CHAPTER 1

INTRODUCTION

1.1 Background

Object detection and learning is brought about by certain machine learning and deep learning algorithms. Acquiring the training dataset requires a lot of efforts in the major computer science domain region. This project aims in minimizing the efforts for object detection and learning. Instead of following the conventional method of training the dataset using the deep learning techniques, a contemporary human-computer interaction system is used for dataset acquisition and training process. The human-computer interaction system involves the learning and localization of objects using eye tracking or hand region segmentation process.

1.2 Problem Statement

Aims at initiating a system of human-computer interaction which can simplify the task of object training using hand region segmentation or eye gaze data. The acquiring of training dataset and learning process in object detection is achieved in a simpler way by human-computer interaction system.

1.3 Specific Objectives

The objectives of this project include:

- Extracting region of interest using human-computer interaction
- Object learning and detection using machine learning algorithms
- Object localization using learned properties

The underlying challenge includes in creating a hybrid system incorporating both eye movement and hand region segmentation.

1.4 Applications

This work, can help supermarkets to make offers available for certain products. With eye tracker devices, it can detect eyes of shoppers and then process a datasheet based on what the person is gazing at. This eye tracking datasheet processed helps in detecting the object using machine learning algorithms. With the eye tracker giving information about regions of interest, object can be learnt and can be used find out which one draws more attention of people. (Refer to Fig 1.1)

Fig 1.1 Shopper's area of interest (Src: Google Images)

Some of the major applications of simplifying the object learning process includes surveillance, medical science and robotic cognition. It requires training of robots for object detection. This robotic cognition process may involve high level CPU consumption. In order to simplify, human-computer interaction is being used.

CHAPTER 2
LITERATURE SURVEY

This chapter briefs about the existing methods on various object detection methodologies. The following sections deal with Human Computer Interaction (Section 2.1), Eye Tracking Process (Section 2.2) and Hand Region Segmentation (Section 2.3)

2.1 Human Computer Interaction

The paper titled "Behavior Evolution of Pet Robots with Human Interaction" [1]by Hiroyuki and Hifumi gives an insight into the challenges and methods of HCI. One vital HCI factor is that diverse clients structure distinctive originations or mental models about their connections and have diverse methods for learning and keeping information and aptitudes. Another thought in contemplating or planning HCI is that UI(User Interface) innovation changes quickly, offering new cooperation conceivable outcomes to which past research discoveries may not matter. As of late, different robots have been utilized in human life. Consequently, it is fundamental that robots learn alluring conduct so as to set up a real existence of the co-activity among robots and humans (Fig 2.1). In this paper, we propose strategies utilizing intuitive hereditary calculations so that robots get the conduct that clients like.

Fig 2.1 Robots in HCI(Src:Google Images)

2.2 Eye Tracking Process:

The paper titled "A Survey of Eye Tracking Methods and Applications By Robert Gabriel Lupu and Florina Ungureanu"[2] gives an overview of eye tracking features. An eye tracking system is in light of a gadget to follow the movement of the eyes to know precisely where the individual is looking and for to what extent. It additionally provides programming calculations for students, picture preparing, information shifting and recording eye d by methods for fixation point, fixation length and saccade. A huge assortment of equipment and programming approaches were executed by research organizations as per mechanical advancement. The following features are mainstream in the eye tracking process:

Fixation − the time taken for processing image by fovea;

Saccade – the time taken by fovea to focus its attention from one image to another (time interval between two fixations),eye following is a strategy whereby the

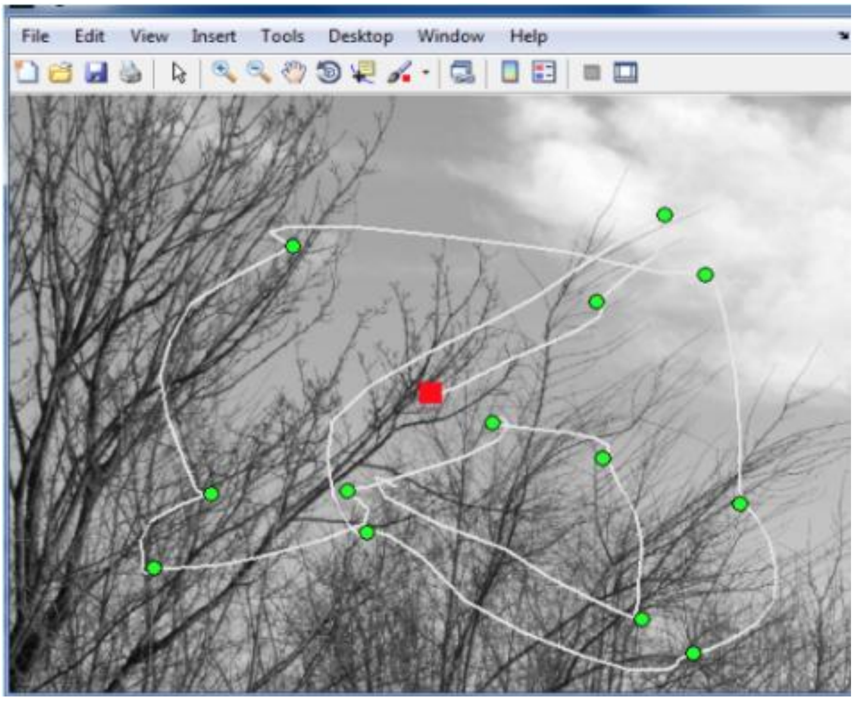


Fig 2.2 Fixation Points plotted in an image

situation of the eye is utilized to decide look course of an individual at a given time and furthermore the arrangement in which there are moved. That is valuable for researchers who attempt to comprehend developments of the eye while an individual is engaged with various exercises.

In the Fig 2.2, the red square shows the first fixation and green dots show the computed fixations. With the study of above-mentioned applications, it is proposed that the eye tracking data can be analysed for various problems. It uses eye tracking metrics, scan paths and fixations to identify glaucomatous changes in high-risk groups.

The advancement of Eye Tracking from an idea to the real world, it is being investigated experimentally nowadays in Human PC Interaction so as to record the eye developments to decide the look course, position of a client on the screen at a given time and the arrangement of their development. The triple goal is to incorporate acquainting the reader with the key perspectives and issues of eye-development innovation, down to earth direction for building up an Eye following application, and different openings and basic difficulties to create (Man and Machine Interfacing) MAMI frameworks utilizing Eye-tracking. Eye GPS beacons joined with physiological information, for example, mind imaging can help distinguish how the data is handled in the cerebrum. Eye following can be utilized to dissect visual advancement and connection it to formative parts of neurological capacities, neurological maladies, and mental harm. Additionally, Reading examples can be cross-referenced with various socioeconomics of individuals and along these lines give knowledge into how they accumulate it[3].

An eye tracking device is playing a major role in the research area due to its feature to capture the eye movements of a human. Eye tracker gives examples of visual consideration of the onlooker which is broken down as far as obsessions and saccades. The proposed work is utilized eye tracker for preparing the arrangement of the programmed discovery of no-section signboard. The entire procedure is isolated into three phases: in the first stage, the fixation point is gathered from an eye-tracker gadget. In the second stage, objects are labelled by k-implies grouping calculation connected on fixation focuses. In the third arrange, the framework is prepared to utilize course preparing calculation to distinguish the objects consequently. No-entry sign sheets are considered as the objective in the proposed work. Recognition of objects depends on obsession focuses as opposed to examining the highlights of the considerable number of pixels of the pictures. Fixation points and fixation duration plays a major role to get features of the area of interest to train the system. This helps to focus on only that region of interest rather than all the pixels of images. Once the system is trained, then the features collected will facilitate to detect no entry signs automatically. It finds the target even if it is more than one.[4]

The paper titled “Eye Communication System for Nonspeaking Patients “gives an overview of eye tracker device used for patients benefits. A stay in an intensive care unit (ICU), although potentially life-saving, may be a traumatic experience to patients. The frustration reported is partly due to communication difficulties caused by the presence of artificial ventilation and intubation. The primary goal of this examination is to build up a medicinal gadget to improve the correspondence capacity of patients who are intubated or getting mechanical ventilation. It will be connected to an eye following framework to control a graphical UI (GUI) for correspondence with the eyes. The primary clear utilization of this ability is in augmentative and elective correspondence field. Truth be told, the primary constant uses of eye-following in human-PC connection tended to individuals with handicaps. People with extreme and significant disabilities can utilize this innovation to talk, send messages, peruse the web and perform other comparative exercises, utilizing just their eyes. The principal concern is to advise patients' indications, torment levels and needs so as to get to their wellbeing condition and improve their daily life at the clinic.[5]

2.3 Hand Region Segmentation:

Humans are specialists at rapidly and precisely distinguishing the most outwardly detectable region in the scene, known as area of interest, and adaptively concentrate on such seen important regions. Conversely, computationally recognizing such striking article locales that coordinate the human annotators' conduct when they have been approached to pick a notable item in a picture, is testing. Having the capacity to naturally, efficiently, and precisely gauge striking article locales, in any case, is profoundly attractive given the quick capacity to describe the spatial help for highlight extraction, seclude the item from conceivably confounding foundation, and specially allot finite computational assets for consequent picture preparing. The saliency of an area principally relies upon its stand out from regard to its adjacent areas, while differentiations to far off locales are less significant(Refer to Fig 2.3). In man-made photos, objects are frequently thought towards the inward areas of the pictures, far from picture limits. Saliency maps ought to be quick, precise, have low memory impressions, and simple to create to permit preparing of extensive picture accumulations, and encourage efficient picture classification and recovery.[6]

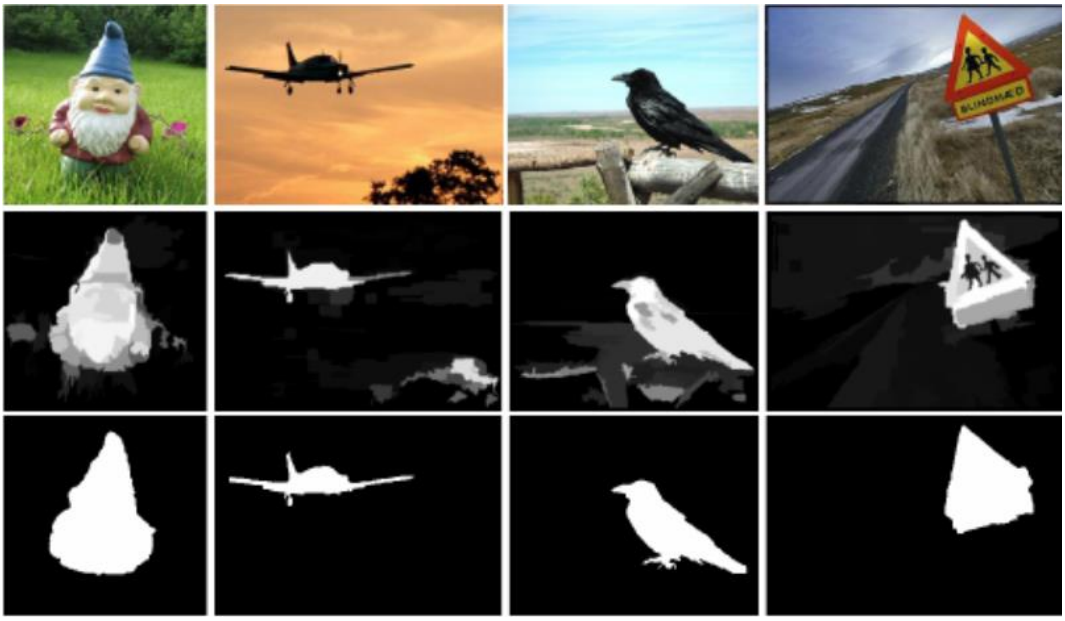


Fig. 2.3. Given an info picture (top), a complexity investigation is utilized to process a high goals saliency map (center), which can be utilized to create an unsupervised segmentation mask (base) for an object of interest. (Src : Research paper [5])

CHAPTER 3
SYSTEM SPECIFICATION

The hardware and software specifications for our project are as follows:

Hardware Specifications:

The two main hardware devices required to acquire data:

