

Stony Brook University



OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Gaze Behavior Detection System Based on the Object Image

A Thesis Presented

by

Yunkai Huang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Electrical Engineering

Stony Brook University

May 2013

Stony Brook University

The Graduate School

Yunkai Huang

We, the thesis committee for the above candidate for the
Master of Science degree, hereby recommend
acceptance of this thesis.

Sangjin Hong – Thesis Advisor
Professor, Department of Electrical and Computer Engineering

Peter Milder – Second Reader
Assistant Professor, Department of Electrical and Computer Engineering

This thesis is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Thesis

Gaze Behavior Detection System Based on the Object Image

by

Yunkai Huang

Master of Science

in

Electrical Engineering

Stony Brook University

2013

This thesis presents a design methodology of a low-cost noninvasive gaze tracking system to detect gaze behavior when user is browsing internet or reading material on computer. The user's face image is captured and processed in real-time. By means of C++ and OpenCV library, the system detects face, eye region with Haar feature-based cascade classifier. Eye center is detected by contouring dark area in eye region and finding the center of largest area among contoured dark areas. The detected eye center is mapped to gaze point on computer screen after four point calibration. The average angular error is 1.96 degree, which is comparable to other proposed techniques. During the experiment, the gaze point is displayed real-time with eye movement, and its coordinate as well as the gazed object are recorded in file. The system represents image information in unit area, object, scene, and frame hierarchy structure. With the gaze point data and image information, it is able to analyze gaze duration among objects and understand user's gaze behavior.

To my dearest parents

Quanming Huang and

Yiyi Chen

For their continuous support, encouragement, and love

Table of Contents

List of Figures	vii
List of Tables	viii
Acknowledgements.....	ix
1. Introduction	1
1.1 Overview	1
1.2 The Nature of the Eye	2
1.2.1 Human Visual System.....	2
1.2.2 Types of Eye Movement	3
1.3. Gaze Tracking and Gaze Estimation	3
1.3.1 Eye Tracking Techniques.....	3
1.3.2 Gaze Estimation	6
1.4 Gaze Tracking Application.....	6
2. Gaze Tracking System	11
2.1 Eye Center Detection	11
2.1.1 Face and Eye Region Detection.....	11
2.1.2 Eye Center Detection	12
2.2 Calibration.....	13
2.3 Gaze point mapping	14
3. Gaze Behavior Detection System.....	15
3.1 Gaze Behavior Detection System Model	15
3.2 Image Data Representation Method	16
3.3 Rule based analysis	19
4. Experiment and Evaluation	21
4.1 Gaze Tracking Experiment	21
4.2 Gaze Tracking Accuracy Evaluation.....	22
4.3 Gaze behavior representation	24

4.4 Gaze behavior with gaze point error	29
5. Conclusion and Future Work	32
Reference	33

List of Figures

Figure 1.1. Human visual system[3].....	2
Figure 1.2. Eye movement velocity[4]	3
Figure 1.3 Head-mount gaze tracker[13].....	5
Figure 1.4 Dark and bright pupil images with IR light [17].	6
Figure 1.5 Seven records of eye movements by the same subject. Each record lasted 3 minutes[18]	7
Figure 1.6 Eye tracking system benefits motor-skill disabilities[26].....	8
Figure 1.7 Golden triangle viewing pattern[30].....	9
Figure 1.8 a. Efficient style of scanning a SERP, viewing less than three results, b. Exhaustive style of scanning a SERP, reading all the surrogates in more detail compared with efficient style[36]	10
Figure 2.1 Eye center detection operation flow	11
Figure 2.2 Eye region detection	12
Figure 2.3 Eye center detection	13
Figure 2.4 Eye center detection example. a. eye ROI, b. eye region, c. binary image, d. contoured image, e. bounding rectangular of the contours, f. largest rectangular and eye center	13
Figure 2.5 Gaze mapping	14
Figure 3.1 Gaze behavior system schematics	16
Figure 3.2 Visual information structure	17
Figure 3.3 Data structure relation.....	17
Figure 3.4 Frame information	18
Figure 3.5 Data Structure of frame, scene, obj, and unit area	19
Figure 3.6 Analysis criteria element and related data	20
Figure 4.1 Gaze tracking system user interface	21
Figure 4.2 Gaze object test	22
Figure 4.3 Gaze tracking accuracy evaluation.....	23
Figure 4.4 Gaze evaluation result	24
Figure 4.5 Gaze behavior test Frame 1	25
Figure 4.6 Unit area and gaze point representation for Frame 1	26
Figure 4.7 Gaze duration in Frame 1.....	26
Figure 4.8 Gaze duration by object in Frame 1.....	27
Figure 4.9 Gaze behavior test Frame 2	27
Figure 4.10 Unit area and gaze point representation for Frame 2	28
Figure 4.11 Gaze duration in Frame 2.....	28
Figure 4.12 Gaze duration of object 1 with different gaze point error	30
Figure 4.13 Gaze duration of object 8 with different gaze point error	30
Figure 4.14 Gaze duration difference relative to no error case (%)	31

List of Tables

Table 4.1 Position error at each reference point.....	23
Table 4.2 Gaze duration with different gaze error relative to no gaze error	31

Acknowledgements

The past two years in Department of Electrical and Computer Engineering of Stony Brook University is memorable experience in my life. Thanks to many people who help me to go through this period.

First, I would like to express my deepest gratitude and appreciation to my advisor and mentor Professor Sangjin Hong for his guidance in the research work. His inspiration offers me the opportunity to do the development and research in this interesting project. His academic rigor and research passion impress me and lead me to achieve the thesis as well as to pursuit my future life.

Also, I grateful that Professor Peter Milder takes time from his busy schedule to examine my thesis report and suggests me some valuable advice. I would like to thank Ph.D. candidate Kyung Hoon Kim for his help in the lab as well as my research. I also extend my thanks to Dr. Jung-Min Oh for her assistance to the project.

Finally and most importantly, a heartfelt thanks to my parents without whom I could not have made it anywhere. Although they are far from me in China, their endless love and encouragement is the motivating power for me to continue my life and education in Stony Brook.

1. Introduction

1.1 Overview

There are significantly increase of visual information source all over the world. Most of visual information contains a lot of information such as special story, specific scene, important object, or useful information. People consume these various information almost every day from internet, mobile phone and television. Although visual information has various contents, there is few feedbacks from observers. Content providers have little knowledge about observers. They get to know users' external behaviors such as content selection and channel switching with remote controller or mouse. However, users do not always move mouse to the position where they are browsing and sometimes they leave the television on while doing other things. So these behaviors are still limited to understand users' internal perception. Lack of ability to capture how user perceives certain visual information makes it difficult to know users' intention and what the observer really wants to see is. The development of visual sensor system such as smart TV and eye tracking make it possible to detect observer's nonverbal behavior, such as whether user is looking at the screen or not, where user is gazing at on the screen. However there still remains the problem that we cannot detect how user perceives specific information and interacts with that.

To solve this problem, we propose gaze tracking system by detect eye ball center with web camera and map the eye location to gaze position on screen. In addition to that, we produce gaze behavior analysis platform to connect gaze position with image frame and object in frame. Through our proposed system, it is able to invasively detect users' gaze behavior and makes it possible to understand users' intention.

The organization of the thesis is as follows. Chapter I presents the motivation of the thesis and describes the physiology of eye and eye movement. This chapter concludes the techniques in gaze tracking and gaze estimation. It introduces gaze tracking application in on-line advertisement and search engine. Chapter II introduces the proposed gaze tracking system with the components of eye center detection, calibration and gaze estimation. Chapter III proposes the gaze behavior detection system based on the gaze point data collected in chapter II. The system divides the image information in hierarchy structure as unit area, object, scene and frame. So that it is easy to analyze gaze behavior with rules. Chapter IV provides the procedure to operate the proposed gaze tracking system and evaluates the system. Besides that, it presents the way of analyzing gaze behavior with the proposed system. Chapter V summarizes the thesis and the future work.

1.2 The Nature of the Eye

1.2.1 Human Visual System

Eyes are the windows of our mind. They are the most direct path to interact with the outside world. We use our eyes to perceive the world and obtain surrounding information by gazing objects of interest directly. Eyes not only play roles of input organs but also express our emotions and interests. We could find ones sadness from tearing eyes, or read curiosity from wide eyes. In general, the gaze point reveals the focus of our attention. By recording the direction of gaze, we can deduce what holds the attention of the observer.

Although eyeball is a small spherical object with a radius of around 12 mm, it is regarded one of the most complex organs in our body. There are three major structures that could be observed from outside to the center of eye, including the sclera (white part), the iris (colored part) and the pupil (the black part), as shown in Figure 1.1. Light rays, from light source or reflected from object, enter the eyes, go through cornea and into pupil. The cornea is transparent layer covering pupil and iris and refracts light before it enters the eye. Then the light is refracted by the lens, a transparent structure behind pupil, and is focused onto retina. The retina is a membrane, located at rear part of eye, containing numerous photoreceptors. Photoreceptors are a type of neuron, which are sensitive to light and transform light energy to electrical impulses that are sent to brain through the optic nerve. There are two types of photoreceptors, rods and cones. Rods, about 94% of the total photoreceptors, are sensitive to lights, which helps us to see in dim light and at night. While the cones are less sensitive to rods, but they are good at capture colors and fine details of an object. The highest concentration of cones locates in the fovea, which is on the retina and provide sharp vision to brain [1, 2].

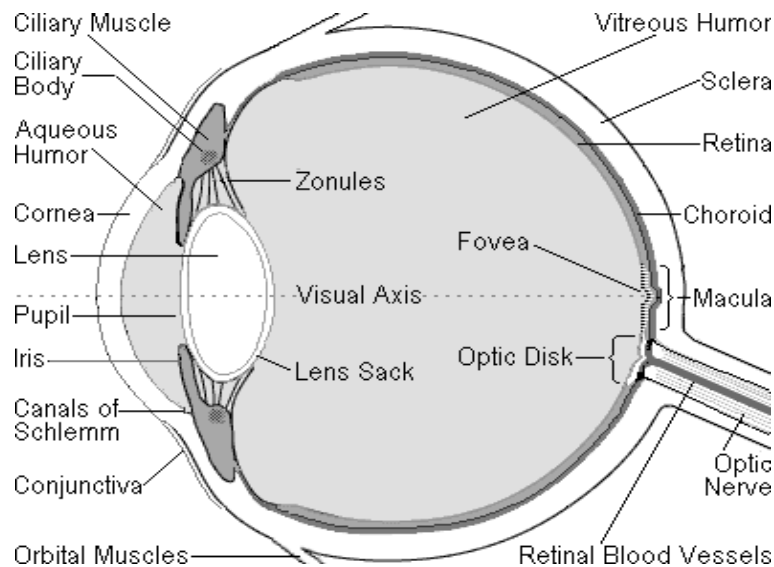


Figure 1.1. Human visual system[3]

The Operation of eye is same as the operation of camera. The iris, similar to shutter of camera, adjust the diameter of pupil in order to regulate amount of light passing through. The retina plays the role of

film in traditional camera or CCD (charged coupled device) in digital camera. The thickness of lens in eyes is changed by the ciliary muscles. So that focused image from various distance is projected on retina.

1.2.2 Types of Eye Movement

Eye movements are classified as three types, saccadic movement, smooth pursuit and fixation.

Saccadic movement is fast sudden jump to relocate eyes. It usually occurs between two focus points. During saccadic movement no information is obtained by visual system in order to maintain perception stable and smooth, namely saccadic suppression. The fast and large movement of saccadic achieves 700 degree per second peak velocity[1], as shown in Figure 1.2. Smooth pursuit occurs when our eyes track a moving object in a range of 1 to 30 degree per second [2]. When the object moving too fast, smooth pursuit performs as saccadic movement. Fixation takes place when eyes stare at static object of interest and obtain information from surroundings. The average fixation time is around 250ms and the velocity of eye is lower than 20 degree per second. The fixation is relative static while there is still micro movement occurring, including drifts, tremors and micro saccades. These movements are within 1 degree of visual angle.

In gaze tracking system, it is easy to distinguish saccade movement with others due to fast movement. Although it is challenging to classify fixation and smooth pursuit from overlapped velocity distributions, low velocity movement reveals that eyes are focusing at either static or moving object [4].

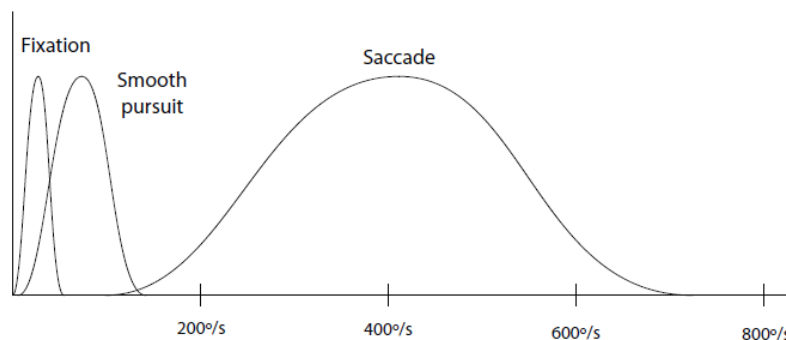


Figure 1.2. Eye movement velocity[4]

1.3. Gaze Tracking and Gaze Estimation

1.3.1 Eye Tracking Techniques

The movement of eyes can be detected by means of electrical, mechanical and visual technologies. With these technologies, there are three eye tracking methods, electro-oculography (EOG), scleral contact lens / search coils, and video-oculography (VOG).

