor screen. When the viewer looks at the screen through shuttering eyewear, each shutter is synchronized to occlude the unwanted image and transmit the wanted image. Thus each eye sees only its appropriate perspective view. The left eye sees only the left view, and the right eye only the right view.

5.3 Color Filter Glasses

Color filter glasses are one of the oldest methods of viewing 3D images or movies. The system works by feeding different images into your eyes. The different color filters allow only one of the images to enter each eye, and your brain does the rest. There are two color filter systems: Red/Blue and Red/Green.

5.4 Polarizing Glasses

This method is more commonly used in today's 3D movie projections. The audience must wear special glasses which have two polarizing lenses which have their polarization directions adjusted to be 90 degrees different. This makes is possible that left eye sees it's picture without problems but everything ment to right eye (sent out at different polarization) seems to be black. Same applies also to right eye.

Figure 5.1: STEREO GLASS MODULE

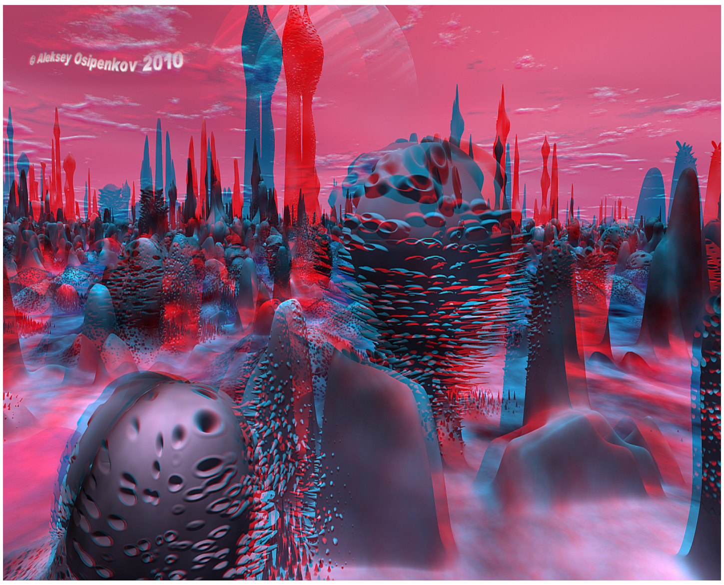


Figure 5.2: IMAGE WITH 3D PERCEPTION

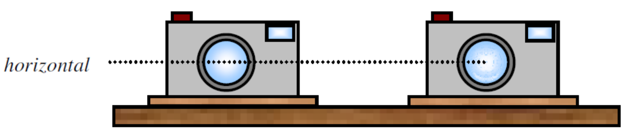
Chapter 6

Implementation

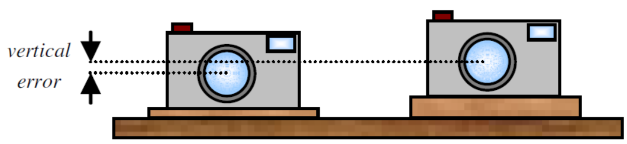
Stereoscopy is a technique used for recording and representing stereoscopic (3D) images. It can create an illusion of depth using two pictures taken at slightly different positions.

Stereoscopic picture can be taken with a pair of cameras similarly to our own eyes. The most important restrictions in taking a pair of stereoscopic pictures are the following:

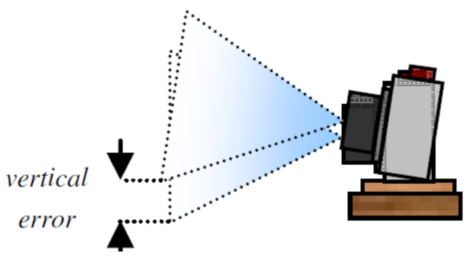
- cameras should be horizontally aligned (see Figure 1), and
- the pictures should be taken at the same instant.