Tobii Eye Tracker 4C- Refer to Fig 3.1

Dragonfly Express – Refer to Fig 3.2

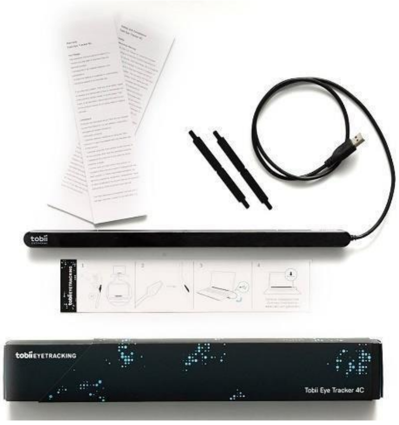


Fig 3.1 Tobii Eye Tracker

(Src: Google Images)

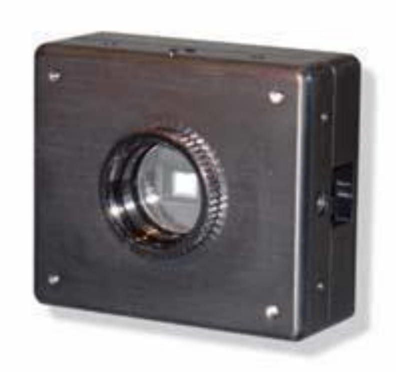


Fig 3.2 Dragonfly Express

(Src: Google Images)

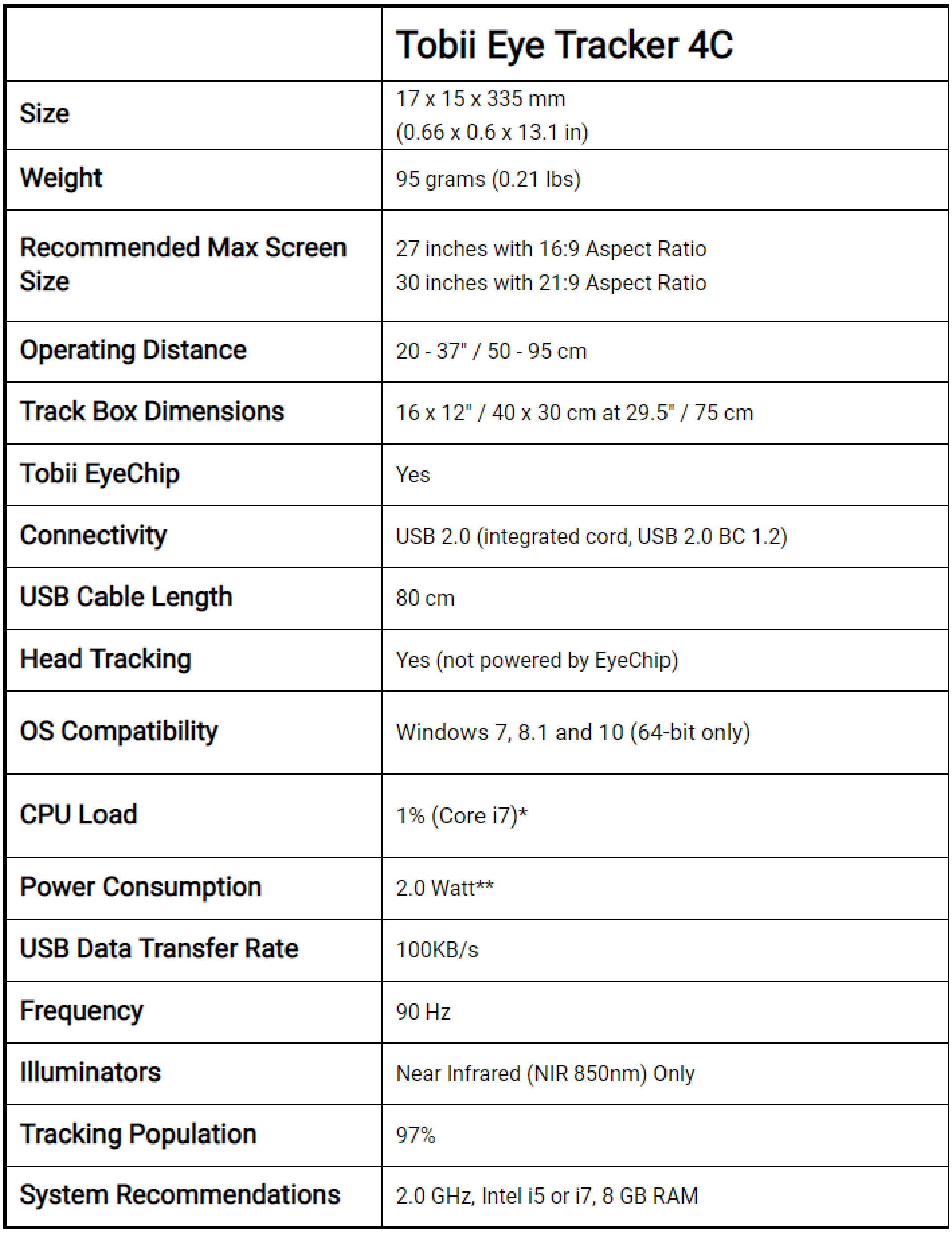
System - Intel i5 or i7

Hard Disk – 1TB

RAM – 4GB

Tobii Eye Tracker 4C:

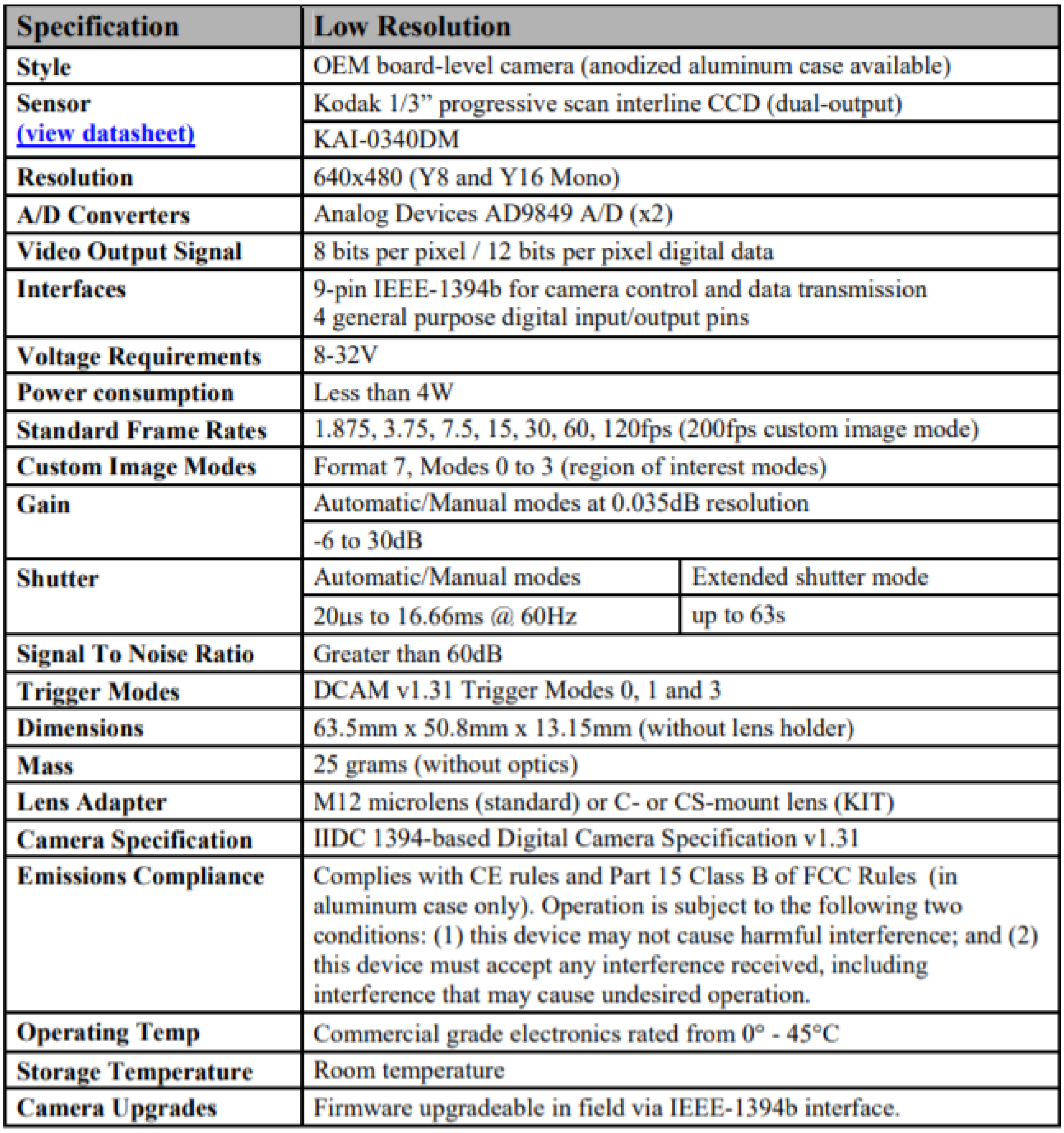
The eye tracker used in our project is theTobii Eye Tracker 4C. It is considered to be a static eye tracker since the eye movement cannot be captured when the person is moving back and forth. It can be captured when a person statically gazes at an image. The average power consumption during usage is 2.0W, but the eye trackers power consumption can spike up to 5-6W depending on the use case. For example, when the illuminators first turn on, the consumption can briefly spike (matter of milliseconds) at around 5-6W. The Tobii Eye Tracker 4C officially works with USB 2.0 BC1.2 which can handle spikes up to 7.5W. (Refer Table 1)



	Tobii Eye Tracker 4C
Size	17 x 15 x 335 mm (0.66 x 0.6 x 13.1 in)
Weight	95 grams (0.21 lbs)
Recommended Max Screen Size	27 inches with 16:9 Aspect Ratio 30 inches with 21:9 Aspect Ratio
Operating Distance	20 - 37" / 50 - 95 cm
Track Box Dimensions	16 x 12" / 40 x 30 cm at 29.5" / 75 cm
Tobii EyeChip	Yes
Connectivity	USB 2.0 (integrated cord, USB 2.0 BC 1.2)
USB Cable Length	80 cm
Head Tracking	Yes (not powered by EyeChip)
OS Compatibility	Windows 7, 8.1 and 10 (64-bit only)
CPU Load	1% (Core i7)*
Power Consumption	2.0 Watt**
USB Data Transfer Rate	100KB/s
Frequency	90 Hz
Illuminators	Near Infrared (NIR 850nm) Only
Tracking Population	97%
System Recommendations	2.0 GHz, Intel i5 or i7, 8 GB RAM

Table 1 Tobii Eye Tracker Specifications(https://help.tobii.com/hc/en-us/articles/213414285-Specifications-for-the-Tobii-Eye-Tracker-4C)

Dragonfly Express: The iCub dataset is being captured using Dragonfly express Camera. (Refer to Table 2)



Specification	Low Resolution	
Style	OEM board-level camera (anodized aluminum case available)	
Sensor (view datasheet)	Kodak 1/3" progressive scan interline CCD (dual-output)	
	KAI-0340DM	
Resolution	640x480 (Y8 and Y16 Mono)	
A/D Converters	Analog Devices AD9849 A/D (x2)	
Video Output Signal	8 bits per pixel / 12 bits per pixel digital data	
Interfaces	9-pin IEEE-1394b for camera control and data transmission 4 general purpose digital input/output pins	
Voltage Requirements	8-32V	
Power consumption	Less than 4W	
Standard Frame Rates	1.875, 3.75, 7.5, 15, 30, 60, 120fps (200fps custom image mode)	
Custom Image Modes	Format 7, Modes 0 to 3 (region of interest modes)	
Gain	Automatic/Manual modes at 0.035dB resolution	
	-6 to 30dB	
Shutter	Automatic/Manual modes	Extended shutter mode
	20µs to 16.66ms @ 60Hz	up to 63s
Signal To Noise Ratio	Greater than 60dB	
Trigger Modes	DCAM v1.31 Trigger Modes 0, 1 and 3	
Dimensions	63.5mm x 50.8mm x 13.15mm (without lens holder)	
Mass	25 grams (without optics)	
Lens Adapter	M12 microlens (standard) or C- or CS-mount lens (KIT)	
Camera Specification	IIDC 1394-based Digital Camera Specification v1.31	
Emissions Compliance	Complies with CE rules and Part 15 Class B of FCC Rules (in aluminum case only). Operation is subject to the following two conditions: (1) this device may not cause harmful interference; and (2) this device must accept any interference received, including interference that may cause undesired operation.	
Operating Temp	Commercial grade electronics rated from 0° - 45°C	
Storage Temperature	Room temperature	
Camera Upgrades	Firmware upgradeable in field via IEEE-1394b interface.	