1.3.1.1 Electro-oculography (EOG)

Electro-oculography is a technique to measure the potential change of electrical field with the movement of eyes. The eye ball acts as a dipole with approximately 1mV electrical potential difference from Cornea and Fundus. To detect the potential change, electrode pairs are placed on the skin of either right, left, above or below of the eye [5]. When eye moves to one direction, the electrical potential of the direction results a positive-going change. Electro-oculography is used in application of eye movement tracking for forty years and still used in clinical diagnose. The wide usage thanks to its cheap and non-invasive advantage. As the electrode pairs are attached to face, its detection is not affected by head movement and wearing glasses or lens [6]. However, the technique requires accurate pair placement and reliable adaptive electrical signal processing, that depends on individual tester.

1.3.1.2 Scleral Contact lens /Search coils

The technique introduces specific contact lens wired with coil. During the experiment, a large magnetic field is applied to detect the electrical potential difference with the moving of coil on lens. The technique provides very high temporal and spatial resolution as 10 arc-seconds [7]. Since the technique is invasive and uncomfortable to use, it is limited in research laboratory usage.

1.3.1.3 Video-oculography (VOG)

Video-oculography is a non-invasive technique in eye tracking. By processing images from camera, it is able to capture the location of eye. Usually it does not require extra equipment and is tolerant to head movement that makes it the most popular technique.

VOG technique uses computer vision for eye detection and gaze estimation. Eye image could be identified by means of intensity distribution of different part of eye structure, including cornea, iris, pupil or its circular shape. The algorithm can be classified into three types: shape-based, appearance-based and hybrid. The shape-based method takes advantage of prior model of eye shape, which is extracted in either local point features or the contours around eye region. The features of eye corner, pupil and iris plays major role in shape-based methods. The appearance-based method detects eye by matching template to eye images. The templates are trained from a large set of images based on different image intensity or various subspace. Hybrid method combines shape-based method and appearance-based method to compensate each other and exploit benefits from each methods [8].

Although there are numerous algorithm dealing with eye detection, the detection still remains challenge due to variation change in appearance and low resolution in image. That includes viewing partial of eyeball due to eyelids occlusion and eye openness, deviation in eye size and color, darkness of eye due to external light, and head pose.

Based on hardware configuration of VOG eye tracking system, it is possible to classified into two typed: head-mount tracker and table-mount tracker.

Head-mount gaze tracker requires the user to wear specific helmet or a pair of glasses that are equipped with camera or light source [9-11], as shown in Figure 1.3. The camera is stable relative to head position, so that the system is able to apply in the situation of fast or wide range of head movements. Moreover, as the camera is close to eye, it could capture high resolution eye movement video [12].

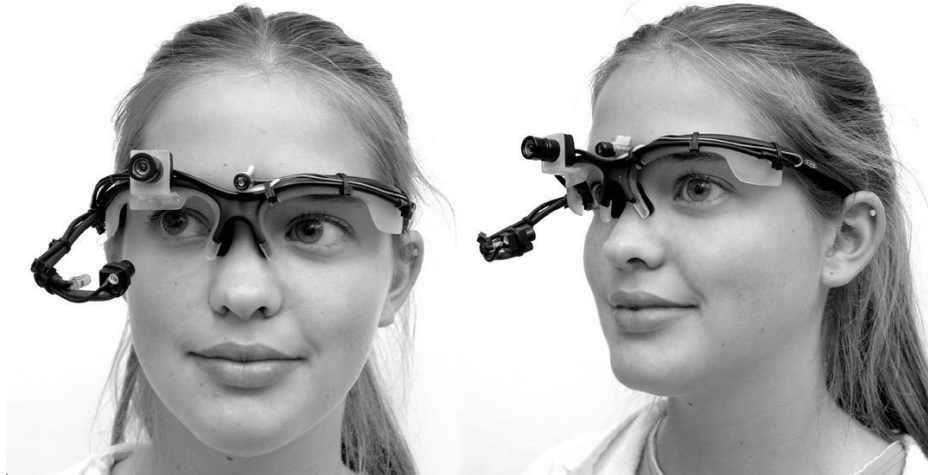


Figure 1.3 Head-mount gaze tracker[13]

Table-mount gaze tracker is usually placed around the computer screen or uses the embedded web camera instead. In order to set up relative position between light source and camera, both of them are designed within a customized bar in some commercial products [14, 15]. There is no equipment attaching user and allows free head movement in certain extent. That makes the gaze system really non-invasive.

In VOG, there are two illumination methods applied, passive nature light and active infrared (IR) light. With natural light, appearance-based methods are employed to detect iris with high contrast compared to sclera. However, the method requires a light environment to obtain clear contrast. When the user's eye is black pupil and dark or brown iris, it becomes problematic to detect pupil. System based IR light improves the imaging contrast and provides a solution for stable illumination. The IR light reflection or refraction formed on cornea distinguishes pupil from iris with infrared sensitive camera [16]. It appear bright pupil or dark pupil depending on the location of IR light source, as shown in Figure 1.4. When the light source is placed co-axial with camera, it occurs bright pupil. That is same as "red eye", when light from light source reflecting from retina along same direction back to camera as retroreflector. While when the light source is off-axial with camera, most light will refract on corneal and be not obtained by camera. In this case, the pupil appears dark. As IR light is not visible to human, the method does not distract user's attention nor affect gaze tracking experiment. However the high contrast image by IR method is not robust, when there is external light source in outdoor condition. Moreover, when the user wears glasses or contact lenses, the reflection will be too complicated to analyze.

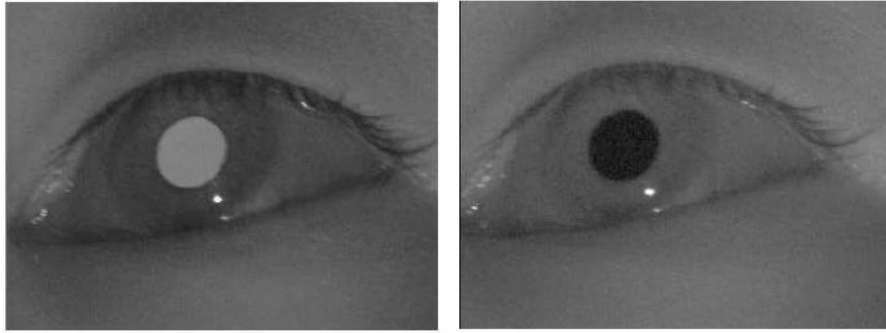


Figure 1.4 Dark and bright pupil images with IR light [17].

1.3.2 Gaze Estimation

Gaze Estimation is to translate detected eye features into corresponding image data. In the VOG technique, these features include the position of pupil center or eye corner, contour of cornea, and reflections from cornea. With these information, gaze estimation will result the coordinate of eye center and gaze direction. Two types of featured-based gaze estimation are developed: geometry-based methods and interpolation methods. The application of these methods depends on the hardware setup and available knowledge of relative geometric relationship of the system. The geometry-based method obtains 3-D model in advance, with the geometric parameter among eyes, camera and screen. So that it builds the relationship among them and calculates the gaze point on screen. The geometry-based method is usually used in head-mount tracker where the relative position of head and camera is stable. On the other hand, the interpolation method estimates linear or polynomial relationship between eye center position and gaze coordinate. It calculates gaze points on screen by means of intersection [8].

It is essential to practice calibration before test in both of the methods, in order to obtain model parameters or coefficient of polynomial. The user is required to follow specific instruction in the calibration process. When the user is gazing at four or nine calibration points, a set of coordinate data will be recorded as the reference data of gaze estimation.

1.4 Gaze Tracking Application

Eye tracking research and application has developed for over fifty years [2]. As gaze is a most important way to obtain information from the environment and to express attention of human beings, gaze tracking can provide good understanding of people's behavior. In 1967, Yarbus tries to find eye-gaze pattern affected with different task [18], as shown in Figure 1.5. He presents same picture and instructs a subject to answer seven different questions. The records of eye movements reveals the observation and thought process of the subject.

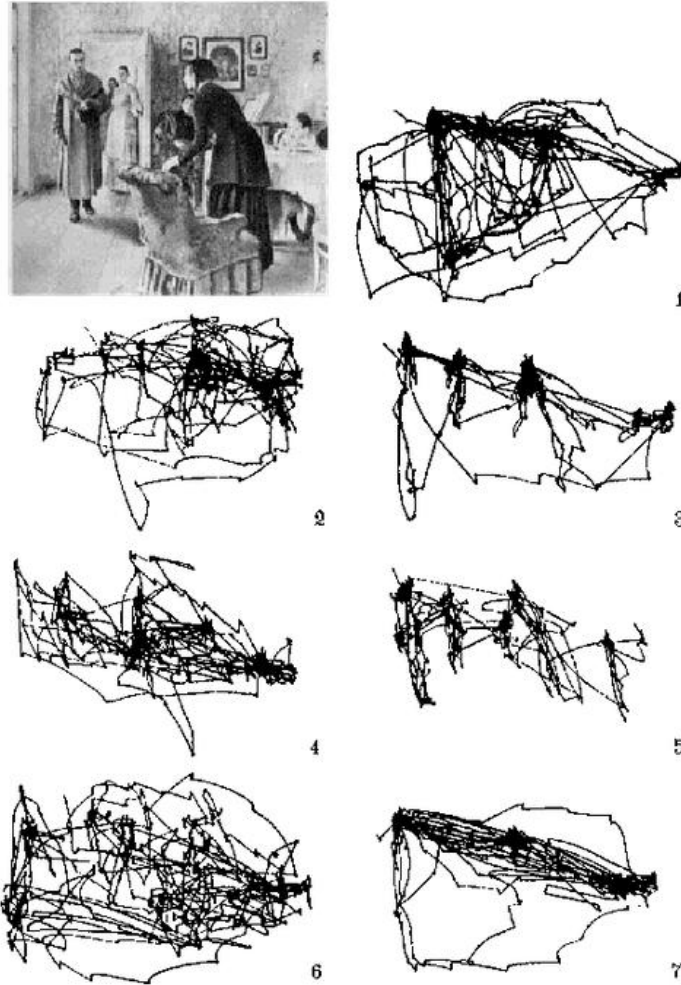


Figure 1.5 Seven records of eye movements by the same subject. Each record lasted 3 minutes[18]

The first well-known application of eye tracking is in the research of reading experiments. Eye fixation and saccade duration when reading is analyzed. Textual and typographical content influence eye movement. The way of reading silently and aloud will also affect eye movement [2]. In the perception of scene and perception of film research, four basic features: color, size, orientation and direction of motion influence eye movement [19]. Based on these research, eye tracking is able to detect human performance in vision, cognition and attention. Now eye tracking is applying in flight simulation for Primary Flight Display (PFD) which is an important source of information during flight [20]. It also helps when developing new flight electronic maps [21]. Eye movement plays role in analyzing driver behavior, in order to prevent deficiencies in visual attention and improve driver training as well as accident countermeasures.

In computer science field, gaze begins to take the place of mouse and keyboard. We are able to select items on interface or objects in Virtual Reality (VR) [22]. People could scroll text on cell phone with eye movement [23]. Instead of game controller, our eyes would play computer interactive games [24]. The application of eye tracking within Human-computer Interaction (HCI) also bring benefits for server

motor-skill disabilities in Figure 1.6. Eye typing offers an invaluable communication tool for them [4]. Besides selective application of eye tracking, Gaze-Contingent Displays (GCDs) extends its roles as a passive indicator. With GCDs, system can display details where the user is gazing, while tailor the periphery contents. That reduces bandwidth requirement for video telephony or graphical applications [25].



Figure 1.6 Eye tracking system benefits motor-skill disabilities[26]

It is the era of eyeball economy. Marketing and advertisement focus on how to obtain how attention from customers. From Yellow Pages [27], to magazines [28], from television commercials to web search site, gaze is always the interest of both business and research. Webpage content providers want to have insights into user behaviors to homepage marketing, online news and online search environments. The investigation of eye tracking for online information retrieval has become focus of researchers since 2003 [29].