(a) HORIZANTAL SETUP



(b) VERTICAL SETUP



(c) ANGLE SETUP

Figure 6.1: STEREO CAMERA SETUP

Stereoscopic pictures allow us to calculate the distance between the camera(s) and the chosen object within the picture. Let the right picture be taken in location SR and the left picture in location SL. B represents the distance between the cameras and is cameras horizontal angle of view. Objects position (distance D) can be calculated by doing some geometrical derivations. We can express distance B as a sum of distances B_1 and B_2:

$$B = B_1 + B_2 = D \tan \varphi_1 + D \tan \varphi_2 \tag{6.1}$$

if optical axes of the cameras are parallel, where φ_1 and φ_2 are angles between optical axis of camera lens and the chosen object. Distance D is as follows:

$$D = \frac{B}{\tan \varphi_1 + \tan \varphi_2} \tag{6.2}$$

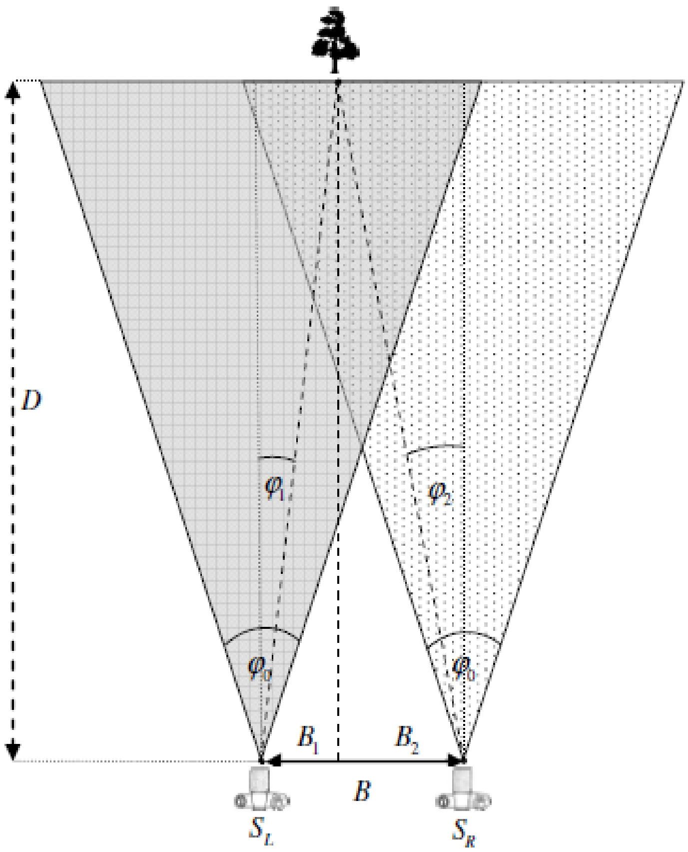
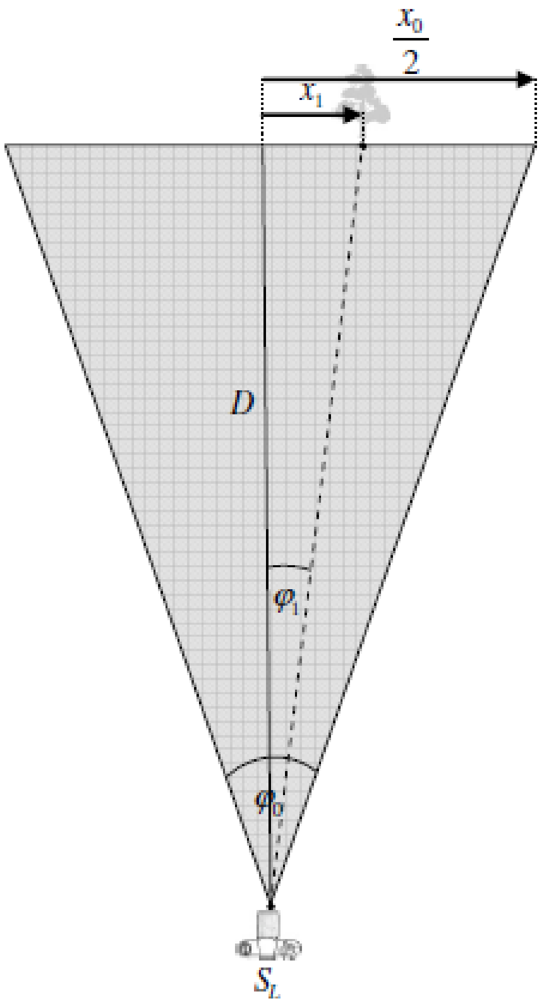
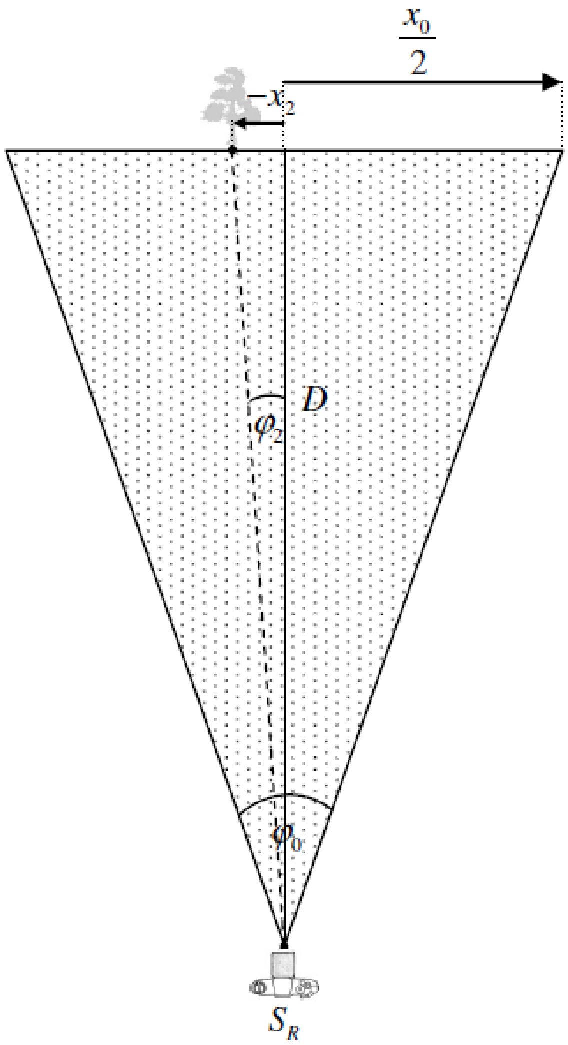