Table 2 Dragonfly Express specifications(http://www.cs.unc.edu/Research/stc/FAQs/Cameras_Lenses/PtGrey/DragonflyEXPRESSGettingStartedManual.pdf)

Software Specifications:

Operating System : Windows 10, Ubuntu

Tools : R studio, MATLAB 2018A , Python

Python Libraries : six,numpy,scipy,Pillow,matplotlib,scikit-image,opencv-python,imageio,shapely

CHAPTER 4

PROPOSED METHODOLOGY

The proposed methodology of this project is as follows: (Refer to Fig 4.1)

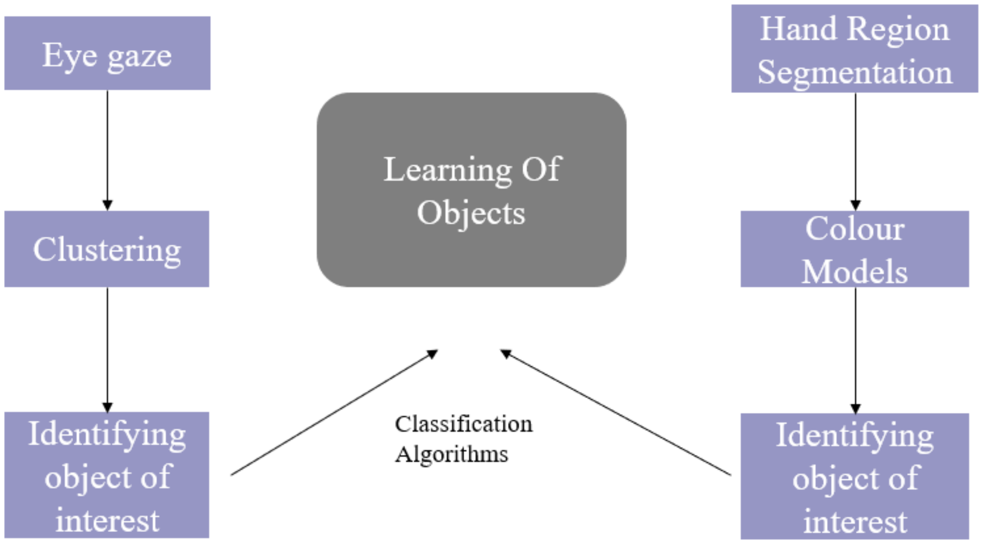

Eye gaze
Clustering
Identifying object of interest
Learning Of Objects
Classification Algorithms
Hand Region Segmentation
Colour Models
Identifying object of interest

Fig 4.1 Proposed Methodology

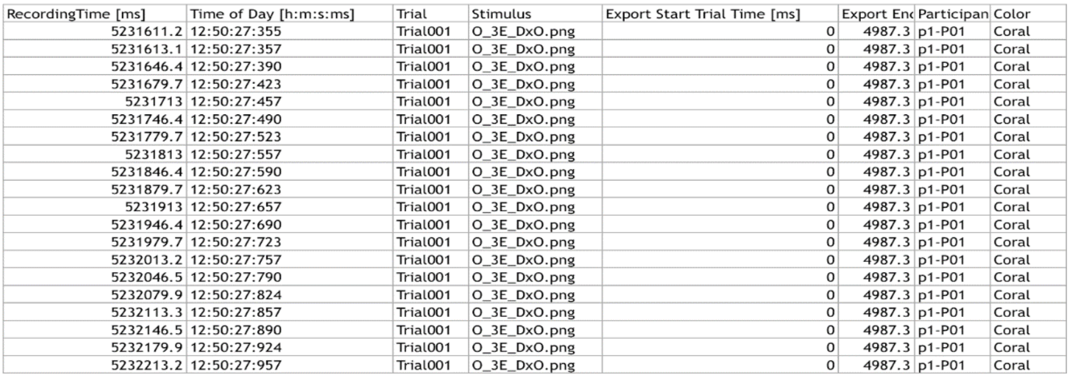
CHAPTER 5

EYE TRACKER

5.1 Eye Tracking Phase:

Dataset Description:

The experimental setup for acquiring input for eye tracking process includes the eye tracker device Tobii Eye Tracker 4C, fitted into the person’s eye and an image on the computer screen. The person is supposed to look at the image for a minute (60 seconds).When he is gazing at the image, a dataset is being generated by the eye tracker device. The eye gaze dataset captured by the device has a lot of fields about his eye movements at particular X and Y coordinates of the image. To be more precise, the dataset gives us an idea about the person’s interest in the image in the form of eye movements. His eye movements on the image are being detected by the device and produced on the dataset in the form of excel sheet.



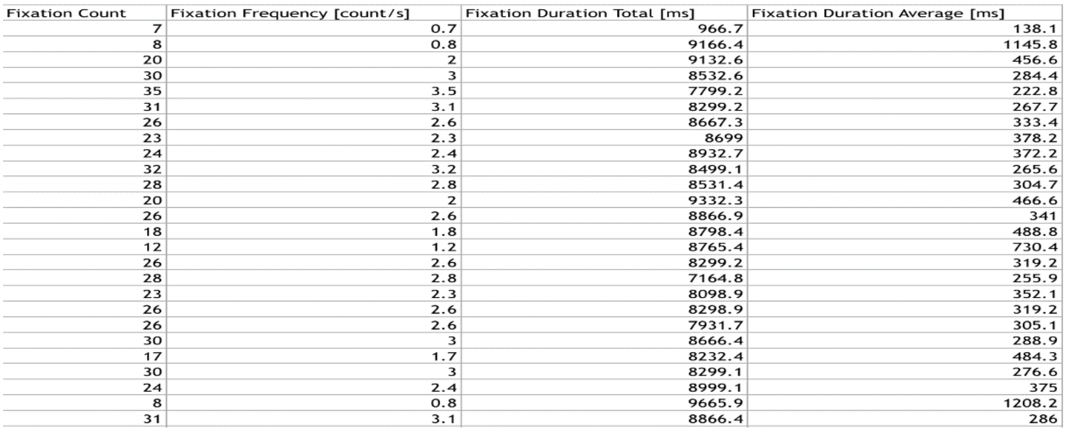
RecordingTime [ms]	Time of Day [h:m:s:ms]	Trial	Stimulus	Export Start Trial Time [ms]	Export Enc	Participan	Color
5231611.2	12:50:27:355	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231613.1	12:50:27:357	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231646.4	12:50:27:390	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231679.7	12:50:27:423	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231713	12:50:27:457	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231746.4	12:50:27:490	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231779.7	12:50:27:523	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231813	12:50:27:557	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231846.4	12:50:27:590	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231879.7	12:50:27:623	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231913	12:50:27:657	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231946.4	12:50:27:690	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5231979.7	12:50:27:723	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232013.2	12:50:27:757	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232046.5	12:50:27:790	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232079.9	12:50:27:824	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232113.3	12:50:27:857	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232146.5	12:50:27:890	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232179.9	12:50:27:924	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral
5232213.2	12:50:27:957	Trial001	O_3E_DxO.png	0	4987.3	p1-P01	Coral

Fig 5.1 Snapshot of Eye Gaze Datasheet

Recording Time - It is the time duration for which his eye movements has been recorded for a particular image. (Refer to Fig 5.1)

Trial - We may have a set of trials for each person corresponding to an image. Among them, this is the first proceeding of the person on the image.

Participant- It corresponds to the person into whose eyes the eye tracker device is being fitted into for experimentation.



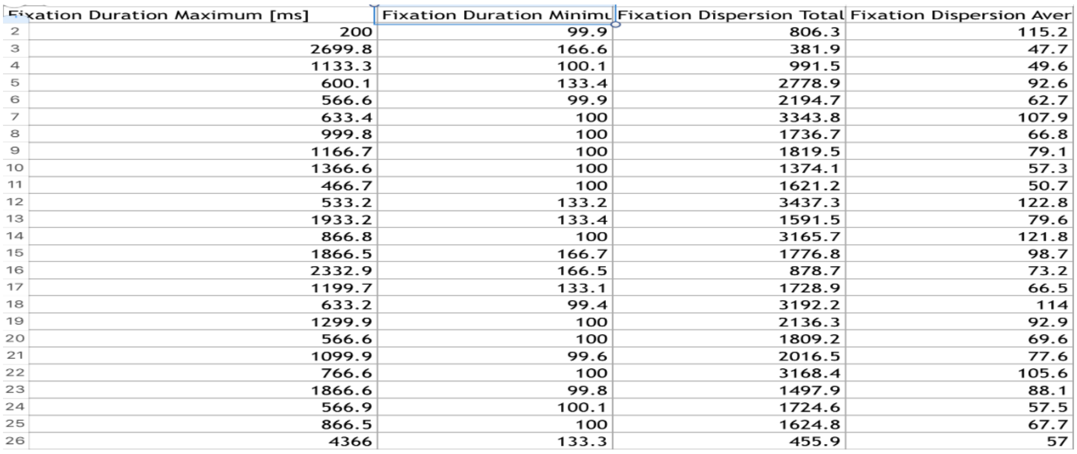
Fixation Count	Fixation Frequency [count/s]	Fixation Duration Total [ms]	Fixation Duration Average [ms]
7	0.7	966.7	138.1
8	0.8	9166.4	1145.8
20	2	9132.6	456.6
30	3	8532.6	284.4
35	3.5	7799.2	222.8
31	3.1	8299.2	267.7
26	2.6	8667.3	333.4
23	2.3	8699	378.2
24	2.4	8932.7	372.2
32	3.2	8499.1	265.6
28	2.8	8531.4	304.7
20	2	9332.3	466.6
26	2.6	8866.9	341
18	1.8	8798.4	488.8
12	1.2	8765.4	730.4
26	2.6	8299.2	319.2
28	2.8	7164.8	255.9
23	2.3	8098.9	352.1
26	2.6	8298.9	319.2
26	2.6	7931.7	305.1
30	3	8666.4	288.9
17	1.7	8232.4	484.3
30	3	8299.1	276.6
24	2.4	8999.1	375
8	0.8	9665.9	1208.2
31	3.1	8866.4	286

Fig 5.2 Snapshot of Eye Gaze Datasheet

Fixation Frequency: It is defined as the number of fixation counts per second.it varies from image to image based on the region of interest on the image. (Refer to Fig 5.2)

Fixation Duration Time: It is measured in milliseconds as the amount of time spent in gazing at a certain area. It is believed to change from person to person. Since one person might gaze at a certain fixation point for a longer time than the other due to his/her interest.

Fixation Average Time: It is measured in milliseconds as the average time spent in gazing at a particular point of an image.



	Fixation Duration Maximum [ms]	Fixation Duration Minimu	Fixation Dispersion Total	Fixation Dispersion Aver
2	200	99.9	806.3	115.2
3	2699.8	166.6	381.9	47.7
4	1133.3	100.1	991.5	49.6
5	600.1	133.4	2778.9	92.6
6	566.6	99.9	2194.7	62.7
7	633.4	100	3343.8	107.9
8	999.8	100	1736.7	66.8
9	1166.7	100	1819.5	79.1
10	1366.6	100	1374.1	57.3
11	466.7	100	1621.2	50.7
12	533.2	133.2	3437.3	122.8
13	1933.2	133.4	1591.5	79.6
14	866.8	100	3165.7	121.8
15	1866.5	166.7	1776.8	98.7
16	2332.9	166.5	878.7	73.2
17	1199.7	133.1	1728.9	66.5
18	633.2	99.4	3192.2	114
19	1299.9	100	2136.3	92.9
20	566.6	100	1809.2	69.6
21	1099.9	99.6	2016.5	77.6
22	766.6	100	3168.4	105.6
23	1866.6	99.8	1497.9	88.1
24	566.9	100.1	1724.6	57.5
25	866.5	100	1624.8	67.7
26	4366	133.3	455.9	57

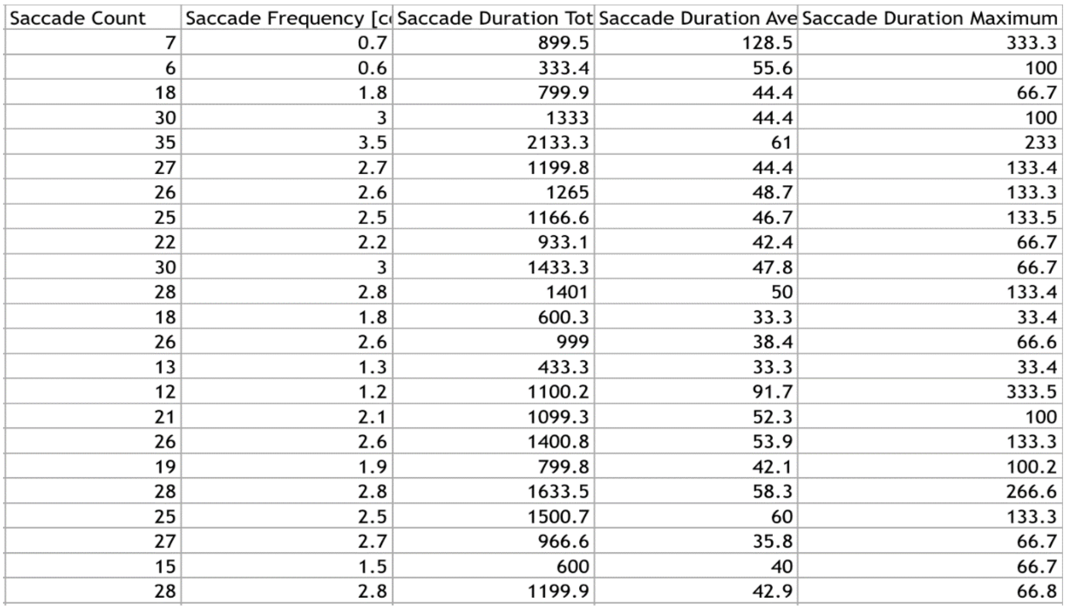
Fig 5.3 Snapshot of Eye Gaze Datasheet

Fixation Duration Maximum: It is measured in milliseconds as the amount of time a person spends maximum in gazing at a particular area of an image. It shows the person's area of interest in the image. (Refer to Fig 5.3)

Fixation Duration Minimum: It is measured in milliseconds. Lower the fixation duration minimum predicts his ignorance towards that particular region of interest.