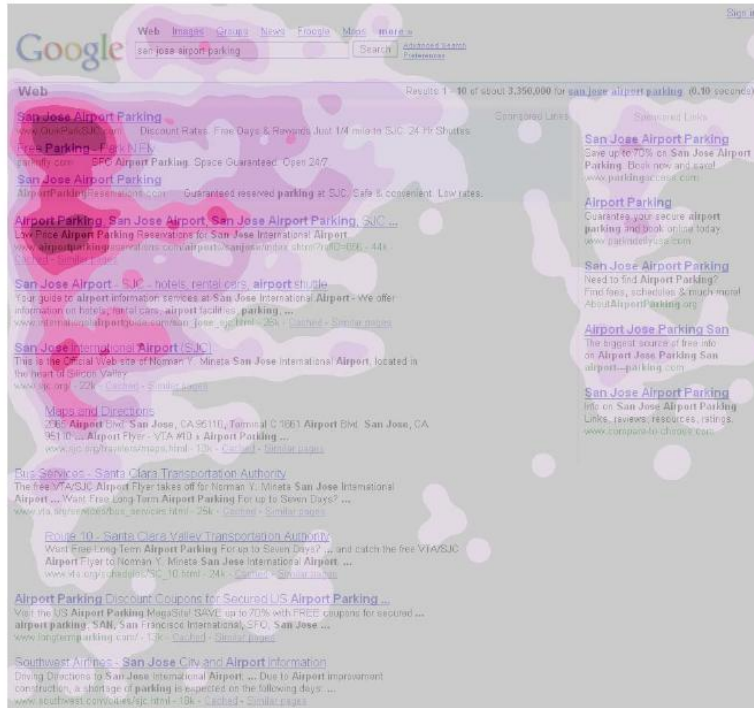
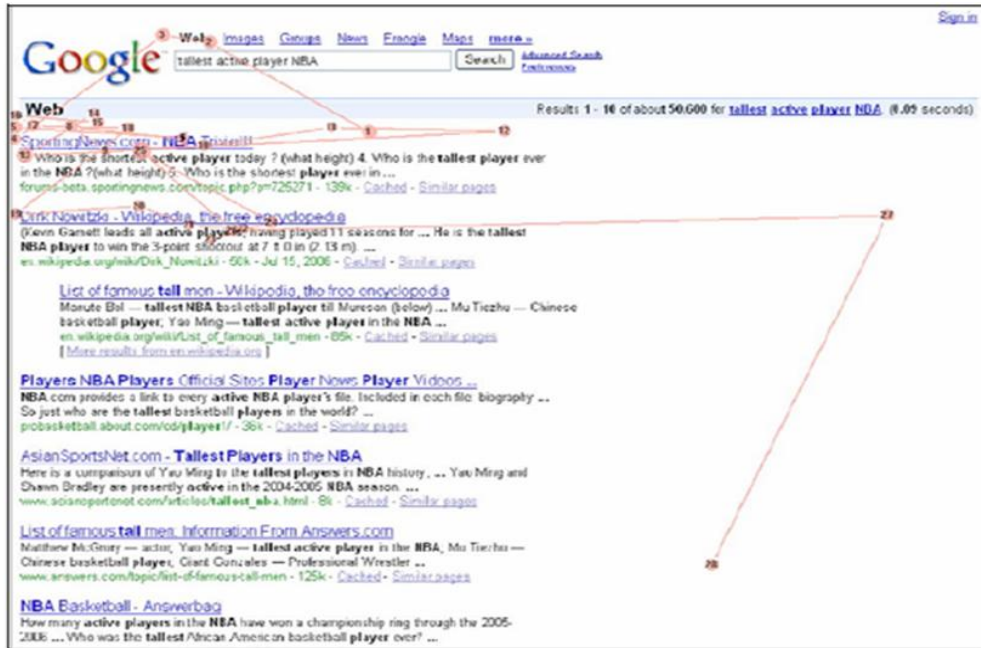


Figure 1.7 Golden triangle viewing pattern[30]

Many metrics is used for web-based eye tracking analysis, including length of scanpath, determining the fixation density in a region, and comparing fixation [31, 32]. The sequence alignment of scanpath is analyzed by measuring the similarity between two scanpaths over same stimulus [33]. Probabilistic models are configured to represent eye movements [34]. Gaze location characteristics model is generated to predict visual attention that user may receive [35]. Gaze pattern for search engine result page (SERP) is found as F shape or a Golden Triangle [30] as shown in Figure 1.7. The heatmap of the gaze pattern shows that people would like to pay major attention on top of few search result. The individual scan patterns are classified as economic and exhaustive style [36], as shown in Figure 1.8. The users' query intent is analyzed into informational, navigational and transactional pattern [37]. Combined with other behavioral data, such as server log data and mouse click data, eye tracking present users' behavior when scanning and searches for online content [38, 39]. The understanding of user behavior in turn benefits the design of website and the content representation [38].



a



b

Figure 1.8 a. Efficient style of scanning a SERP, viewing less than three results, b. Exhaustive style of scanning a SERP, reading all the surrogates in more detail compared with efficient style[36]

2. Gaze Tracking System

The gaze tracking system captures video frames from a web camera. With image processing it is able to detect face, eye region and eye center by frames. The system is composed of calibration stage and test stage. Initially, the user moves eyes by instruction and the system collects the calibration data for reference. In test stage, the system maps the detected eye center to computer screen in order to track the gazing position on screen.

The system is based on C++ and OpenCV 2.43, and is running on a laptop with Intel i7 1.80GHz, RAM 4GB. The resolution of camera is 1280*720.

2.1 Eye Center Detection

Eye center is detected recursively, from face region detection, eye region detection and eye center detection, as shown in Figure 2.1.

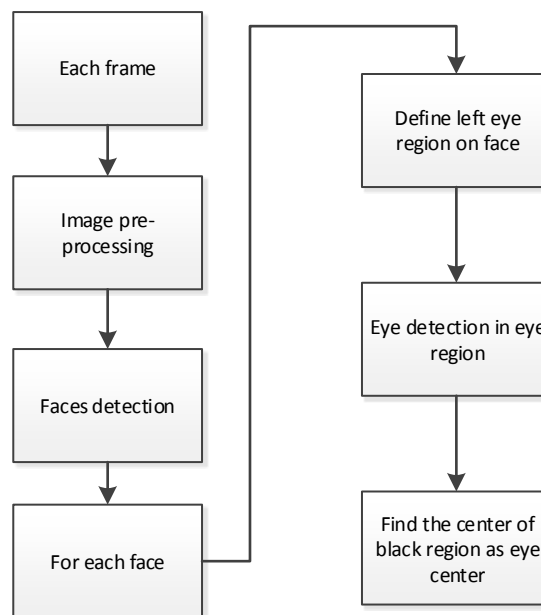


Figure 2.1 Eye center detection operation flow

2.1.1 Face and Eye Region Detection

Each frame of color RGB image is read from a web camera. It is converted into gray image. To increase image contrast, the histogram of gray image is normalized so that the sum of histogram bins is 255.

Within the whole normalized image, faces are detected by applying Haar feature-based cascade classifier, as shown the blue box in Figure 2.2. The classifier encodes the existence of oriented contrasts exhibited by a human face and their special relationships in the image [40, 41]. The classifier is trained

with hundreds of positive and negative examples, provided by OpenCV library in file “lbpcascade_frontalface.xml”. The classifier is moved through the whole image and to search for the matching object. It is resized in multiple scales to find faces with different size. In each of detected face region, both left and right eye regions is able to be detected in the same way with classifier “haarcascade_eye_tree_eyeglasses.xml”. In order to reduce computation, only left eye is used for gaze tracking, as shown red box in Figure 2.2. Besides that, before eye region detection, the area for searching eyes is narrowed in an eye ROI (region of interest), a rectangular with $0.3 \times \text{face width}$ and $0.3 \times \text{face height}$, as shown the pink box in Figure 2.2. The left eye ROI is defined in the position of $0.2 \times \text{face height}$ from top of face region and $0.15 \times \text{face width}$ from left of face region.

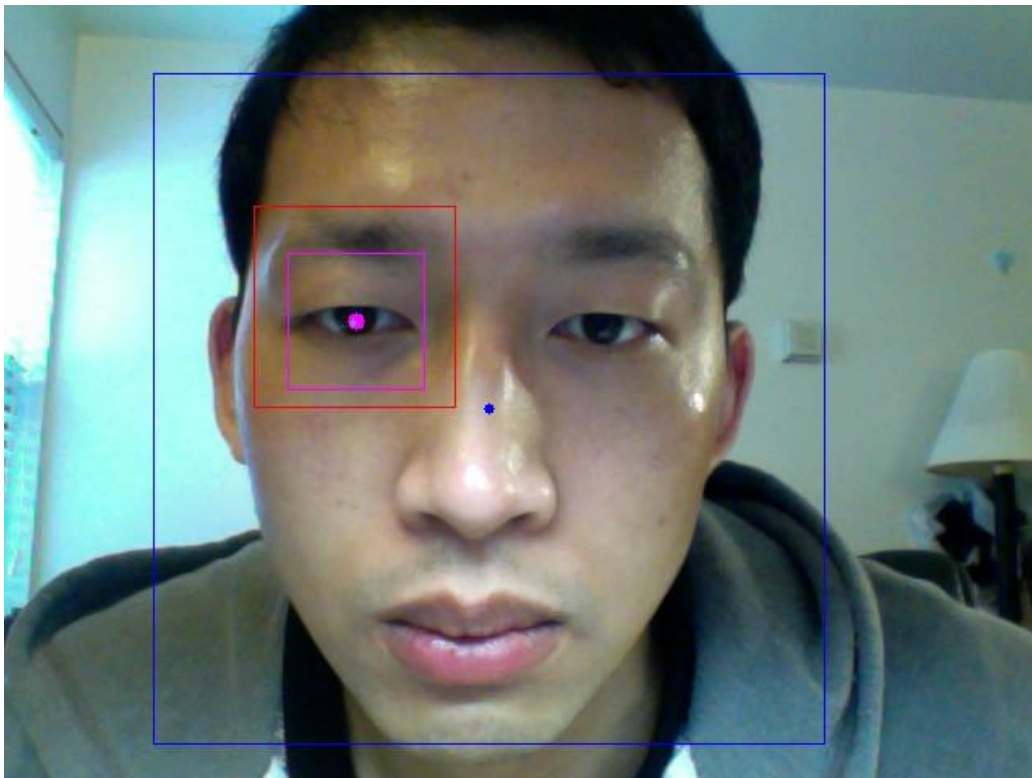


Figure 2.2 Eye region detection

2.1.2 Eye Center Detection

Once the eye region is detected, the eye center is detected as shown in Figure 2.3. A gray-scale threshold is applied to the region in order to get binary image to extract the dark eye ball region. The threshold is set according to the contrast of live frame images. In the binary image, it contours all dark regions and finds the up-right out-bounding rectangular for each dark region. In order to eliminate false dark regions among the contours, such as eyelid and medial canthus, the rectangular with largest area is selected as eye ball. The center of the largest rectangular is defined as eye center. Figure 2.4 shows an example of the eye center detection procedure.

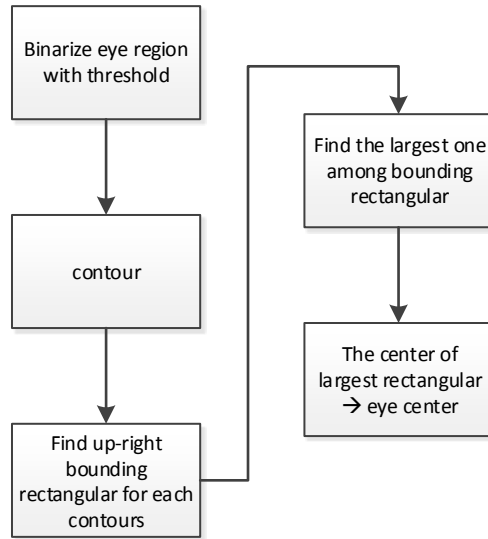


Figure 2.3 Eye center detection

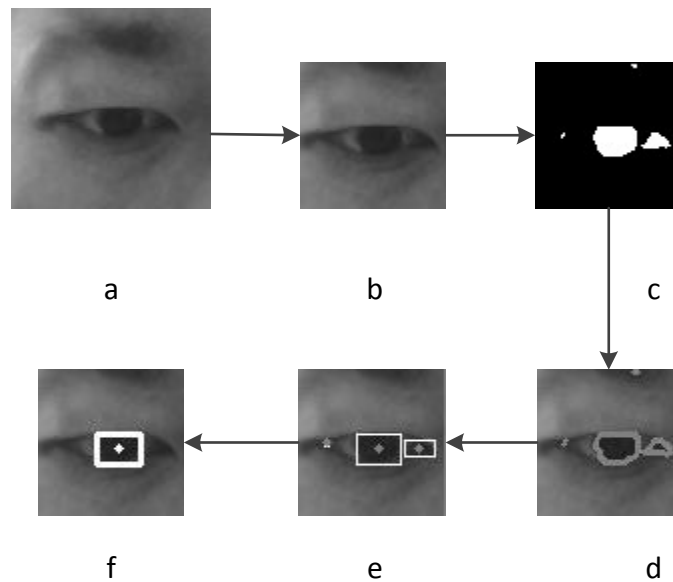


Figure 2.4 Eye center detection example. a. eye ROI, b. eye region, c. binary image, d. contoured image, e. bounding rectangular of the contours, f. largest rectangular and eye center

2.2 Calibration

In the calibration stage, the eye moving region is mapped to the region on screen. So that the eye center movement is able to match the gazing point on screen, as shown in Figure 2.5.