Figure 6.2: OBJECT CAPTURED BY STEREO SETUP

This figure shows the setup installed to capture Image of a *tree*, the two cameras used(S_L and S_R) are similar in aspect of their **Field of View**, **Focal length**.



(a) OBJECT THROUGH LEFT CAMERA



(b) OBJECT THROUGH RIGHT CAMERA

Figure 6.3: INDIVIDUAL CAMERA WORKING

The two separate images, give rise to two different equations as follows:

$$\frac{x_1}{\frac{x_0}{2}} = \frac{\tan\varphi_1}{\tan\frac{\varphi_0}{2}} \tag{6.3}$$

$$\frac{-x_2}{\frac{x_0}{2}} = \frac{\tan\varphi_2}{\tan\frac{\varphi_0}{2}} \tag{6.4}$$

Putting the deduced values of φ_1 and φ_2 in the principle equation gives an resultant form of:

$$D = \frac{Bx_0}{2\tan(\frac{\varphi_0}{2})(x_1 + x_2)} \tag{6.5}$$

Note: Here $(x_1 + x_2)$ is represented in the form of $(x_L - x_D)$ for ease of understanding.

Therefore, if the distance between the cameras (B), number of horizontal pixels (x_0), the viewing angle of the camera(φ_0) and the horizontal difference between the same object on both pictures $(x_L - x_D)$ are known, then the distance to the object (D) can be calculated as given in final expression. The accuracy of the calculated position (distance D) depends on several variables. Location of the object in the right picture can be found within accuracy of one pixel. Each pixel corresponds to the following angle of view:

$$\delta = \frac{\varphi_0}{x_0} \tag{6.6}$$

one pixel angle of view $\delta\varphi$ results in distance error δD. In image we find:

$$\frac{\tan\varphi}{\tan(\varphi - \delta\varphi)} = \frac{\delta D + D}{D} \tag{6.7}$$

Using basic trigonometry identities, the distance error can be expressed as follows:

$$\delta D = \frac{D^2}{B}\tan(\delta\varphi) \tag{6.8}$$

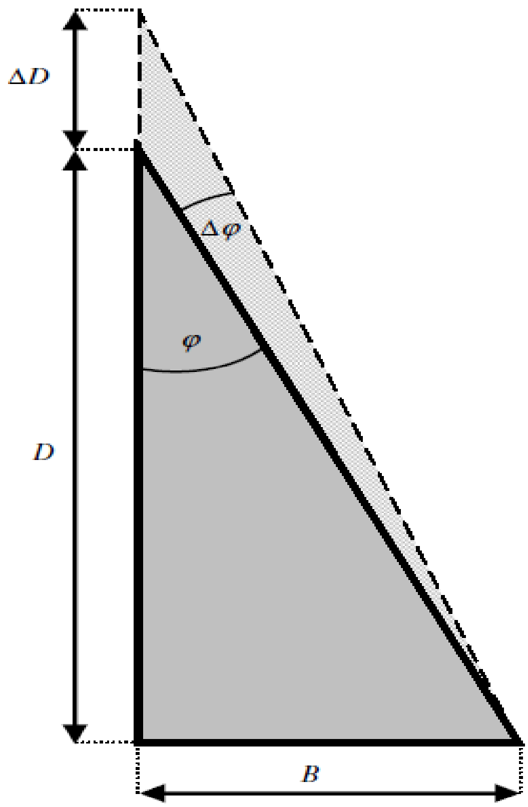


Figure 6.4: DISTANCE ERROR CAUSED BY 1 PIXEL

Chapter 7

Object Recognition

When the certain object is selected on the left picture, the same object on the right picture has to be automatically located. Let the square matrix I_L represent selected object on the left picture. Knowing the properties of stereoscopic pictures (the objects are horizontally shifted), we can define search area within the right picture matrix I_R. Vertical dimensions of I_R and I_L should be the same, while horizontal dimension of I_R should be higher.

Our next step is to find the location within the search area IR, where the picture best fits matrix I_L. We do that by subtracting matrix I_L from all possible submatrixes (in size of matrix I_L) within the matrix I_R. When size of the matrix IL is NxN and size of the matrix I_R is MxN, where $M > N$, $M - N + 1$ submatrixes have to be checked.