Fixation Dispersion Total: This metric emphasizes on the dispersion (spread distance) of fixation points assuming that fixation points maybe close to one another in an image.

Fixation Dispersion Average: This defined as an average value of dispersion of fixation points on an image for a person.



Saccade Count	Saccade Frequency [c	Saccade Duration Tot	Saccade Duration Ave	Saccade Duration Maximum
7	0.7	899.5	128.5	333.3
6	0.6	333.4	55.6	100
18	1.8	799.9	44.4	66.7
30	3	1333	44.4	100
35	3.5	2133.3	61	233
27	2.7	1199.8	44.4	133.4
26	2.6	1265	48.7	133.3
25	2.5	1166.6	46.7	133.5
22	2.2	933.1	42.4	66.7
30	3	1433.3	47.8	66.7
28	2.8	1401	50	133.4
18	1.8	600.3	33.3	33.4
26	2.6	999	38.4	66.6
13	1.3	433.3	33.3	33.4
12	1.2	1100.2	91.7	333.5
21	2.1	1099.3	52.3	100
26	2.6	1400.8	53.9	133.3
19	1.9	799.8	42.1	100.2
28	2.8	1633.5	58.3	266.6
25	2.5	1500.7	60	133.3
27	2.7	966.6	35.8	66.7
15	1.5	600	40	66.7
28	2.8	1199.9	42.9	66.8

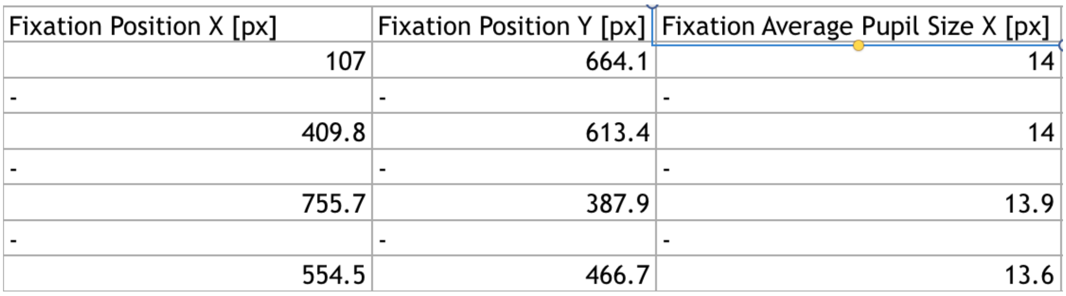
Fig 5.4 Snapshot of Eye Gaze Datasheet

Saccade Count: It is defined as the number of rapid eye movements between two fixation points. (Refer to Fig 5.4)

Saccade Frequency: It is the number of simultaneous movements of both the eyes while moving from one fixation point to another per second

Saccade Duration Total: It is the total amount of time spent in the movement of eyes from one fixation point to another.

Saccade Duration Average: The average time taken in the eye movement while moving between two fixation points.



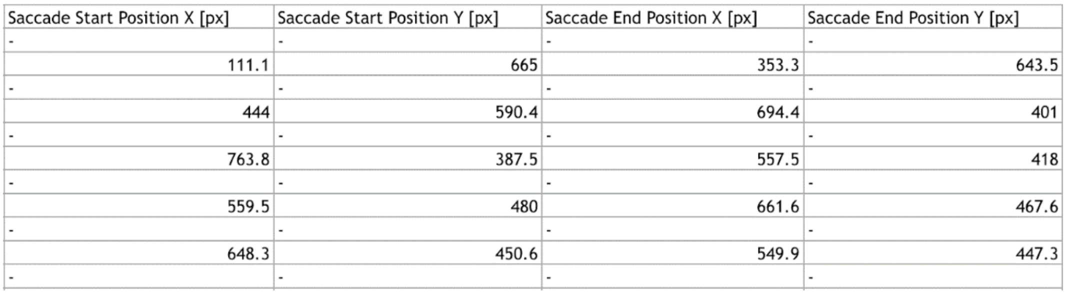
Fixation Position X [px]	Fixation Position Y [px]	Fixation Average Pupil Size X [px]
107	664.1	14
-	-	-
409.8	613.4	14
-	-	-
755.7	387.9	13.9
-	-	-
554.5	466.7	13.6

Fig 5.5 Snapshot of Eye Gaze Datasheet

Fixation Position X[px]: It is measured in pixels as the position of eye on the image at a particular instant pertaining to X-coordinate of the image. (Refer to Fig 5.5)

Fixation Point Y[px]: It is measured in pixels as the position of the eye on the Y-coordinate of the image.

Fixation Average Pupil Size: A person's pupil enlarges when he is looking at some area in an image. This size corresponds to the pupil's size when he is gazing at a fixation point.



Saccade Start Position X [px]	Saccade Start Position Y [px]	Saccade End Position X [px]	Saccade End Position Y [px]
-	-	-	-
111.1	665	353.3	643.5
-	-	-	-
444	590.4	694.4	401
-	-	-	-
763.8	387.5	557.5	418
-	-	-	-
559.5	480	661.6	467.6
-	-	-	-
648.3	450.6	549.9	447.3
-	-	-	-

Fig 5.6 Snapshot of Eye Gaze Datasheet

Saccade Start Position[x] [y]: It corresponds to the eye movement on the image as X and Y coordinates. The eye movement captured while moving from one fixation point is Saccade Start Position. (Refer to Fig 5.6)

Saccade End Position[x] [y]: It corresponds to the eye movement to the image as X and Y coordinates. The eye movements captured when a person's eyes stops moving at a fixation point is Saccade End Position.

Parameters Selection for Eye Tracking Process:

With the above mentioned attributes in the dataset, we have to choose some attributes as eye tracking parameters which will help us in region identification and processing. Fixation points[X] and Fixation points[Y] are the two parameters to be chosen for further processing of eye tracking. These two points are being chosen since fixation points are maintaining visual gaze on a certain location. A person's visual gaze for a longer time tells us that he is more interested in that area of an image amongst the background. Our project works deals with extracting the region of interest. Among the attributes of eye gaze datasheet, only fixation points denote the area of interest in an image.

Pre-processing Of Eye Tracking Parameters:

After choosing the parameters for the eye tracking process, the data has to be pre-processed to remove any incomplete, inconsistent errors in the dataset. The missing values can be found in the dataset. In order to get rid of missing values, the dataset is being processed and remove missing rows. After the missing values are being removed, we have to remove outliers in the dataset. Outliers in input data can mislead the training process resulting in less accurate models and ultimately poorer results. The dataset is being processed in matlab and remove the outliers. With the Input Acquisition being done, we have to move to the eye tracking phase.

5.2 Training Process:

With the fixation points being used as an input dataset for the training, we have to predict the region of interest on an image. Different people will have different fixation points. Thus, it requires us to find the region of interest from fixation points. This can be done by the following steps:

1) K means clustering can be applied on the fixation points obtained after pre-processing.

2) Each cluster obtained from K means will have a centroid point.

3) The centroid point of the fixation point can be used for cropping and labelling the clusters as boundary boxes.

k-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. In our project, k cluster corresponds the various regions of interest processed by the fixation points(Refer to Fig 5.7). After the clustering process, we need to crop the area of region from the background. With the centroid of each cluster, we have to capture those clusters. These clusters are labelled as boundary boxes and processed in the next validation phase.

5.3 Validation Phase:

The boundary boxes obtained from clustering process are tested across a set of positive and negative images. By positive images, it means the object or region of interest. By negative image, it includes images without the object to be identified or area of interest. This validation process is brought about machine learning algorithms such as k-means, SVM Classifier and Naïve Bayes Classifier Algorithms.

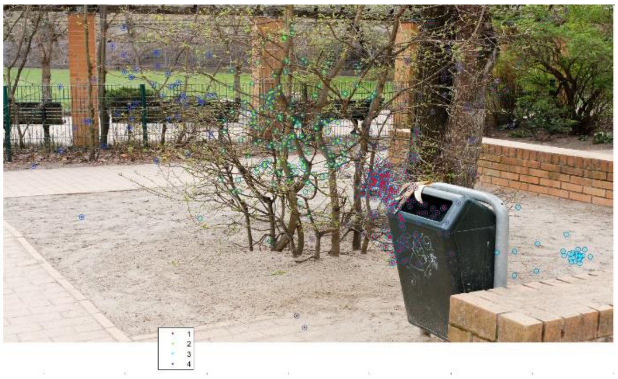


Fig 5.7 Boundary labels as a k-means cluster

CHAPTER 6
HAND SEGMENTATION PROCESS

6.1 Introduction

Hand segmentation becomes a challenging task due to uncontrolled environmental conditions, lighting, rapid motion of the hand and skin colour detection. This paper's objective is to propose saliency-based colour model algorithm for hand segmentation under constrained and non-constrained environments. Researchers are undergoing various works on hand segmentation to attain natural interaction with a machine. The objective of this paper is to excel in the region of skin colour detection for human-like interaction between the end user and the computer. In order to identify peoples with varying skin tones, a suitable colour range for the skin segmentation is inevitable.

6.2 Input Acquisition

The iCub image dataset was used. All the images are of the same size 128*128(pixels) in JPG format. Every image in the dataset is an object in the hand. The performance of the hand segmentation models were tested using factors like 'True positive rate (TPR)', 'false positive rate (FPR)', 'True Negative Rate (TNR)' and 'False Negative rate (FNR)'.The above mentioned factors are measured in percentage for comparison between ground truth image and image obtained from colour models.

6.3 Colour Models for Hand Segmentation

A) RGB Model:

The Red, Green and Blue (RGB) colour model has its key additive colours as red, green and blue which in different combinations will produce a wide range of distinct colours. It is the primary additive colour model from which the following colour models are defined.

B) HSV Model:

HSV colour model is constituted of Hue; indicating the way in which RGB colours are mixed together to form new colours; Saturation measuring the limit of the colour (how light green is separated from dark green) and Value indicating the level of light and dark colours. This model is mostly used in web applications as a colour selection tool but may prove ineffective in low brightness backgrounds for they cannot be differentiated from low contrast background.

C) YCbCr Model (Yellow,Chromaticity of Blue, Chromaticity Of Red):

In the YCbCr color space,Y is the luma component of the colour. Luma can be defined as brightness of a colour, for which the human eye is most sensitive to Cb and Cr are defined as chromaticities with respect to green component. Cb is blue component relative to green and Cr is the red component relative to red.It is used for component digital video.

D) HSI Model:

The HSI model represents colour similar to humans perceiving it in the eyesThe Hue component describes the colour which is measured in angle(0 degree means red),the Saturation colour indicated the colour’s mixture with white colour(measured in the range [0,1]).The Intensity ranges from 0(black) to 1(white). When the input for the dataset (hand with an object) is acquired through the webcam as video, the values of Hue for skin colour for the same person changes in each frame. The acquired frames are highly dependent on background illumination. In order to detect the skin region accurately, the range of Hue is varied by the device capturing the image in each frame with the help of information obtained from the previous frame. HSI is independent of luminance and reflectance.