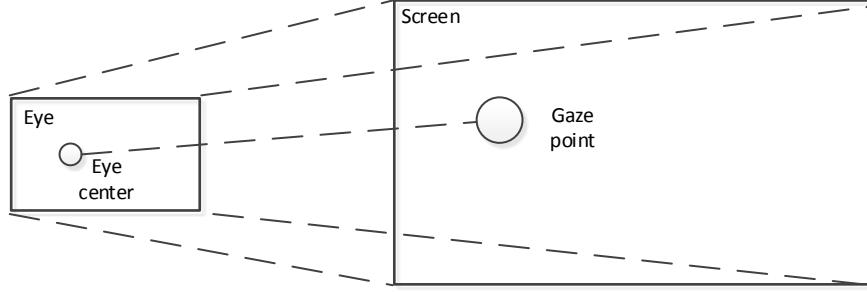


Figure 2.5 Gaze mapping

The left eye center is selected to track gaze point. Four buttons at UI edge, including UP, DOWN, RIGHT and LEFT, are used as reference. During the calibration, a user presses one button (e.g., UP) by instruction, while keeping gazing at the button. After the system collects eye center points, it will instruct to gaze at next button. In order to eliminate eye movement instability and errors in image processing, it collects multiple eye center points consecutively when the user gazing at UP and averages the coordinates of these points to get the UP gaze reference (x_{UP}, y_{UP}) as follow.

$$x_{UPRef} = \frac{\sum_{i=1}^n x_i}{n}, y_{UPRef} = \frac{\sum_{i=1}^n y_i}{n}, \quad (2.1)$$

where (x_i, y_i) is eye center coordinate at i th collection, (x_{UPRef}, y_{UPRef}) is UP gaze reference coordinate. The system obtains the mapping between the coordinates of UP button and UP gazing reference coordinates. Repeat the procedure of gaze and collection, the system obtains the gazing reference coordinates of UP, DOWN, LEFT and RIGHT.

2.3 Gaze point mapping

In gaze tracking test stage, each frame when eye center is detected, it is mapped to the eye gaze point coordinates as follow,

$$x_{gaze} = \frac{x_{eye} - x_{LEFTRef}}{x_{RIGHTRef} - x_{LEFTRef}} \times (x_{RIGHTBut} - x_{LEFTBut}) + x_{LEFTBut}, \quad (2.2)$$

$$y_{gaze} = \frac{y_{eye} - y_{DOWNRef}}{y_{UPRef} - y_{DOWNRef}} \times (y_{UPBut} - y_{DOWNBut}) + y_{DOWNBut}, \quad (2.3)$$

where (x_{gaze}, y_{gaze}) is eye gaze point coordinate, (x_{eye}, y_{eye}) is the detected eye center coordinate, $x_{RIGHTRef}$ and $x_{LEFTRef}$ are RIGHT and LEFT gaze reference x coordinates, y_{UPRef} and $y_{DOWNRef}$ are UP and DOWN gaze reference y coordinates, $x_{RIGHTBut}$ and $x_{LEFTBut}$ are RIGHT and LEFT button x coordinates, y_{UPBut} and $y_{DOWNBut}$ are UP and DOWN button y coordinates. The coordinates of eye gaze point is checked whether it is within boundary of screen region. Then it is stored in file with system sampling time.

3. Gaze Behavior Detection System

The gaze behavior detection system is to map the detected gaze point into the information on image. The image information is classified in hierarchy as unit area, object, scene and frame. The system can analyze user's gaze behavior by time as well as the image information.

3.1 Gaze Behavior Detection System Model

The basic structure of the system model is constructed with five module as shown in Figure 3.1. The definition of each module is as follows:

M1. Data capture module. In this module, image data, that the user is watching, is loaded from the content provider. Meanwhile user's eye movement data is captured by a web camera.

M2. Object data identification module. This module identifies the relations of content and object created by content provider. Then they are stored into database.

M3. Eye data identification module. This module maps the eye tracking data to the object area information from M2.

M4. Gaze behavior data profiling module. This module combines the relations between objects and image area from M2 with user's gaze data from M3. As a result, it produces user's gaze behavior data based on object.

M5. Rule checking and monitoring module. This module uses inference rules and statistic rules show user's gaze behavior.

From M1 to M5, each module is closely connected with data flow. M1 and M2 stores image data and their metadata respectively. Metadata is the data of block information. It present the set concept in group, such as frame set, object set and rule set. M1 sends content information, block information and image information to M2. Block information is composed of multi frame units that express specific context in the content. Then, M2 identifies general metadata and image metadata by means of information from M1. M1 transfers user information and eye information data to M3. M3 discriminates user's gaze area and gaze time from eye information with object area information from M2. Next M4 receives metadata from M2 and gaze data from M3, and builds the profiles of user's gaze behavior data on base of objects. Researchers can use the profiling data to find the statistical user's behavior pattern. Besides that, M5 is defined as a rule checker to infer user's intention within time period.

In this thesis, we assume that time element has two perspective. One is the time for image loading (T_i) and the other is the time for detection user's eye information (ST_s). At time T_i , image (I_i) is delivered to user. User watches the image frame at ST_s and at same time the system captures user's eye movement as gaze (G_e). I_i and G_e compose one set.

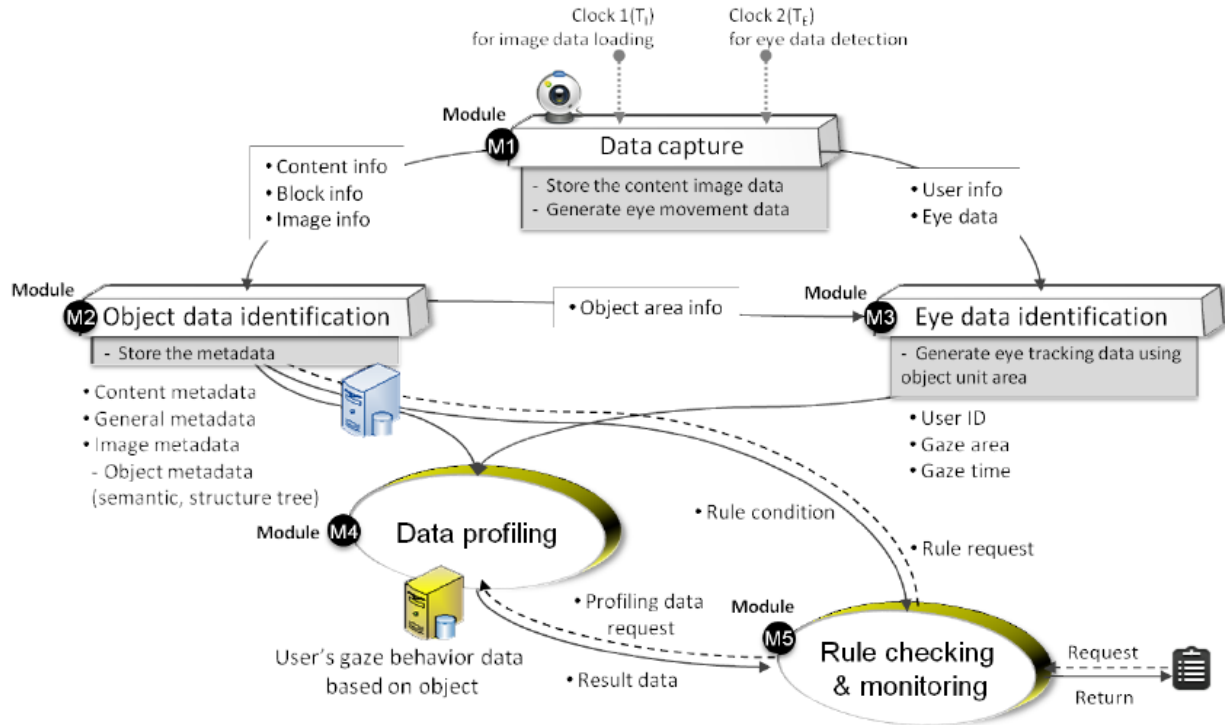


Figure 3.1 Gaze behavior system schematics

3.2 Image Data Representation Method

Images are composed of pixels which is defined as unit point in this thesis. Unit points are too small and too many to define an object. Moreover, even user gazes at same position, the gaze point will not locate at same unit point due to gaze detection error. Thus, unit area is defined as groups of unit points to distinguish objects. The visual information of object represents in hierarchy model from unit point, unit area, object and image. Image may have only one object or a lot of objects. An object has at least one unit area based on the size of object. The size of unit area is not fixed, so that it is flexible to the size of objects. When the objects included in the image are large, the size of unit area is also divided into big scale. If the object is tiny, unit area has to split it into small size. As it reduces calculation consumption when seeking object with unit area, system process would be efficient.

When considering characteristic of information in images, scene block is defined by groups of unit areas with common scene content. The scene block can be interpreted into content type like action, romance, eating or item type like fashion, food and so on. The single image may have single or various scene block in it. Scene comprised of several objects which share same content.

So the visual information structure of one single image frame (F_i) consists of scene block (S_j), object (O_n), unit area (A_m) and unit point (P_z) as shown in Figure 3.2. In hierarchy, a frame consists objects. Individual object is composed of unit areas. Unit area is represented by multiple unit points. While several objects

with common character compose one scene. The hierarchy structure leads to the inter-related data structure as shown in Figure 3.3

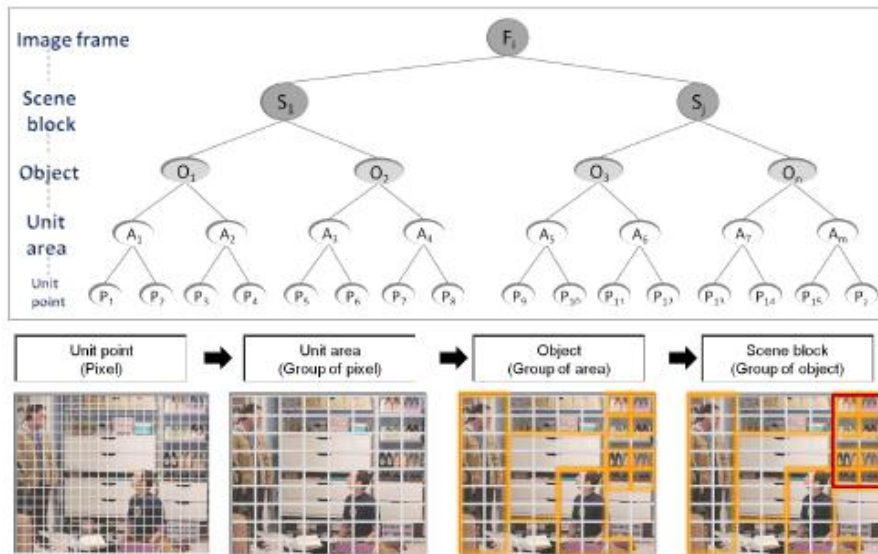


Figure 3.2 Visual information structure

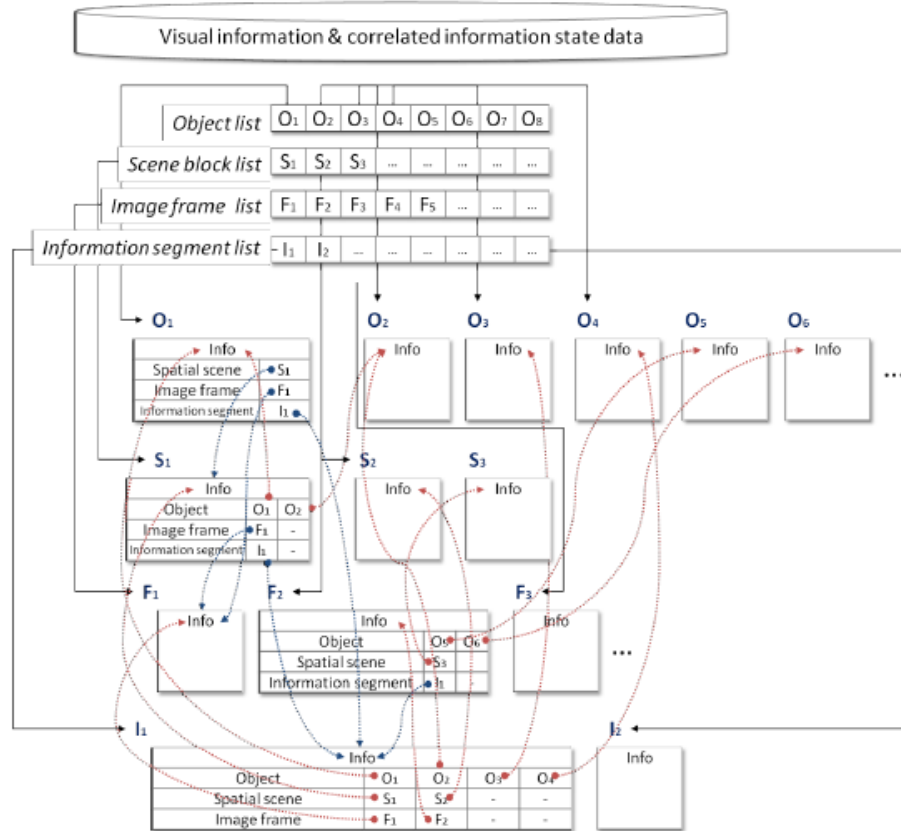


Figure 3.3 Data structure relation

When considering multiple image frames, there is correlation between image frames. Multiple image frames bring context to light such as same episode, same object, or same spatial scene. To represent shared information, shared correlation is defined to link the image frames according to the object and spatial scene. For example, if F_1 and F_2 contains same object O_1 , shared information between F_1 and F_2 is represented as $CO_1=[F_1, F_2]$, and is included in image information. When F_1 and F_2 contains same spatial scene S_1 , $CS_1=[F_1, F_2]$ will be added to image information. Thus, when an image frame is loaded, following structure information, in Figure 3.4, is necessary in order to maintain data within the system.