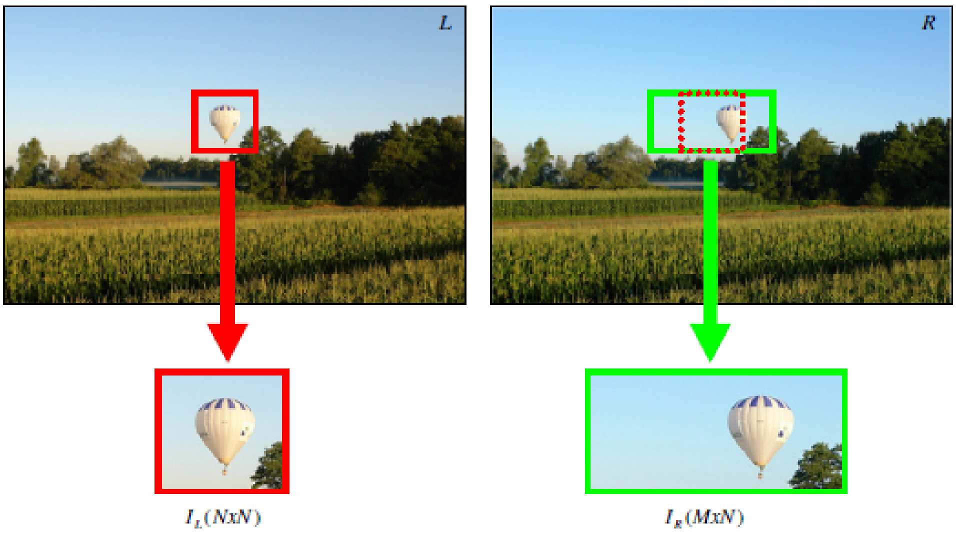


Figure 7.1: SELECTED OBJECT ON LEFT PICTURE AND SEARCH AREA ON RIGHT PICTURE.

The result of each subtraction is another matrix Ii which tells us how similar subtracted images are. More similar two subtracted images are, lower is the mean absolute value of matrix I_i.

Images in the first example do not match, while in the second example images match almost perfectly.

The matrix I_K with the lowest mean of its elements, represents the location where matrix I_L best fits to matrix I_R. This is also the location of the chosen object within the right image. Knowing the location of the object in the left and right image allow us to calculate the distance between the pictures:

$$x_L - x_D = \frac{N}{2} + K - 1 - \frac{M}{2} \tag{7.1}$$

The distance $(x_L - x_D)$ is the most vital and a bit complex in calculation. This value determines the amount of match between frames captured from both cameras. Rest of the values are pre-defined and constant, the only value varying is $(x_L - x_D)$. Our work is in finding this value and in turn calculating value of D, which gives the depth.

Chapter 8

Tools and Libraries

OpenCV is an open-source library for computer vision. It is used for capturing image, and extract features from it which ultimately leads to feature extraction and object detection.

Grayscale and thresholding of image is also done using OpenCV. Also capping by colour, Canny edge detection, erosion and dilation, contour search, and moments (to find centre of frame) are implemented using CV.

Figure 8.1: OpenCV

Python language is used along with numpy distribution for numerical computation. Also Scipy is used for final distance/depth calculation.

Figure 8.2: Numpy & Python

Chapter 9

Algorithm & Code

9.1 Importing modules

```
from scipy.spatial import distance as dist
import numpy as np
import cv2
```

This are the distributions to be used.

9.2 Capture and Work on image

```
cap0=cv2.VideoCapture(0)
while(1):
  ret,img1=cap0.read()
  image_blur = cv2.GaussianBlur(img1, (7, 7), 0)
  image_blur_hsv1 = cv2.cvtColor(image_blur, cv2.COLOR_BGR2RGB)
```

In this section, OpenCV is capturing live feed from HDcam, and the individuals frames are extracted, Gaussian Blur filter is applied and the image format is changed from BGR to HSV.

9.3 Colour grading

```
max_orange = np.array([245,136,99])
min_orange = np.array([150,45,10])
max_white= np.array([225,224,228])
min_white=np.array([119,115,136])
mask1 = cv2.inRange(image_blur_hsv1,min_orange,max_orange)
mask2 = cv2.inRange(image_blur_hsv1,min_white,max_white)
mask=mask1+mask2
```

In this section, OpenCV is thresholding orange an white color, with some predefined values of R,G,B combination. The mask is created and applied on the blurred image.

9.4 Canny edge and Contour

```
edged1 = cv2.Canny(mask, 50, 100)
edged1 = cv2.dilate(edged1, None, iterations=1)
edged1 = cv2.erode(edged1, None, iterations=1)

 _,contours1,_ = cv2.findContours(edged1, cv2.RETR_TREE,cv2.CHAIN_APPROX_SIMPLE)
```

Here, Canny edge detection done and image dilated and eroded for finding features and then contours are scratched out by edges.