E) CMYK Model:

CMYK is a subtractive colour model used as one of the hand segmentation algorithms. The primary colours are cyan (C), magenta (M), and yellow (Y). Sometimes black (K) is also considered a primary colours, although black can be obtained by combining pure cyan, magenta, and yellow in equal and considerably large amounts.

F) Saliency Model:

Saliency describes how an object or region outstands from the rest of the background in an image. It was more frequently used to predict eye movements during image viewing. When humans view an image without any task in mind, their eyes are drawn towards objects that stand out amid the background, these areas are described as salient.

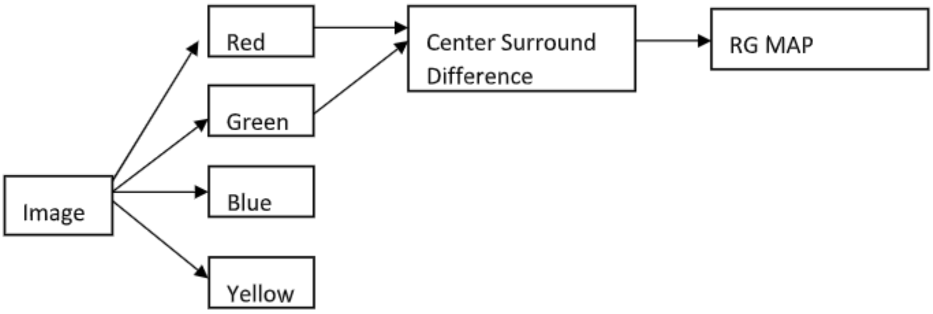


Fig 6.1 Red and Green Salient Model For Hand Segmentation

RGMAP-red and green salient model

GRMAP-green and red salient model

RYMAP-red and yellow salient model

RYADD-red and yellow additive salient model

RY-red and yellow additive model

```
rarry =filtr_img(:,:,1);
garry=filtr_img(:,:,2);
barry=filtr_img(:,:,3);
rarry=double(rarry);
garry=double(garry);
barry=double(barry);
R=rarry-(garry+barry)/2;
G=garry-(rarry+barry)/2;
B=barry-(garry+rarry)/2;
Y=rarry+garry-2*(abs(rarry-garry)+barry);
R=uint8(R);R = uint8(255*mat2gray(R));
```

```
G=uint8(G);G = uint8(255*mat2gray(G));
B=uint8(B);B = uint8(255*mat2gray(B));
Y=uint8(Y);Y = uint8(255*mat2gray(Y));
map{1}=imresize(R, [128 128]);
map{2}=imresize(G, [128 128]);
map{3}=imresize(B, [128 128]);
map{4}=imresize(Y, [128 128]);
[cr sr]=Pyr_Rescale(R);
[cg sg]=Pyr_Rescale(G);
[cb sb]=Pyr_Rescale(B);
[cy sy]=Pyr_Rescale(Y);
f_rgmap=abs((cr-cg)-(sg-sr));  // RGMAP
f_grmap=abs((cg-cr)-(sr-sg));  //GRMAP
f_rymap=abs((cr-cy)-(sr-sy)); //RYMAP
f_rymapadd=abs((cr-cy)+(sr-sy));//RYADD
f_RY=R+Y; //RY
```

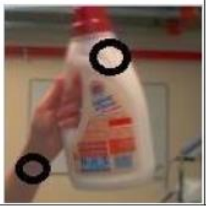
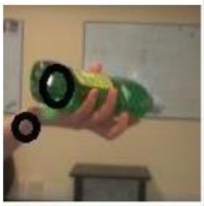
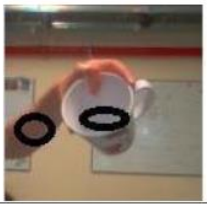
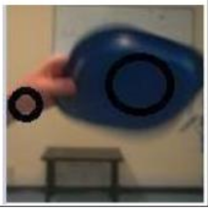
Image Dataset	1-30	31-60	61-90	91-120
Representative Image				

Table 3 Comparison of Image Dataset alongside Object orientations

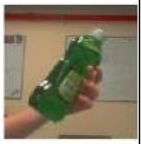
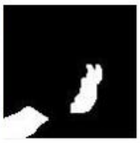
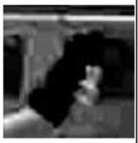
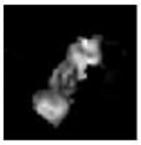
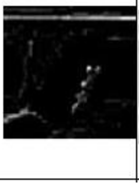
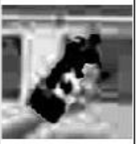
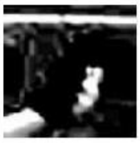
Original Image	Ground Truth Image	RGMAP	GRMAP	RYMAP	RY	RYADD
						

Table 4. Comparison between different salient models

CHAPTER 7

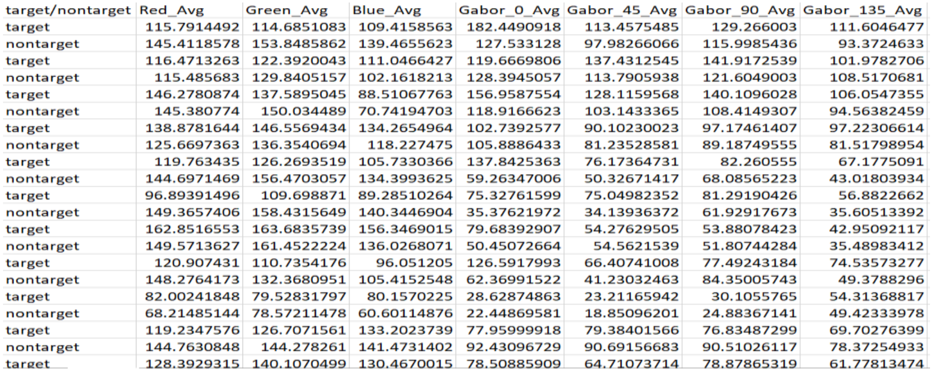
REGION IDENTIFICATION AND OBJECT DETECTION

7.1 Input Acquisition

The input dataset used for region identification includes boundary labels derived from the k-means cluster of the eye tracking phase. The boundary labels will be trained to detect the target image from the non-target background. This classification of target image and non-target background is achieved by machine learning classification algorithms. Some algorithms include Naïve Bayes, Support Vector Machine(SVM) and k-means clustering algorithms. The dataset is trained with fifty positive images of target and fifty negative images.

Processing the training images for generating an excel sheet

1) The training set will have attributes inserted for each image. Each image will have 2 rows in the excel sheet.
2) One row will deal with the target class and the other row will have the attribute values for the non-target class. Therefore we will have a training sheet of 2*no of images. The feature extraction would have Gabor filter and RGB colour model.
3) The attributes would be average of the Gabor filter on 0, 45, 90, 135 degrees and an average of the red, green and blue components of the image**.** (Refer to Fig 7.1)

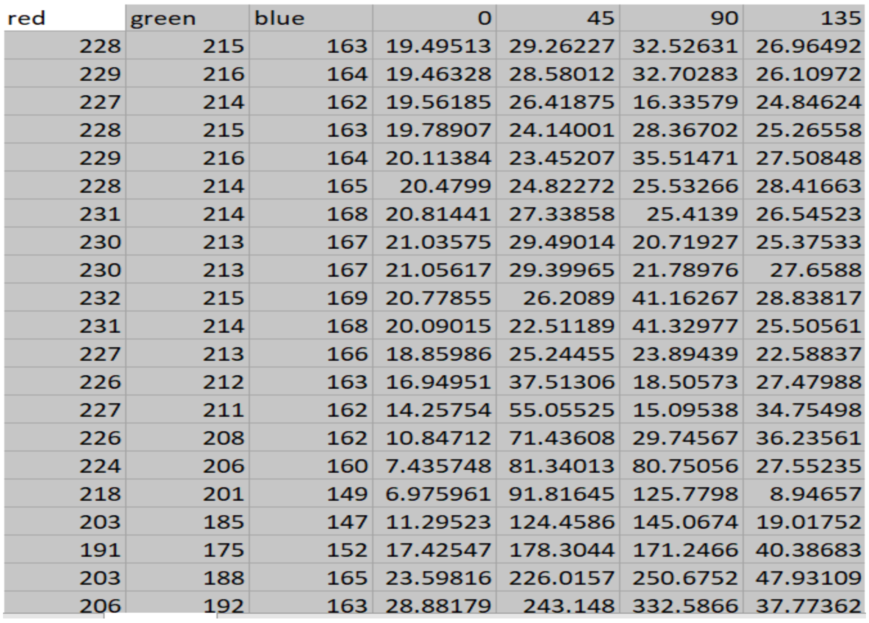


target/nontarget	Red_Avg	Green_Avg	Blue_Avg	Gabor_0_Avg	Gabor_45_Avg	Gabor_90_Avg	Gabor_135_Avg
target	115.7914492	114.6851083	109.4158563	182.4490918	113.4575485	129.266003	111.6046477
nontarget	145.4118578	153.8485862	139.4655623	127.533128	97.98266066	115.9985436	93.3724633
target	116.4713263	122.3920043	111.0466427	119.6669806	137.4312545	141.9172539	101.9782706
nontarget	115.485683	129.8405157	102.1618213	128.3945057	113.7905938	121.6049003	108.5170681
target	146.2780874	137.5895045	88.51067763	156.9587554	128.1159568	140.1096028	106.0547355
nontarget	145.380774	150.034489	70.74194703	118.9166623	103.1433365	108.4149307	94.56382459
target	138.8781644	146.5569434	134.2654964	102.7392577	90.10230023	97.17461407	97.22306614
nontarget	125.6697363	136.3540694	118.227475	105.8886433	81.23528581	89.18749555	81.51798954
target	119.763435	126.2693519	105.7330366	137.8425363	76.17364731	82.260555	67.1775091
nontarget	144.6971469	156.4703057	134.3993625	59.26347006	50.32671417	68.08565223	43.01803934
target	96.89391496	109.698871	89.28510264	75.32761599	75.04982352	81.29190426	56.8822662
nontarget	149.3657406	158.4315649	140.3446904	35.37621972	34.13936372	61.92917673	35.60513392
target	162.8516553	163.6835739	156.3469015	79.68392907	54.27629505	53.88078423	42.95092117
nontarget	149.5713627	161.4522224	136.0268071	50.45072664	54.5621539	51.80744284	35.48983412
target	120.907431	110.7354176	96.051205	126.5917993	66.40741008	77.49243184	74.53573277
nontarget	148.2764173	132.3680951	105.4152548	62.36991522	41.23032463	84.35005743	49.3788296
target	82.00241848	79.52831797	80.1570225	28.62874863	23.21165942	30.1055765	54.31368817
nontarget	68.21485144	78.57211478	60.60114876	22.44869581	18.85096201	24.88367141	49.42333978
target	119.2347576	126.7071561	133.2023739	77.95999918	79.38401566	76.83487299	69.70276399
nontarget	144.7630848	144.278261	141.4731402	92.43096729	90.69156683	90.51026117	78.37254933
target	128.3929315	140.1070499	130.4670015	78.50885909	64.71073714	78.87865319	61.77813474

Fig 7.1 Feature Extraction for Training Dataset

Processing an excel sheet for a single test image:

The testing excel (Refer to Fig 7.2) is created in a unique manner. One single image is taken and all the attribute values are got for each pixel in that image. If the image has a height of 'h' and width of 'w' then the excel will have h*w rows inserted. The novelty dealt here is each pixel will be classified as target class labels and a confusion matrix will be created with true positive, true negative, false positive and false negative values.



red	green	blue	0	45	90	135
228	215	163	19.49513	29.26227	32.52631	26.96492
229	216	164	19.46328	28.58012	32.70283	26.10972
227	214	162	19.56185	26.41875	16.33579	24.84624
228	215	163	19.78907	24.14001	28.36702	25.26558
229	216	164	20.11384	23.45207	35.51471	27.50848
228	214	165	20.4799	24.82272	25.53266	28.41663
231	214	168	20.81441	27.33858	25.4139	26.54523
230	213	167	21.03575	29.49014	20.71927	25.37533
230	213	167	21.05617	29.39965	21.78976	27.6588
232	215	169	20.77855	26.2089	41.16267	28.83817
231	214	168	20.09015	22.51189	41.32977	25.50561
227	213	166	18.85986	25.24455	23.89439	22.58837
226	212	163	16.94951	37.51306	18.50573	27.47988
227	211	162	14.25754	55.05525	15.09538	34.75498
226	208	162	10.84712	71.43608	29.74567	36.23561
224	206	160	7.435748	81.34013	80.75056	27.55235
218	201	149	6.975961	91.81645	125.7798	8.94657
203	185	147	11.29523	124.4586	145.0674	19.01752
191	175	152	17.42547	178.3044	171.2466	40.38683
203	188	165	23.59816	226.0157	250.6752	47.93109
206	192	163	28.88179	243.148	332.5866	37.77362

Fig 7.2 Single excel sheet for a test image

7.2 Algorithms Used For Feature Learning:

A) Naïve Bayes:

The dataset is divided into two parts, namely, **feature matrix** and the **response vector**.

- Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of **dependent features**. In above dataset, features are ‘Red’, ‘Green’, ‘Blue’ ,‘Gabor 0’,’Gabor 45’,’Gabor 90’ and ‘Gabor 135’.
- Response vector contains the value of **class variable** (prediction or output) for each row of feature matrix. In above dataset, the class variable name is ‘target/non-target’.

With relation to our dataset, this concept can be understood as:

- It is assumed that no pair of features are dependent. For example, the Red value being ‘228’ has nothing to do with the green value or the blue value. Hence, the features chosen are considered to be **independent**.
- Secondly, each feature is given the same weight (or importance). For example, knowing only red value and Gabor 45 alone can’t predict the outcome accurately. None of the features is irrelevant and assumed to be contributing **equally** to the outcome.

Inferring about Naïve Bayes Algorithm, They require a small amount of training data to estimate the necessary parameters.
Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods. The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution.

B) Support Vector Machine:

In machine learning, support-vector machines (**SVMs**, also support-vector networks) are supervised learning **models** with associated learning algorithms that analyse data used for classification and regression analysis.

A SVM show is a portrayal of the models as focuses in space, mapped with the goal that the instances of the different classifications are separated by a reasonable hole that is as wide as would be possible.

Notwithstanding performing straight characterization, SVMs can proficiently play out a non-direct grouping, certainly mapping their contributions to high-dimensional component spaces.

C) Decision Tree:

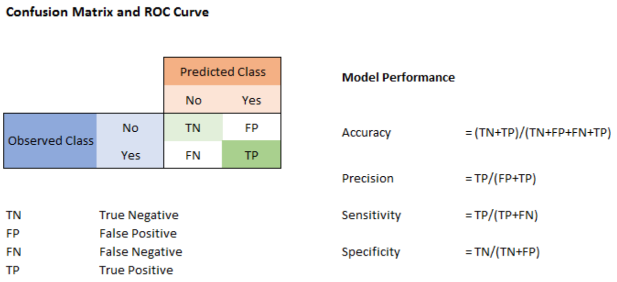
A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Comparison for the Best Algorithm:

The different classification algorithms are being compared using the confusion matrix parameters such as Accuracy, Precision, True Positive Rate and False Negative rates. The performance of the classification algorithms were tested using factors like ‘True positive rate (TPR)’, ‘false positive rate (FPR)’, ‘True Negative Rate (TNR)’ and ‘False Negative rate (FNR)’.The above mentioned factors are measured in percentage for comparison between ground truth image and target obtained from classifiers. (Refer to Fig 7.3)

- Accuracy is a degree of closeness of measurements of the ground truth image and processed image from the different types of classifiers.

- Precision demonstrates how much percent of the results are relevant. Recall and precision are extremely important and poles apart since we have recall determining how much percent of the results are totally classified in the algorithm correctly.



Confusion Matrix and ROC Curve

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

TN	True Negative
FP	False Positive
FN	False Negative
TP	True Positive

Model Performance

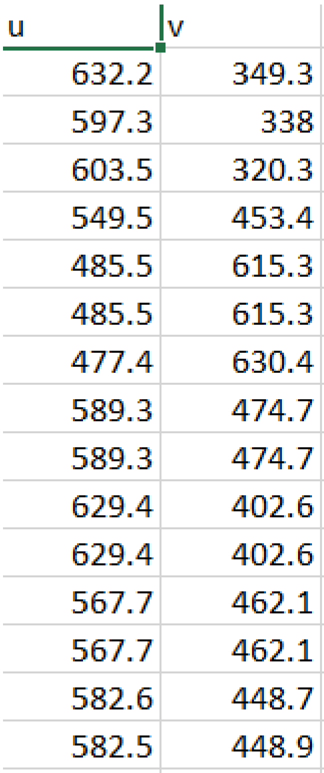
Accuracy	= (TN+TP)/(TN+FP+FN+TP)
Precision	= TP/(FP+TP)
Sensitivity	= TP/(TP+FN)
Specificity	= TN/(TN+FP)

Fig 7.3 Confusion Matrix for Best Algorithm Prediction(Src: Google Images)

CHAPTER 8

RESULTS AND DISCUSSION

8.1 Eye Tracking Phase:



u	v
632.2	349.3
597.3	338
603.5	320.3
549.5	453.4
485.5	615.3
485.5	615.3
477.4	630.4
589.3	474.7
589.3	474.7
629.4	402.6
629.4	402.6
567.7	462.1
567.7	462.1
582.6	448.7
582.5	448.9

Fig 8.1 Fixation Points

Fixation points[X] and Fixation points[Y] are the two parameters to be chosen for further processing of eye tracking. These two points are being chosen since fixation points are maintaining visual gaze on a certain location. A person's visual gaze for a longer time tells us that he is more interested in that area of an image amongst the background. Our project works deals with extracting the region of interest. Among the attributes of eye gaze datasheet, only fixation points denote the area of interest in an image.