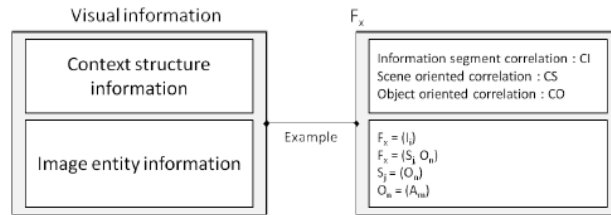


Figure 3.4 Frame information

Based on the proposed data representation model, the hierarchy data structure of frame, scene, object and unit area are constructed as shown in Figure 3.5 . Each unit area has geometry members of x, y, width and height, as well as its image information such as which object or which scene it is covered by. Each object contains an index list of unit areas. Each scene has an index list of objects. Each frame is composed of an index list of objects and an index list of scene.

First, unit areas of whole frame are initialized with the unit area resolution by assigning geometry members of x, y, width and height. Then, based on image information, the unit area list of object is filled with the index of unit areas that the object belongs to. Meanwhile the inObjIdx of the corresponding unit area is assigned with the index of the object. Scenes are initialized in the same way with index list of objects and index list of unit areas that the scene belongs to. The index of the scene is filled into the inScenIdx of the corresponding unit area. Finally, index list of objects and scene are assigned into the frame.

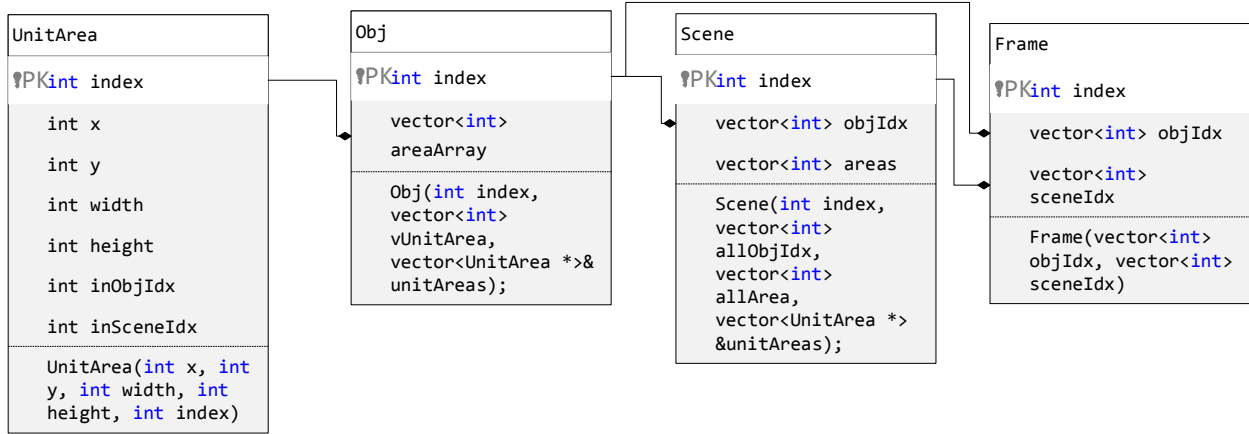


Figure 3.5 Data Structure of frame, scene, obj, and unit area

When gaze points of the frame is obtained from eye tracking module, each gaze point has coordinate on the frame. The corresponding index of the unit area is calculated as follow,

$$\text{Unit area index} = \frac{x}{\text{unit width}} + \frac{y}{\text{unit height}} \times \frac{\text{width of frame}}{\text{unit width}} + 1 \quad (3.1)$$

where gaze point coordinate is (x, y) , size of unit area is $\text{unit width} \times \text{unit height}$, width of frame is obtained from image information and the plus one at end is due to index of unit area starting from 1. So that each gaze point is mapped to one unit area in the frame.

Then by checking the inObjIdx of the unit area in Figure 3.5, it is easy to acquire the corresponding object index where the gaze point locates. After traversing all gaze points by gaze sampling time, we can count gazing times on each object in the frame and sum into gaze duration by object. The comparison of the ratio of gaze duration could reveals user interest among objects.

3.3 Rule based analysis

Image data incorporate rules to represent hierarchy structure and contextual structure. These rules are defined as representation rule. It supports the distinction of object which is composed with combination of unit area information in a uniform way, and spatial scene which is consist of objects. Furthermore, it shows the correlation between image and temporal context.

The system model has analysis rule to interpret gaze behavior to show an implicit perception of the observer with profiling data. Users' interest could be obtained by running rule check module in queries like: how long observer looks specific entity during play the visual information, or who are the observers look specific entity first? The analysis criteria element and related data is shown in Figure 3.6. The element of "where" indicates monitoring and searching area. The element of "what" means target to analysis of the gaze behavior. The element of "who" is the user of eye tracking data. The element of

“how long” and “what order” describe gaze duration and gaze order by index. Through the perspective, users’ behavior could be quantified in order to analyze his/her interest.

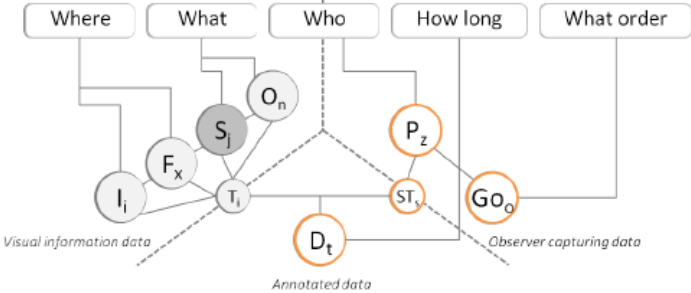


Figure 3.6 Analysis criteria element and related data

4. Experiment and Evaluation

4.1 Gaze Tracking Experiment

When the system starts, the user sits in front of the computer with his/her gaze point at the center of the screen. The user interface presents detected face, detected eye and eye center detection in lower right corner as shown in Figure 4.1. The user can observe the detected eye center real-time and set the gray scale threshold according to the light and refraction environment. So that the system is able to contour eye ball and finds eye center.

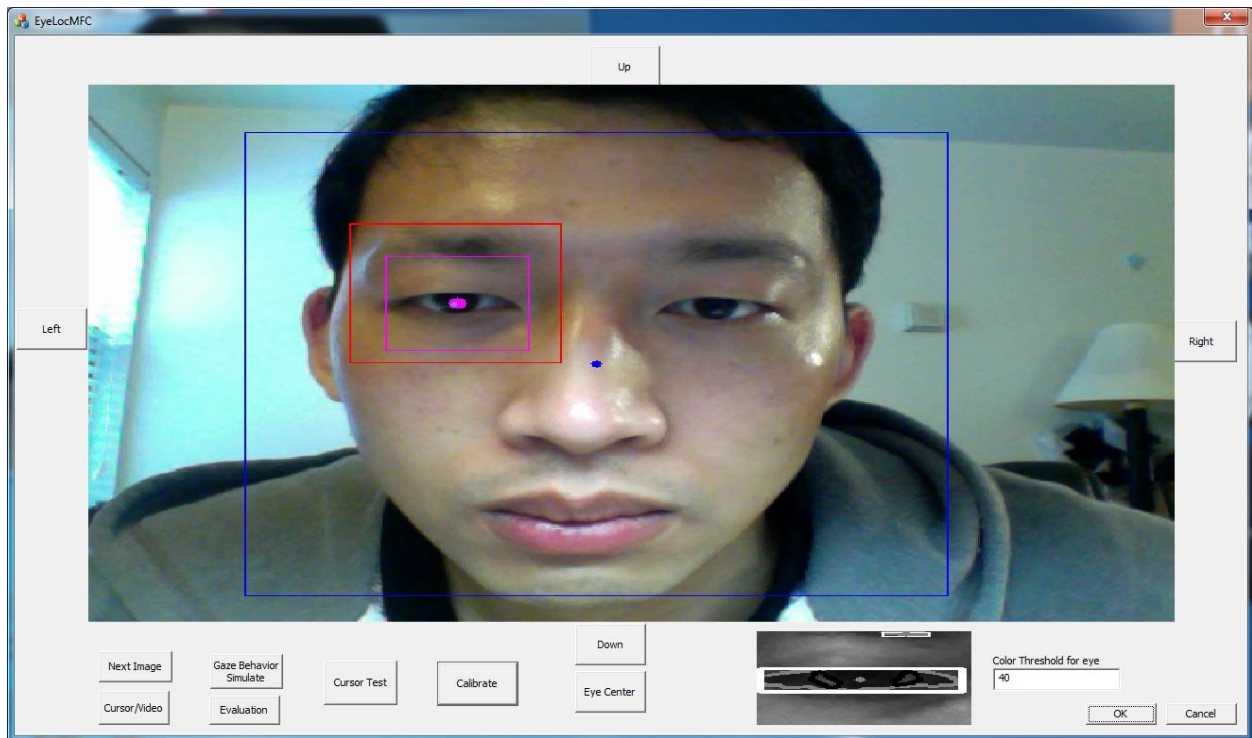


Figure 4.1 Gaze tracking system user interface

When the eye center is detected correctly as well as the user keeps still and ready for test, he or she can start calibration test by pressing “Calibrate”. Under the instruction, the user gazes at one button (e.g. ‘Up’) and then presses the button. The system begins to collect gaze point for ‘Up’. When it finishes collection, the button will show as ‘Done’. Then the user can switch to another button for calibration.

After finishing the calibration of ‘Up’, ‘Down’, ‘Left’ and ‘Right’, the user can begin gaze tracking test by pressing ‘Cursor Test’. In Figure 4.2, the interface is divided into unit areas. Each area has its own unique index. Some of the unit area is covered by objects, which are represented with different colors (e.g., blue and green). The gaze point real-time moves with eye movement. When the gaze point locates in one object, it will show the index of the object along with the gaze point. The coordinates of the gaze

point as well as its corresponding object index is stored in file with time stamp. We can analyze user gaze behavior based on the stored gaze point data.

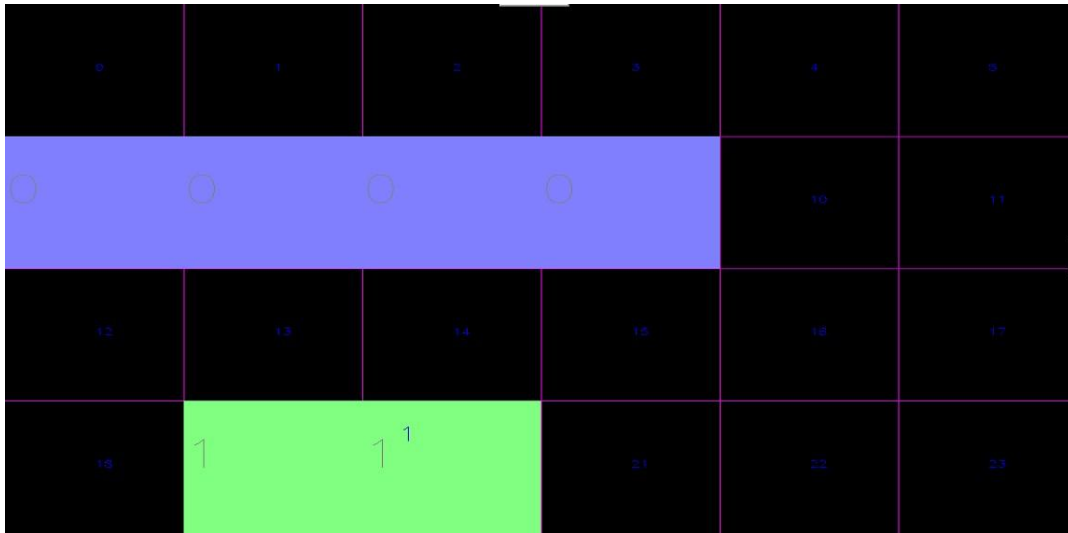


Figure 4.2 Gaze object test

4.2 Gaze Tracking Accuracy Evaluation

The gaze tracking evaluation process is to examine the difference between desired gaze point and obtained gaze point at various screen region and with different user.

The evaluation is started by press “Evaluation” button in Figure 4.1 after calibration. We use six defined points as reference points and measure error between gaze points and reference points as shown in Figure 4.3. Initially, the test interface is empty. When the evaluation test begins, one reference point appears on the interface. Counting down number shows beside the reference point, so as to instruct user to gaze at the reference point. When it counts down to zero, the system begins to collect user’s gaze points. These points will be shown on interface. After collecting ten gaze points, the system finishes the collection on this reference point, and shows the next reference point with counting down number in order to instruct the user to gaze at that point. All reference points appear one by one. The corresponding gaze points are stored in the system.