9.5 Finding co-ordinates

```
M1=cv2.moments(contours1)
 if M1["m00"] != 0:
   cX1=int(M1['m10']/M1["m00"])
   cY1=int(M1["m01"]/M1["m00"])
   cv2.drawContours(img1,c1, -1, (0, 255, 0), 2)
   cv2.circle(img1, (cX1, cY1), 7, (255, 255, 255), -1)
   cv2.putText(img1, "center", (cX1 - 20, cY1 - 20),cv2.FONT_HERSHEY_SIMPLEX,
       0.5, (255, 255, 255), 2)
 else:
   cX1,cY1=0,0
   c2 = max(contours2, key = cv2.contourArea)
```

Here the co-ordinate of the detected image is obtained, and the distance of object from the center of frame is determined.

9.6 Finding co-ordinates

```
 D=cX1+cX2
```

Finally the depth is calculated as sum of C_{x1} and C_{x2} which are the distance returning $x_L - x_D$.

Chapter 10

Conclusion and Future Scope

10.1 Conclusion

This distance measurement is based upon the pictures taken from two horizontally displaced cameras. The user should select the object on left camera and the algorithm finds similar object on the right camera. From displacement of the same object on both pictures, the distance to the object can be calculated. Although the method is based on relatively simple algorithm, the calculated distance is still quite accurate. Better results are obtained with wider base (distance between the cameras). This is all according to theoretical derivations.

- Visible light is used as mode of communication, hence no hustle of Radio Frequencies.
- This is an accurate way of calculating depth of object in passive manner.
- It has wide range of application, from medical science to research.
- Now a days, it is implemented in Smartphone cameras to improvise autofocus.

10.2 Future Scope

Recent advances in 3-dimensional (3-D) stereoscopic imaging have enabled 3-D display technologies in the operating room. We find 2 beneficial applications for the inclusion of 3-D imaging in clinical practice. The first is the real-time 3-D display in the surgical theater, which is useful for the neurosurgeon and observers. In surgery, a 3-D display can include a cutting-edge mixed-mode graphic overlay for image-guided surgery. The second application is to improve the training of residents and observers in neurosurgical techniques. This article documents the requirements of both applications for a 3-D system in the operating room and for clinical neurosurgical training, followed by a discussion of the strengths and weaknesses of the current and emerging 3-D display technologies. An important comparison between a new autostereoscopic display without glasses and current stereo display with glasses improves our understanding of the best applications for 3-D in neurosurgery. Today's multiview autostereoscopic display has 3 major benefits: It does not require glasses for viewing; it allows multiple views; and it improves the workflow for image-guided surgery registration and overlay tasks because of its depth-rendering format and tools. Two current limitations of the autostereoscopic display are that resolution is reduced and depth can be perceived as too shallow in some cases. Higher-resolution displays will be available soon, and the algorithms for depth inference from stereo can be improved. The stereoscopic and autostereoscopic systems from microscope cameras to displays were compared by the use of recorded and live content from surgery. To the best of our knowledge, this is the first report of application of autostereoscopy in neurosurgery.

References

[1] M. Vidmar, ABC sodobne stereofotografije z maloslikovnimi kamerami, Cetera, 1998.
[2] H. Walcher, Position sensing Angle and distance measurement for engineers, Second edition, Butterworth- Heinemann Ltd., 1994.
[3] D. Vrancic and S. L. Smith, Permanent synchronization of camcorders via LANC protocol, Stereoscopic displays and virtual reality systems XIII : 16-19 January, 2006, San Jose, California, USA, (SPIE, vol. 6055).
[4] Navigation 2, Radio Navigation, Revised edition, Oxford Aviation Training, 2004.
[5] P. Gedei, Ena in ena je tri, Monitor, july 2006, http://www.monitor.si/clanek/ena-in-ena-je-tri/, (21.12.2007).
[6] J. Carnicelli, Stereo vision: measuring object distance using pixel offset, http://www.alexandria.nu/ai/blog/entry.asp?E=32, (29.11.2007).
[7] A. Klein, Questions and answers about stereoscopy, http://www.stereoscopy.com/faq/index.html, (28.1.2007).
[8] A. Woods, Image distortions and stereoscopic video systems, http://www.3d.curtin.edu.au/spie93pa.html, (27.1.2007).