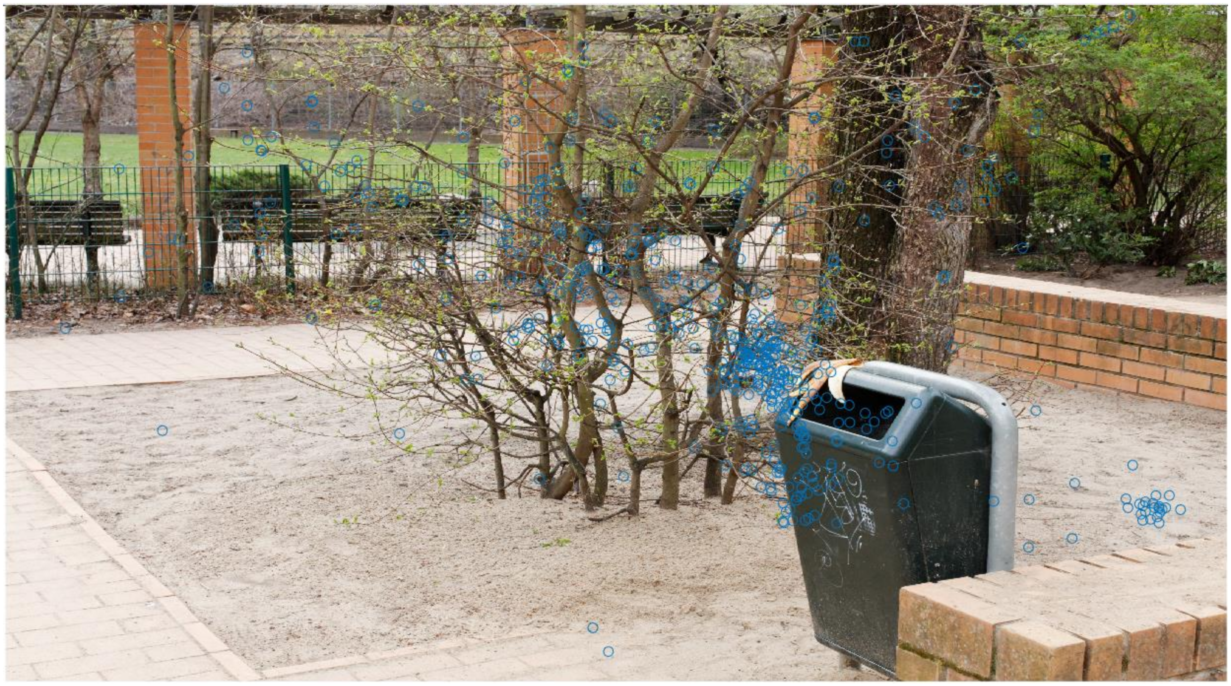


Fig 8.2 Plotting Of Fixation Points

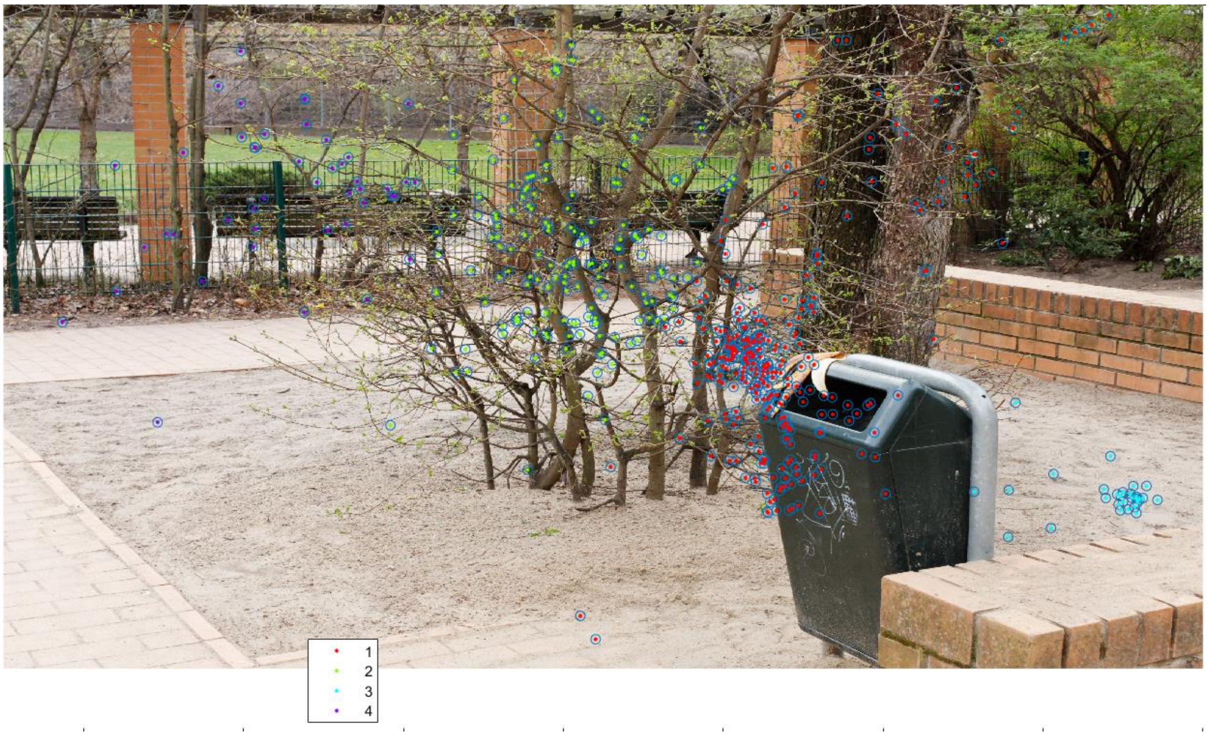


Fig 8.3 K means clustering

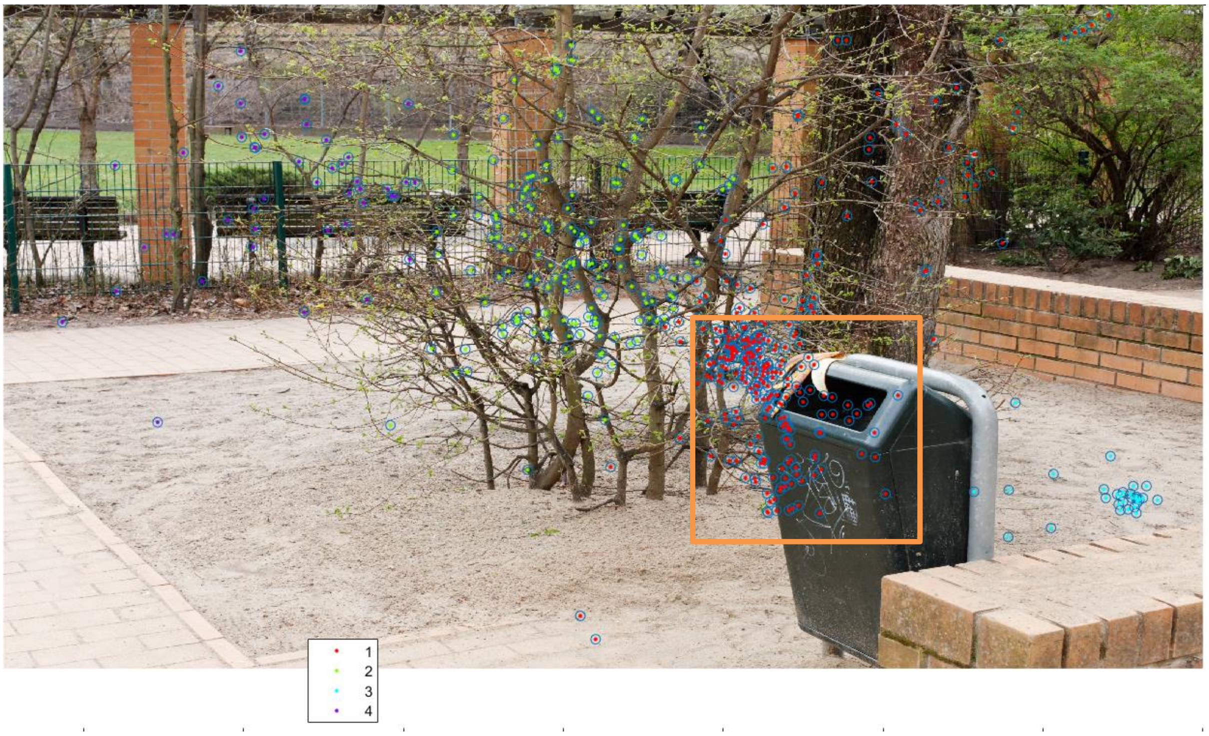
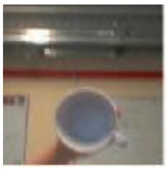
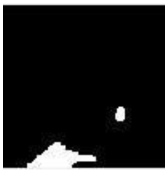
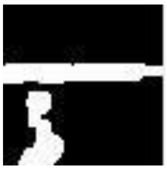
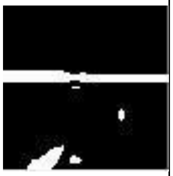
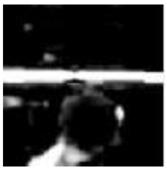
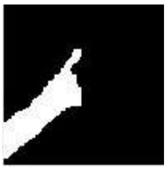
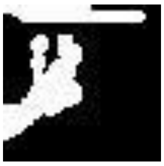
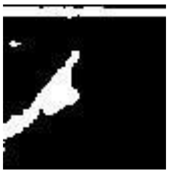
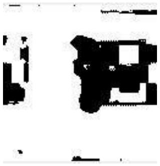
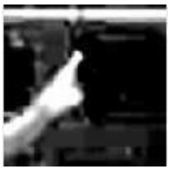
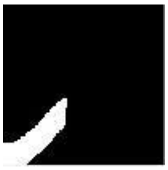
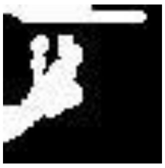
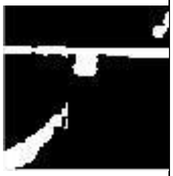
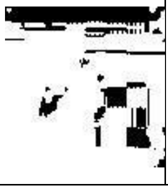
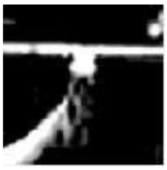


Fig 8.4 Clusters as boundary boxes

Original Image	Ground Truth Image	HSV	CMYK	HSI	Salient Model (RY)
					
					
					

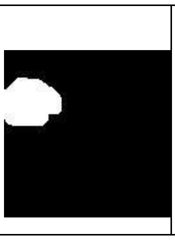
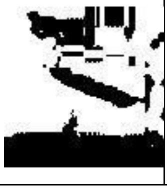
					

Table 5 Result of hand segmentation algorithms for comparison with the original image

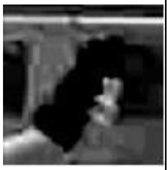
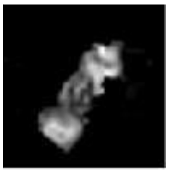
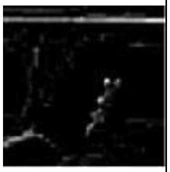
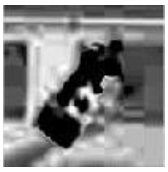
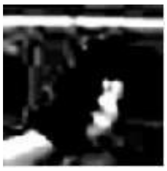
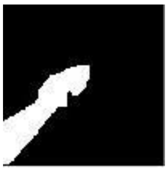
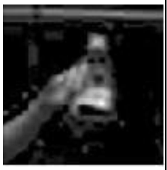
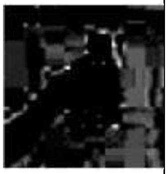
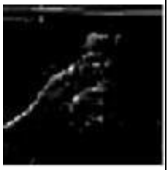
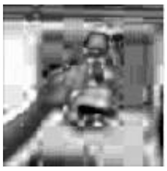
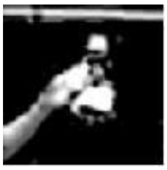
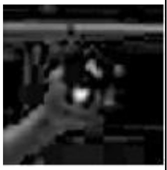
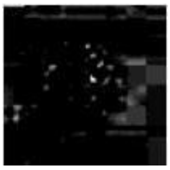
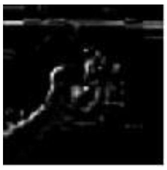
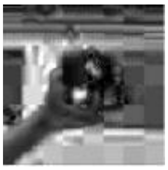
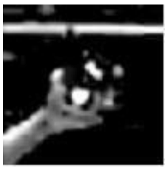
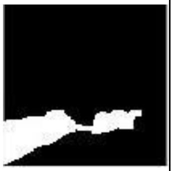
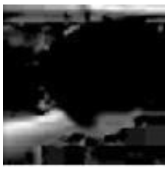
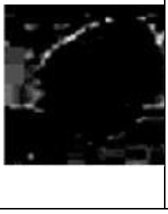
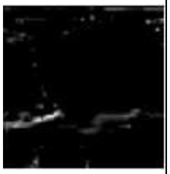
Original Image	Ground Truth Image	RGMAP	GRMAP	RYMAP	RY	RYADD
						
						
						
						

Table 6. Comparison between different salient models

Color model	True positive rate	False positive rate	True negative rate	False negative rate
HSV	82.56 %	15.11 %	84.89 %	17.44 %
CMYK	67.19 %	15.75 %	84.25 %	32.81 %
HSI	98.65	71.56 %	28.44 %	1.35 %
RY Salient Model	75.75 %	73.19 %	26.81 %	24.25 %

Table 7 Confusion Matrix Parameters For colour models

Discussions:

Accuracy is a parameter which demonstrates how close the estimated image is to the original ground truth image. Among the three algorithms of HSV, CMYK and HSI, HSV outstands the other algorithms in the tests on accuracy. Figure 8.5, is a graph plotted with the images being the X-axis and the Y-axis being the corresponding values.

Accuracy for the first quarter of the image dataset seems to be less since the object color overlaps with the hand color. Moving towards the last quarter, the accuracy seems to increase since the object color is completely different from the hand color.

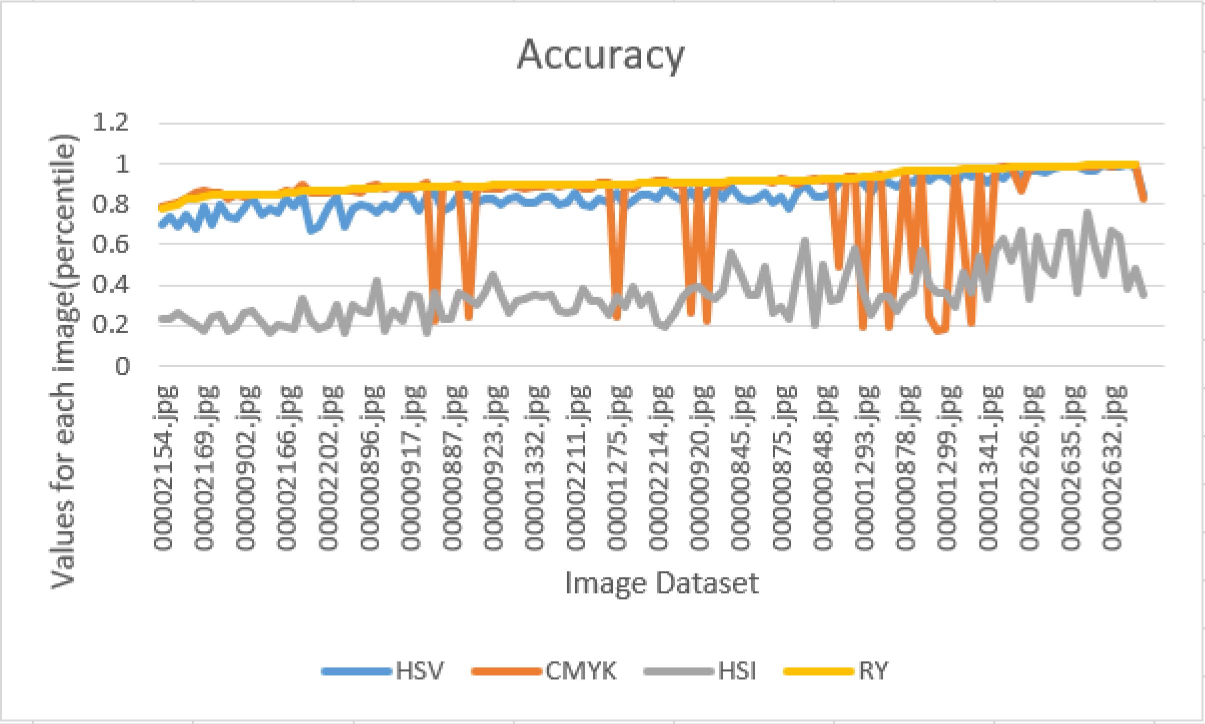


Fig. 8.5 Accuracy graph of image dataset along the values

Accuracy for the first quarter of the image dataset seems to be less since the object color overlaps with the hand color. Moving towards the last quarter, the accuracy seems to increase since the object color is completely different from the hand color.

“What proportion of actual positives was identified correctly” is being demonstrated with **recall** parameter. With reference to Fig 8.6, denoting the graph of image dataset (X-axis) against the values (Y-axis), it can be inferred that HSI hand segmentation algorithm outstands the other algorithms to perform the best in estimating the positives or the original correctly.

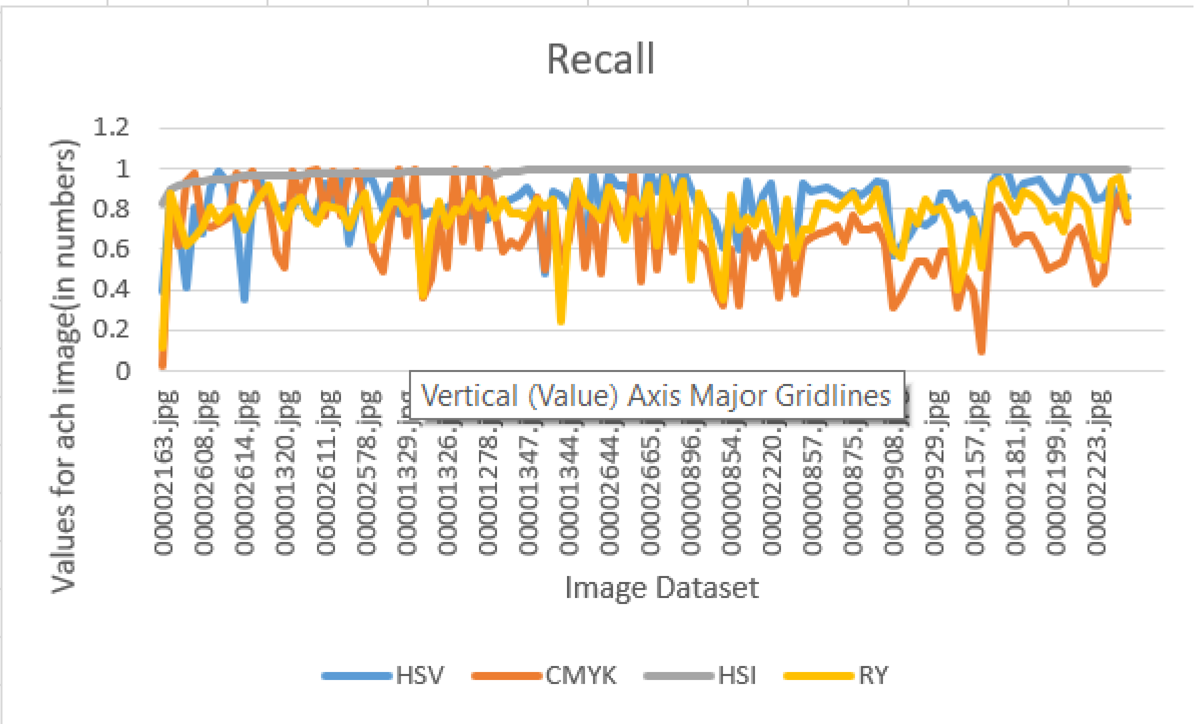


Fig. 8.6. Recall graph of image dataset along the values

There is no deep increase or decrease in case of HSI algorithm in estimating the positives for the graph is very regular and smooth for the entire image dataset. Recall values for HSI algorithm lie between the raranges.9-1 which makes it the best algorithm in detecting positives correctly. HSV and CMYK fail to predict the positives correctly. High recall value of HSI denotes that we are not going to miss the hand in the hand segmentation process while lower and inconsistent values of recall infer that we might indeed lose the true positives of the image. It is better to choose an algorithm with high recall but not at the cost of accuracy.