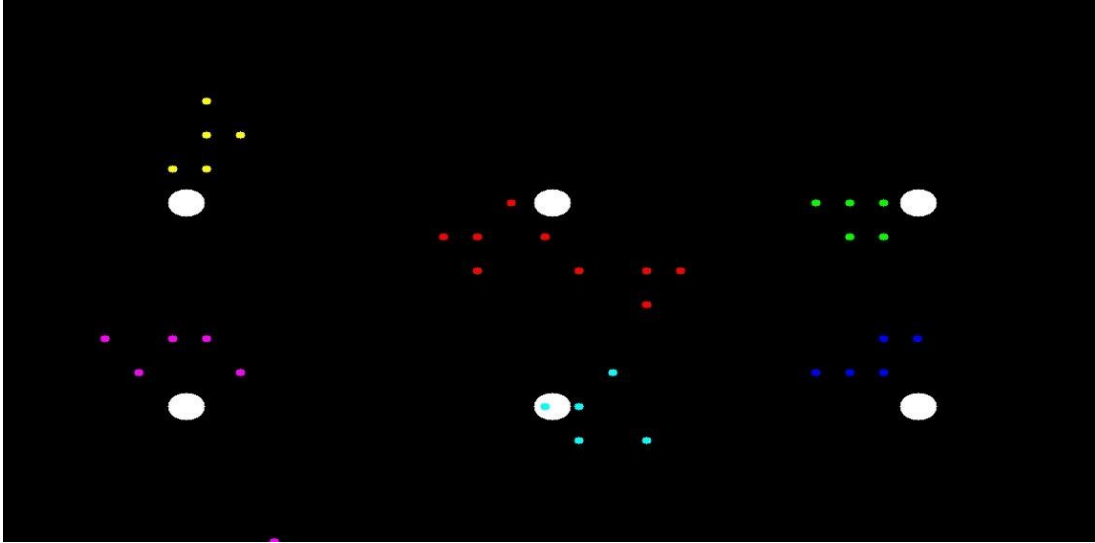


Figure 4.3 Gaze tracking accuracy evaluation

In the experiment, the rectangular test interface on computer screen is 28.8cm wide and 18cm high. The interval of neighbor gaze point on screen is 0.7 cm in width and 0.7 cm in height. The user sits 40 cm away from the computer screen. The camera locates at the top of the screen and within same plan. We assume that user's sight line is perpendicular to the screen. Initially his/her gaze point locates at the center of the screen. The position error at each reference point is computed as follow

$$Position\ error = \sqrt{(x_{gaze} - x_{ref})^2 + (y_{gaze} - y_{ref})^2} \quad (4.1)$$

where, (x_{gaze}, y_{gaze}) is the position of the gaze point on computer screen, (x_{ref}, y_{ref}) is the position of reference point on computer screen. Table 4.1 shows the position error at each reference point.

Table 4.1 Position error at each reference point

Trial	Position error (cm)					
	Ref 1	Ref 2	Ref 3	Ref 4	Ref 5	Ref 6
1	0.70	0.70	2.10	0.99	0.00	2.21
2	0.70	0.70	1.40	1.98	0.70	1.57
3	1.40	1.57	0.70	1.40	0.99	0.99
4	2.10	2.21	1.57	1.40	0.99	1.57
5	1.57	1.98	0.99	0.99	1.57	1.40
6	0.70	1.57	2.10	0.99	0.00	2.21
7	0.70	2.52	1.40	1.98	0.70	1.57
8	1.40	2.97	0.70	1.40	0.99	0.99
9	2.10	3.13	1.57	1.40	0.99	1.57
10	1.57	0.70	0.99	0.99	1.57	1.40

The gaze tracking system is to map eye center to gaze point on screen, which is a fan-shape system with retina as rotation center and sight line as axis. The accuracy of system can be evaluated with angular error [42], as follow:

$$Angular\ error = \tan^{-1} \frac{Position\ error}{Distance\ from\ Screen} \quad (4.2)$$

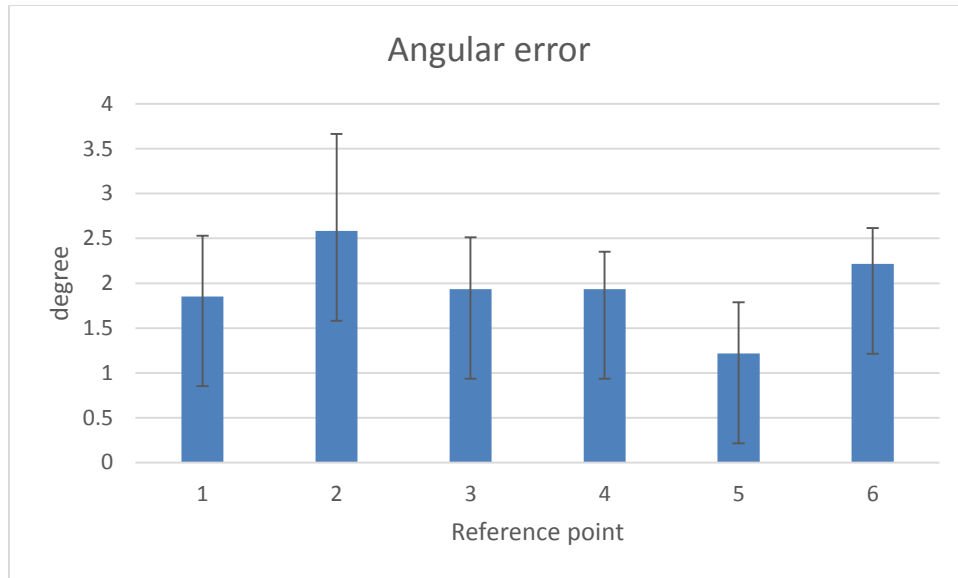


Figure 4.4 Gaze evaluation result

The angular accuracy averages from 1.2 degree to 2.5 degree and deviates 1.08 degree maximum as shown in Figure 4.4. That is 0.85 cm to 1.80 cm average position error and 0.43 cm maximum deviation. Comparing various tracker, the gaze tracking accuracy differs from 0.5 degree to 4 degree with different gaze tracking method and system setup [8]. The proposed eye tracking system only with normal web camera is comparable to existing eye trackers.

Moreover, the distance of user's eye away from camera is critical to position error. When user is close to camera, it will have high resolution of eye movement and result less position error. In the experiment the average angular error is 1.96 degree from Figure 4.4. That means 1cm far away from camera will introduce 0.034 cm more position error. Thus, in order to obtain correct gaze point, the closer the user sits from the camera, the better it will be.

4.3 Gaze behavior representation

User's browsing habit could be observed by tracking the gaze points on single frame or within frames. Their specific intention could be concluded after obtaining gaze order pattern. Gaze duration is one of parameters of gaze pattern. After traversing all the gaze point, the occurrence of gazed object is obtained by accumulating all gaze point.

To demonstrate the image information data representation method on gaze behavior analyze, we simulate gaze tracking result in individual frame and with fine gaze point resolution. In this study, there are 13 objects in the frame as shown in Figure 4.5. O_1 and O_2 are two men in the frame. O_3 is a closet. The rest of objects are shoes on right. As these shoes are within same category, they consist one scene as S_1 . The unit area index of these objects are pre-defined and stored in file.

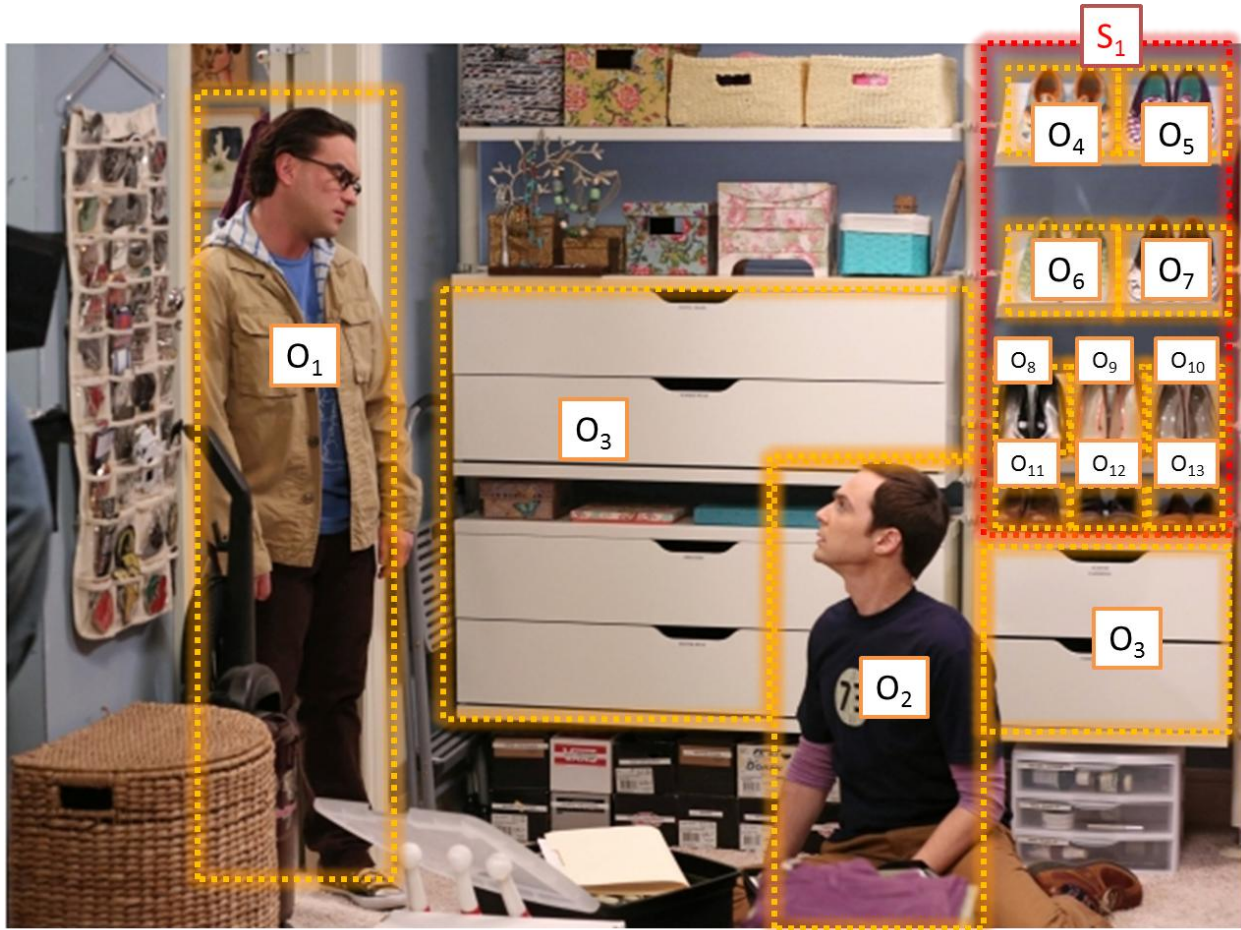


Figure 4.5 Gaze behavior test Frame 1

Each gaze sampling time, the system captures one gaze point. The resolution of Frame 1 is 1260×900 . The unit area resolution is 63×36 as shown in Figure 4.6. On the $28.8\text{cm} \times 18\text{cm}$ screen with resolution of $1440 \text{ pixel} \times 900 \text{ pixel}$, the size of unit area is $1.26\text{cm} \times 0.72\text{cm}$, that is up to the accuracy limit of the proposed gaze tracking system. The gaze point and its corresponding object is shown on interface with the help of data structure shown in Figure 3.5. It is able to track the gaze duration by object and by gaze sampling time, as shown in Figure 4.7. That could summarize into the total gaze duration by object in Figure 4.8. We can conclude that the user has more interest on O_1 , O_2 and O_3 , due to high gaze duration.



Figure 4.6 Unit area and gaze point representation for Frame 1

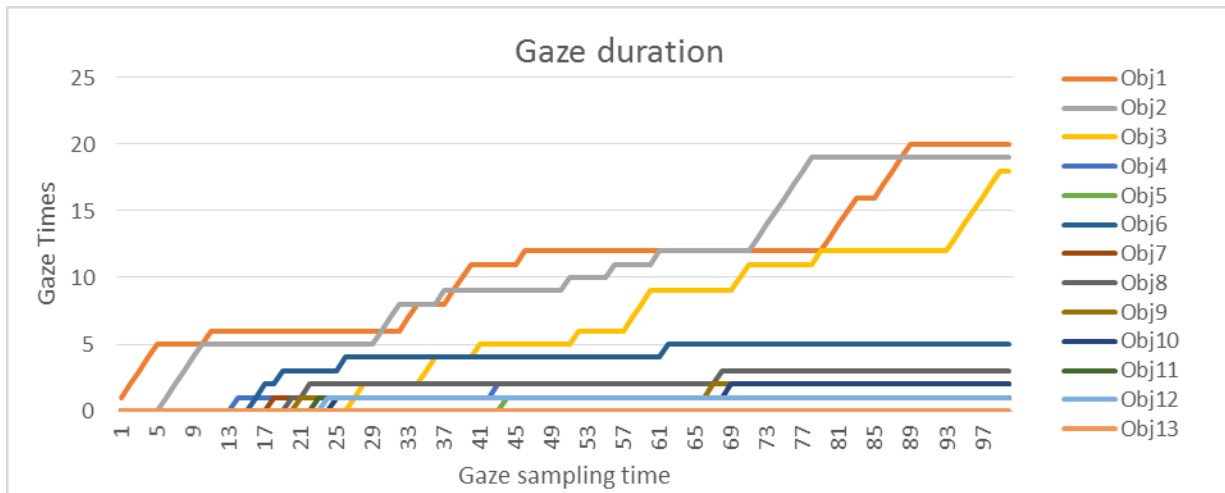


Figure 4.7 Gaze duration in Frame 1

Image data incorporate rules to represent hierarchy structure and contextual structure. These rules are defined as representation rule. It supports the distinction of object which is composed with combination of unit area information in a uniform way, and spatial scene which is consist of objects. Furthermore, it shows the correlation between image and temporal context.

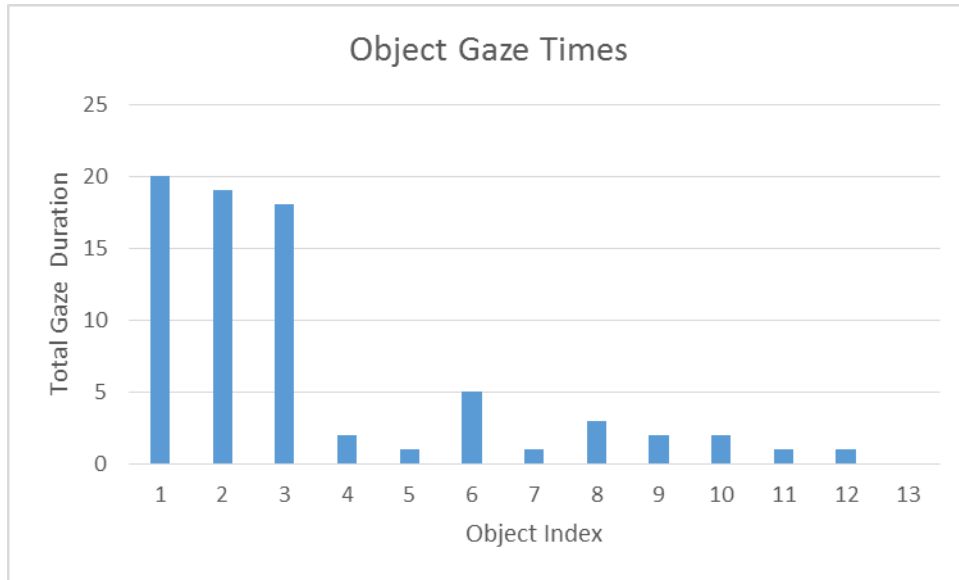


Figure 4.8 Gaze duration by object in Frame 1

The size of unit area is flexible to object size. When the object is big and the interval among objects is distinctive as shown in Figure 4.9, it is possible to use large unit area as basic image description, in order to reduce computation consumption. In Figure 4.10, the image resolution of Frame 2 is 1260×900 , and the unit area resolution is 180×180 larger than that in Frame 1. The number of unit area in Frame 2 is 35 much less than 500 unit areas in Frame 1. With large unit area, it is able to result the gaze duration ratio among objects and reveal user's gaze behavior in Figure 4.11.



Figure 4.9 Gaze behavior test Frame 2

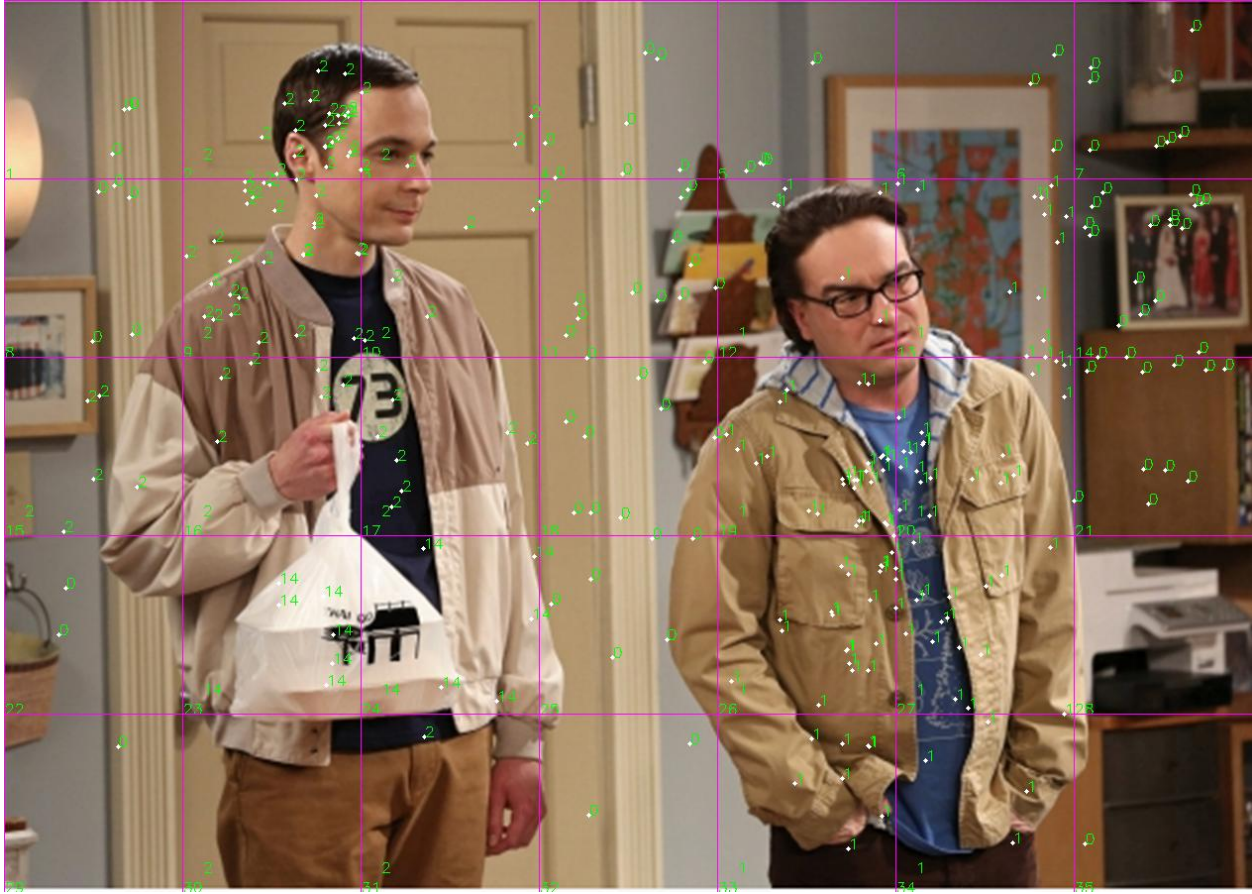


Figure 4.10 Unit area and gaze point representation for Frame 2

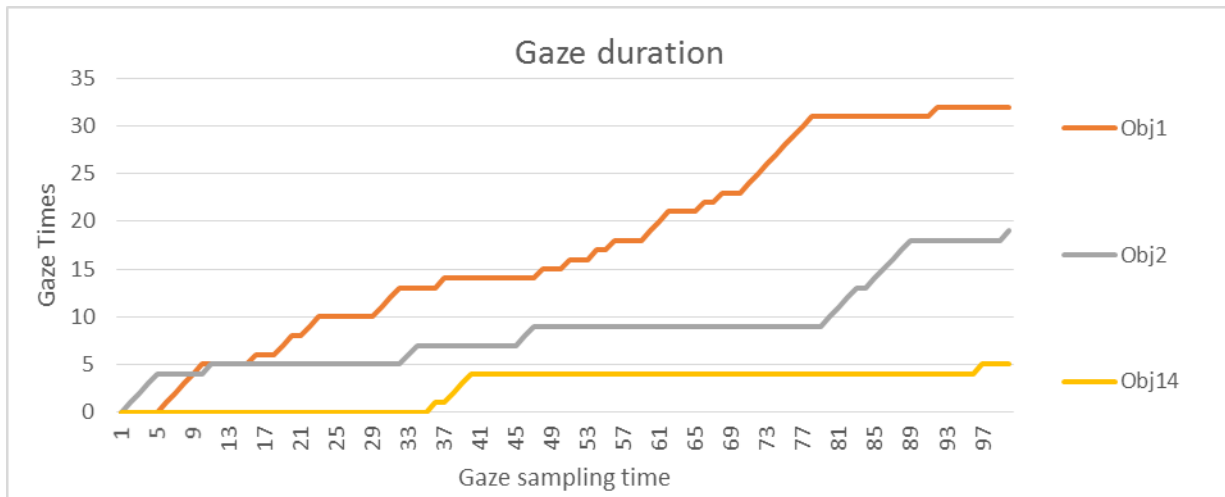


Figure 4.11 Gaze duration in Frame 2

4.4 Gaze behavior with gaze point error

Numerous factors result gaze point error, such as image processing error, gaze point mapping error and user head movement. With these errors, we could not analyze user's gaze behavior correctly. In the simulation, we use the pre-defined gaze point in Figure 4.6 as non-error reference gaze point. Based on these points, we simulate gaze point error to show the effect due to these error. 10-pixel error points, 20-pixel error points, 50-pixel error points and 100-pixel error points are calculated as follow

$$x_{Error} = x_{No\ Error} + 2n \times rand() - n \quad (4.3)$$

$$y_{Error} = y_{No\ Error} + 2n \times rand() - n \quad (4.4)$$

$$if\ x_{Error} > x_{max}, x_{Error} = x_{max}. if\ x_{Error} < x_{min}, x_{Error} = x_{min}.$$

$$if\ y_{Error} > y_{max}, y_{Error} = y_{max}. if\ y_{Error} < y_{min}, y_{Error} = y_{min}.$$

where, error points coordinate is (x_{Error}, y_{Error}) , non-error gaze point is $(x_{No\ Error}, y_{No\ Error})$, n is 10, 20, 50 and 100 for each error simulation, $rand()$ is a random function returning random number in range of $[0, 1]$, (x_{min}, y_{min}) and (x_{max}, y_{max}) are the size of the frame image.

Figure 4.12 and Figure 4.13 shows the gaze duration of object 1 and object 8 with no gaze point error, 10-pixel error, 20-pixel error, 50-pixel error and 100-pixel error. For object 1, the gaze duration decrease with increasing gaze error. However since the gaze duration of object 1 is long relative to duration loss due to gaze error, the gaze duration increment trend is still consistent with different gaze point error. On the other hand for object 8, as the gaze duration is short when no error occurs, the gaze point error introduces large difference for the duration result. That makes it invalid to use gaze duration to analyze gaze behavior when there is gaze error.