It demonstrates how much percent of the results are relevant. Recall and precision are extremely important and poles apart since we have recall determining how much percent of the results are totally classified in the algorithm correctly. With reference to the Fig 8.7(Precision graph), there is graph plotted with image dataset on the X-axis and the precision values on the Y-axis. It can be inferred that HSV outstands the other algorithms. HSI might have a very high recall rate, but in terms of precision it's very low. Hence, HSI cannot be taken as the best algorithm since the recall and precision rate should be taken as a combined parameter for classifying the best and worst algorithms for hand segmentation. There is a gradual increase for the HSV algorithm with CMYK performing a little better than HSI but not as good as HSV.

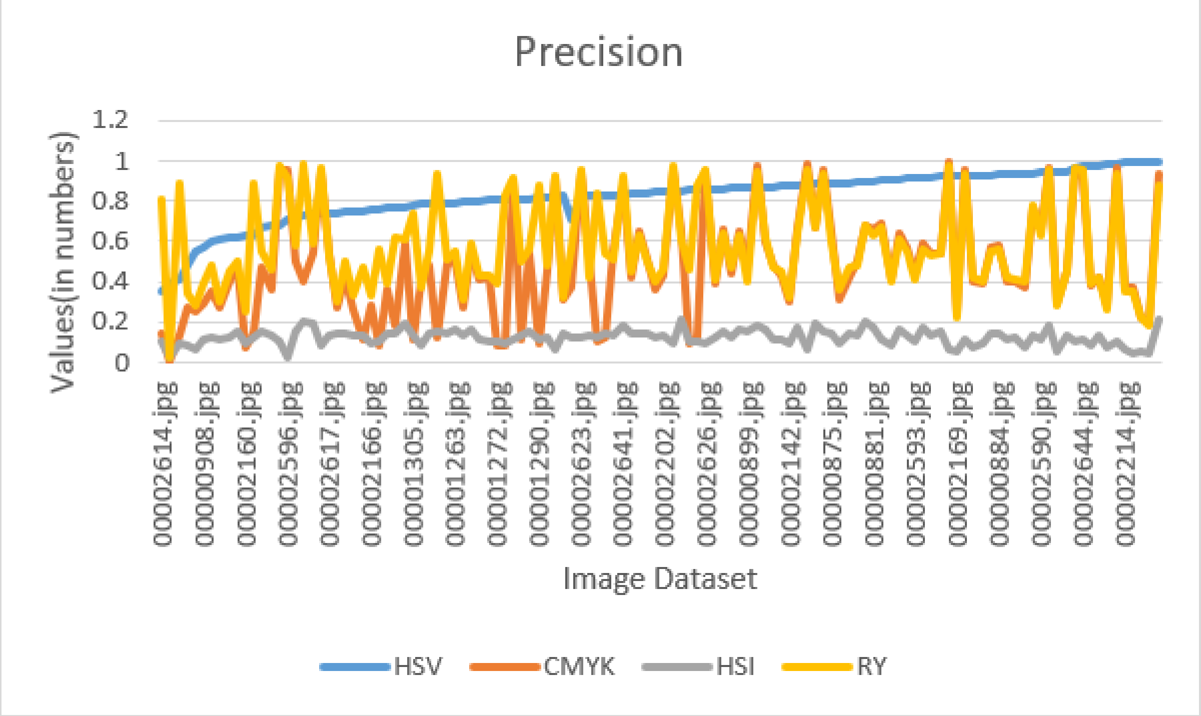


Fig. 8.7. Precision graph of iCub image dataset along the values

From the above experiments and observations, we can conclude that RY additive salient model performs the best among the hand segmentation algorithms. All the objects, irrespective of their color combination values, detected a non-hand area in this algorithm. With the new introduction of the salient model, this paper brings out the additive salient model for hand segmentation. It can also be inferred that HSV performs the second best. In order to choose the best algorithms, we can choose RY additive salient model followed by HSV algorithm. In general, salient models were proposed to find the region of interest, the interesting observations from our experiments stated that different combinations of colours in the salient model can prove to be efficient for determining the skin coloured pixels from the background also. The identification of different coloured pixels lies in choosing the colours of the saliency model. RY additive salient model can be proposed as a hand segmentation colour model in the human-computer interaction field. The future work will be identify different objects based on salient models.

8.3 Region Identification Phase:

Input Dataset:

‘imguag’ is a python library used for augmentation. To create equal number of positive images and negative images, this library is needed. This python library causes you with increasing pictures for your AI ventures. It changes over a lot of information pictures into another, a lot bigger arrangement of somewhat modified pictures.

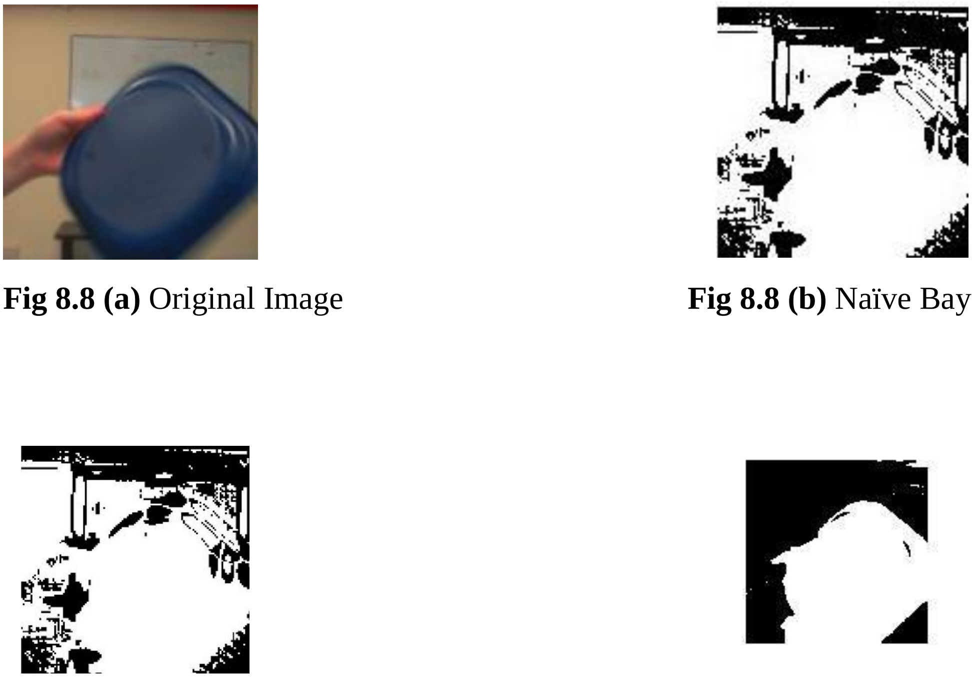


Fig 8.8 (a) Original Image

Fig 8.8 (b) Naïve Bayes

Fig 8.8 (c) SVM

Fig 8.8 (d) Decision Tree

Fig 8.8 Object Detection with Machine Learning Algorithms

Decision Tree Algorithm

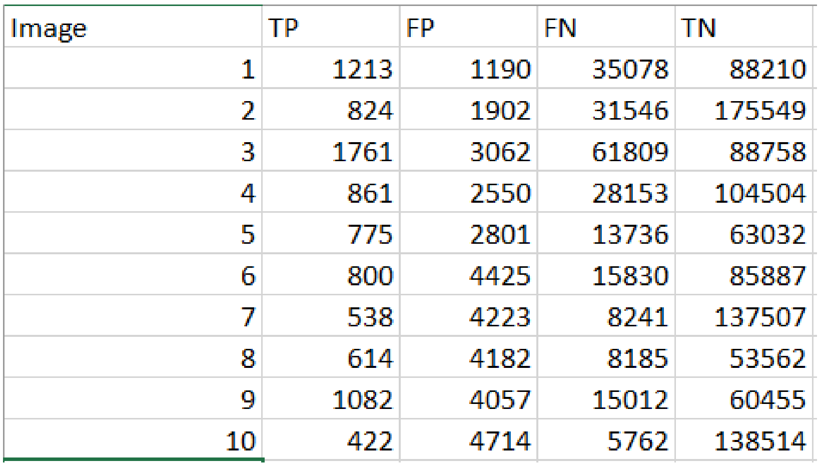


Image	TP	FP	FN	TN
1	1213	1190	35078	88210
2	824	1902	31546	175549
3	1761	3062	61809	88758
4	861	2550	28153	104504
5	775	2801	13736	63032
6	800	4425	15830	85887
7	538	4223	8241	137507
8	614	4182	8185	53562
9	1082	4057	15012	60455
10	422	4714	5762	138514

Fig 8.9 Confusion matrix parameters for Decision Tree

CHAPTER 9

CONCLUSION

9.1 Future Work

The application can be further developed by incorporating the eye movement by using the eye tracking device along with the hand gesture. Extraction of region of interest must be initiated using eye tracking data apart from hand region segmentation. The region of interest is trained with machine learning algorithms for the process of object learning and detection. After analysing the aspects of both hand region segmentation and eye movement, it can be integrated to develop a hybrid system. The hand or eye movement offering highest probability will be chosen as the best to estimate a particular image will be chosen as the gesture for this hybrid system.

9.2 Limitations

The hybrid system has to be tested on a real-time application. Using machine learning and deep learning techniques a generalized system has to developed based on prediction results. The dataset usage sets a boundary on the regions it can be generalized and the results. The incorporation of hand gestures and eye movement would involve a lot of feature extraction and analysis. It is an underlying challenge to build an efficient hybrid system. With the limitation on the hybrid system, it becomes quite challenging to integrate the eye tracking and hand gesture prediction systems. In regard with the eye tracking system, the prediction of the region of interest must be applicable for a wide variety of images. But, it may not be so because training and testing of images would be processed only for a selected image dataset.

REFERENCES

[1]Behavior Evolution of Pet Robots with Human Interaction by Hiroyuki Inoue,Hifumi Miyagoshi, Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)

[2]A Survey Of Eye Tracking Methods And Applications By Robert Gabriel Lupu *and Florina Ungureanu, Publicat de Universitatea Tehnică „Gheorghe Asachi" din Iaşi Tomul LIX (LXIII), Fasc. 3, 2013

[3]Eye Tracking based Human Computer Interaction Applications and their uses by Sushil Chandra and Saloni Maholtra,2015 International Conference on Man and Machine Interfacing (MAMI)

[4]Effectual Training For Object Detection Using Eye Tracking Data Set Sandhya Vishwakarma,D. Radha, Amudha J,Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018)

[5]Eye Communication System for Nonspeaking Patients, Maria Soares da Eira VERSÃO PROVISÓRIA,Dissertação realizada no âmbito do Mestrado Integrado em Bioengenharia Major Engenharia Biomédica

[6]Global Contrast based Salient Region Detection Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi-Min Hu,IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL. XX,NO. XX, XXX. XXXX

[7]S. Krishnan, B.A. Sabarish, Gayathri V., and Dr. Padmavathi S., "Enhanced Defogging System on Foggy Digital Color Images", Computational Vision and Bio Inspired Computing, Part of the Lecture Notes in Computational Vision and Biomechanics book series, pp. 488-495, 2018