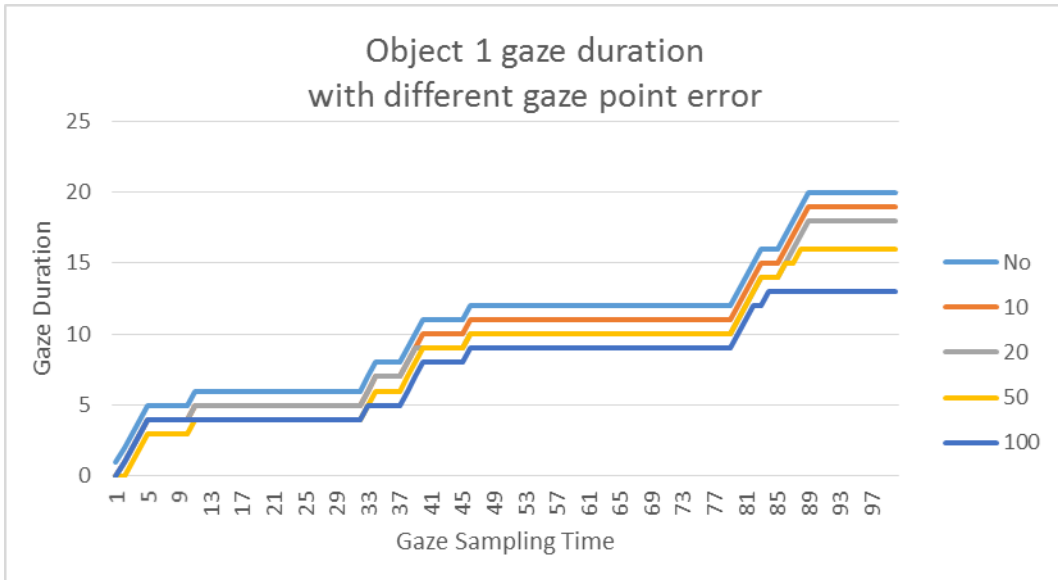


Figure 4.12 Gaze duration of object 1 with different gaze point error

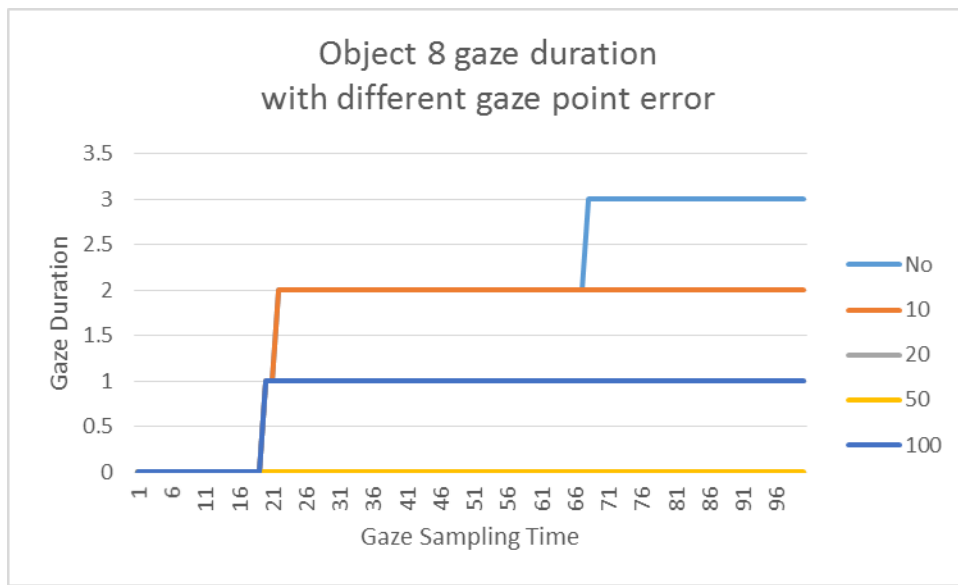


Figure 4.13 Gaze duration of object 8 with different gaze point error

Table 4.2 shows the difference of gaze duration with different gaze error from the duration with no gaze error. With gaze error increase, there is more duration difference from no error case. The object with more gaze duration in no error has more duration difference in gaze point error cases.

Table 4.2 Gaze duration with different gaze error relative to no gaze error

object	Gaze dura	Gaze duration relative to no error			
	no error	10 pixel	20 pixel	50 pixel	100 pixel
1	20	0	-1	-3	-6
2	19	0	1	-6	-4
3	18	0	0	3	3
4	2	0	0	-1	1
5	1	1	0	1	-1
6	5	0	0	0	2
7	1	0	0	2	-1
8	3	-1	-2	-3	-2
9	2	1	3	1	3
10	2	0	-1	0	-2
11	1	0	-1	1	-1
12	1	0	0	1	0
13	0	0	0	0	0

The relative duration is obtained as follow,

$$Relative\ duration\ difference\ (\%) = \frac{Dur_i - Dur_{no}}{Dur_{no}} \times 100\% \quad (4.5)$$

where, Dur_i is the gaze duration at i -pixel gaze error, Dur_{no} is the gaze duration with no gaze error. Figure 4.14 shows the average and deviation of relative duration difference of all objects with each gaze point difference. High gaze point difference results increasing deviation of relative duration. When gaze point error achieves 50-pixel and 100-pixel, the deviation is too large to analyze gaze behavior correctly. Thus, the gaze point detection accuracy is critical to gaze behavior analysis. The more accurate it is, the better analysis will result.

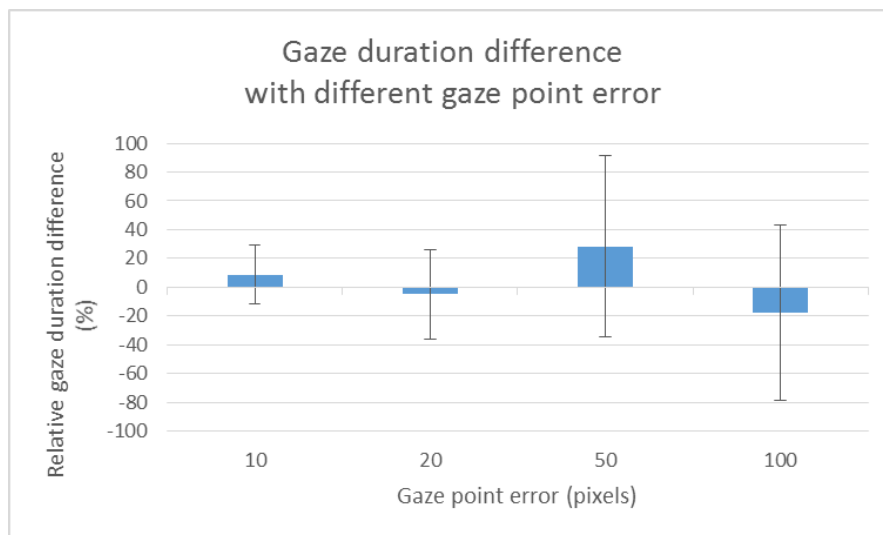


Figure 4.14 Gaze duration difference relative to no error case (%)

5. Conclusion and Future Work

The proposed gaze tracking system is able to real-time detect and display human gaze point on computer screen with normal web camera. That makes it possible to capture gaze data when user is browsing internet or reading material on computer. In the calibration stage, the system maps eye center position to gaze point on screen. In the test stage, gaze point as well as the gaze object are recorded for gaze behavior analysis. The average angular error is 1.96 degree, which is comparable to other proposed techniques. The system represents image information in unit area, object, scene, and frame hierarchy structure. In the gaze behavior simulation, the system is able to summarize gaze duration ratio among each objects and change unit area size with object size to reduce computation. Moreover, it is critical to improve gaze point detection accuracy for gaze behavior understanding.

For future work, there can be several attempts to improve gaze tracking. In the image capturing aspect, infrared camera can be used with visual camera as they are integrated in some smart phones and electrical game players. With fixed setup among infrared emitter, infrared camera and high resolution visual camera, the eye movement can captured more precisely. In the image processing, gray scale histogram method is viable to detect eye center in the condition of various light environment and low image resolution. More image algorithm can be introduced with high quality image, such as detect features on iris and locating eye ball with morphology. In calibration, in addition to the four points “Up”, “Down”, “Left” and “Right”, the corner points “Up left”, “Up right”, “Low left” and “Low right” can be used. Based on the multiple points, further calibration can be implemented. Since the screen is rectangular while the eye movement in eye frame is in ellipse, multiple point calibration can lead gaze mapping more correct. In eye detection processing, to overcome limitation of keeping user head still in the experiment, the position of head and relative position of two eyes and noses can be used to make eye center mapping more robust. As for gaze behavior analyze, changing unit area adaptively based on object size will result multiple frame analysis efficient. Further parameter, aside from gaze object ratio, can be introduced such as gaze object sequence and relationship among objects in multiple frames, in order to understand user’s behavior and attention on web page or reading material.

Reference

1. Klingenhoefer, S. and F. Bremmer, *Saccadic suppression of displacement in face of saccade adaptation*. Vision Research, 2011. **51**(8): p. 881-889.
2. Rayner, K., *Eye movements in reading and information processing: 20 years of research*. Psychological Bulletin, 1998. **124**(3): p. 372-422.
3. *Cross section drawing of the eye - (side view) with major parts labeled*. 2005; Available from: <http://www.99main.com/~charlief/Blindness.htm>.
4. Lopez, J.S.A., *Off-the-Shelf Gaze Interaction*. 2009.
5. *Electro nystagmography*. Available from: <http://www.metrovision.fr/>.
6. Savas, Z., *Real-time detection and tracking of human eyes in video sequences*, 2005, MIDDLE EAST TECHNICAL UNIVERSITY.
7. Young, L.R. and D. Sheena, *Survey of eye movement recording methods*. Behavior Research Methods & Instrumentation, 1975. **7**(5): p. 397-429.
8. Hansen, D.W. and J. Qiang, *In the Eye of the Beholder: A Survey of Models for Eyes and Gaze*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(3): p. 478-500.
9. *The system I4Control*. Available from: http://cyber.felk.cvut.cz/i4c/en_system.html.
10. *Mobile eye tracking - Tobii Glasses*.
11. *SMI EYE TRACKING GLASSES*. Available from: <http://eyetracking-glasses.com/>.
12. Li, D., D. Winfield, and D.J. Parkhurst. *Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches*. in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. 2005. IEEE.
13. *How to build low cost eye tracking glasses for head mounted system*. TUESDAY, AUGUST 17, 2010.
14. *Tobii X2-60 Eye Tracker*. Available from: <http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-x2-60-eye-tracker/>.
15. *RED-M World's smallest high performance eye tracking lab*. Available from: <http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/redm.html>.
16. *Development of Gaze Tracking Technology*. Available from: <http://www.gazegroup.org/research#gazeinteraction>.
17. Morimoto, C.H., et al. *Keeping an eye for HCI*. in *Computer Graphics and Image Processing, 1999. Proceedings. XII Brazilian Symposium on*. 1999. IEEE.
18. Yarbus, A., *Eye movements during perception of complex objects*. L. A. Riggs, ed., 'Eye Movements and Vision', Plenum Press, New York, , 1967. **chapter VII**, : p. 171-196.
19. Doll, T.J., S.W. McWhorter, and D.E. Schmieder. *Simulation of human visual search in cluttered backgrounds*. in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 1993. SAGE Publications.
20. Anders, G., *Pilot's attention allocation during approach and landing- Eye-and head-tracking research in an A 330 full flight simulator*. Focusing Attention on Aviation Safety, 2001.
21. Ottati, W.L., J.C. Hickox, and J. Richter. *Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation*. in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 1999. SAGE Publications.

22. Tanriverdi, V. and R.J. Jacob. *Interacting with eye movements in virtual environments*. in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2000. ACM.
23. *Samsung's Galaxy S4 rumored to use eye-tracking software*. 2013; Available from: http://news.cnet.com/8301-1035_3-57572402-94/samsungs-galaxy-s4-rumored-to-use-eye-tracking-software/.
24. *CNET First Look at The Eye Tribe at CES 2013*. 2013; Available from: http://www.youtube.com/watch?feature=player_embedded&v=SyEqMCwJWkw.
25. Loschky, L.C. and G.W. McConkie. *User performance with gaze contingent multiresolutional displays*. in *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000. ACM.
26. *Eye-tracking tool spells new opportunity for disabled*. Nov 26, 2010.
27. Lohse, G.L., *Consumer eye movement patterns on yellow pages advertising*. Journal of Advertising, 1997: p. 61-73.
28. Wedel, M. and R. Pieters, *Eye fixations on advertisements and memory for brands: A model and findings*. Marketing science, 2000. **19**(4): p. 297-312.
29. Salojärvi, J., et al. *Can relevance be inferred from eye movements in information retrieval*. in *Proceedings of WSOM*. 2003.
30. Nielsen, J., *F-shaped pattern for reading web content*. Retrieved January, 2006. **18**: p. 2007.
31. Megaw, E. and J. Richardson, *Eye movements and industrial inspection*. Applied Ergonomics, 1979. **10**(3): p. 145-154.
32. Goldberg, J.H. and X.P. Kotval, *Computer interface evaluation using eye movements: methods and constructs*. International Journal of Industrial Ergonomics, 1999. **24**(6): p. 631-645.
33. Pan, B., et al. *The determinants of web page viewing behavior: an eye-tracking study*. in *Proceedings of the 2004 symposium on Eye tracking research & applications*. 2004. ACM.
34. Salvucci, D.D. *Inferring intent in eye-based interfaces: tracing eye movements with process models*. in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1999. ACM.
35. Buscher, G., E. Cutrell, and M.R. Morris, *What do you see when you're surfing?: using eye tracking to predict salient regions of web pages*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2009*, ACM: Boston, MA, USA. p. 21-30.
36. Aula, A., P. Majaranta, and K.-J. Rähkä, *Eye-tracking reveals the personal styles for search result evaluation*, in *Human-Computer Interaction-INTERACT 2005*. 2005, Springer. p. 1058-1061.
37. González-Caro, C. and M.-C. Marcos, *Different Users and Intents: An Eye-tracking Analysis of Web Search*. Proc. WSDM, 2011: p. 1-8.
38. Joachims, T., et al., *Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search*. ACM Transactions on Information Systems (TOIS), 2007. **25**(2): p. 7.
39. Huang, J., R. White, and G. Buscher, *User see, user point: gaze and cursor alignment in web search*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2012*, ACM: Austin, Texas, USA. p. 1341-1350.

40. Lienhart, R. and J. Maydt. *An extended set of haar-like features for rapid object detection*. in *Image Processing. 2002. Proceedings. 2002 International Conference on*. 2002. IEEE.
41. Schmidt, A. and A. Kasiński, *The performance of the haar cascade classifiers applied to the face and eyes detection*, in *Computer Recognition Systems 2*. 2007, Springer. p. 816-823.
42. Kar-Han, T., D. Kriegman, and N. Ahuja. *Appearance-based eye gaze estimation*. in *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*. 2002.