2011

GAZE ESTIMATION USING SCLERA AND IRIS EXTRACTION

Prashanth Rao Periketi

University of Kentucky, periketi.prashanth@gmail.com

Recommended Citation

http://uknowledge.uky.edu/gradschool_theses/145
ABSTRACT OF THESIS

GAZE ESTIMATION USING SCLERA AND IRIS EXTRACTION

Tracking gaze of an individual provides important information in understanding the behavior of that person. Gaze tracking has been widely used in a variety of applications from tracking consumers gaze fixation on advertisements, controlling human-computer devices, to understanding behaviors of patients with various types of visual and/or neurological disorders such as autism. Gaze pattern can be identified using different methods but most of them require the use of specialized equipments which can be prohibitively expensive for some applications. In this dissertation, we investigate the possibility of using sclera and iris regions captured in a webcam sequence to estimate gaze pattern. The sclera and iris regions in the video frame are first extracted by using an adaptive thresholding technique. The gaze pattern is then determined based on areas of different sclera and iris regions and distances between tracked points along the irises. The technique is novel as sclera regions are often ignored in eye tracking literature while we have demonstrated that they can be easily extracted from images captured by low-cost camera and are useful in determining the gaze pattern. The accuracy and computational efficiency of the proposed technique is demonstrated by experiments with human subjects.

Key words: gaze, sclera, threshold, optical flow and lucas-kanade algorithm.

Multimedia elements used: JPEG (.jpg).

Prashanth Rao Periketi.
08/01/2011
GAZE ESTIMATION USING SCLERA AND IRIS EXTRACTION

By

Prashanth Rao Periketi

Dr. Sen Ching Samson Cheung.
Associate Professor

Dr. Zhi David Chen.
Director General of Graduate Studies
RULES FOR THE USE OF THESIS

Unpublished theses submitted for the Master’s degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this thesis for use by its patrons is expected to secure the signature of each user.

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THESIS

Prashanth Rao Periketi

The Graduate School
University of Kentucky
2011
GAZE ESTIMATION USING SCLERA AND IRIS EXTRACTION

THESIS

A thesis submitted in partial fulfillment of the requirements for the degree of the Master of Science in Electrical Engineering in the College of Engineering at the University of Kentucky

By

Prashanth Rao Periketi

Lexington, Kentucky

Director: Dr. Sen Ching Samson Cheung, Professor of Electrical and Computer Engineering, Lexington, Kentucky

2011

Copyright © Prashanth Rao Periketi 2011
# TABLE OF CONTENTS

Table of Contents ........................................................................................................iv  
List of Figures ...............................................................................................................vi  

## Chapter One: Introduction

1.1. Background ........................................................................................................1  
1.2. Problems with existing techniques .................................................................1  
1.3. Proposed Methods ..........................................................................................3  
1.4. Outline of Thesis ............................................................................................3  

## Chapter Two: Related Work

2.1. Introduction ......................................................................................................5  
2.2. Literature Review ..........................................................................................6  

## Chapter Three: Feature Extraction

3.1. Calibration ........................................................................................................9  
3.2. Selection and tracking of points ...................................................................9  
3.3. Thresholding in different color spaces ..........................................................10  
3.3.1. RGB .........................................................................................................10  
3.3.2. HSV .......................................................................................................11  
3.3.3. Lab .........................................................................................................12  
3.3.4. YUV .......................................................................................................12  
3.3.5. YCbCr .....................................................................................................13  
3.3.6. NTSC .....................................................................................................13  
3.4. Adaptive thresholding technique ................................................................14  

## Chapter Four: Gaze Estimation

4.1. Using sclera area and distance between tracking Points along iris axis .... 18  
4.2. Using sclera and iris areas ...........................................................................20  

## Chapter Five: Experimental Results

5.1. Procedure ........................................................................................................22
5.2. Results

5.2.1. Advantages and Disadvantages of methods

Chapter Six: Conclusions

6.1. Contributions and Limitations

References

Vita
LIST OF FIGURES

Fig1: Basic design model.................................................................2
Fig2: Embedding a Rectangular Mask on eye........................................2
Fig3: Calibration and tracking of Points................................................9
Fig4: Test Image for performing thresholding in different color spaces........10
Fig5: Sclera and other pixel values in RGB color space along different channels.....11
Fig6: Sclera and other pixel values in HSV color space along different channels.....11
Fig7: Sclera and other pixel values in Lab color space along different channels........12
Fig8: Sclera and other pixel values in YUV color space along different channels........12
Fig9: Sclera and other pixel values in YCbCr color space along different channels......13
Fig10: Sclera and other pixel values in NTSC color space along different channels.....13
Fig11: Histogram of a gray-level image with dark objects on light background........14
Fig12: Histogram of gray-level distribution of an image................................15
Fig13: Sclera Areas of each eye................................................................18
Fig14: Variation of sclera areas with respect to different points on the screen.......19
Fig15: Variation of distance between tracking points along iris axis with respect to different points on screen.........................................................20
Fig16: Sclera and iris areas....................................................................20
Fig17: Variation of iris areas with respect to different points on the screen.......21
Fig18: Experimental set-up........................................................................22
Fig19: Experimental results for gaze estimation..........................................25
Fig20: Results of sclera and iris extraction using adaptive thresholding..........26
Fig21: Results using global and adaptive thresholding....................................27
1. INTRODUCTION

1.1. Background:

Gaze estimation is the process of measuring either the point of gaze or the motion of an eye relative to the head. The device to measure gaze is called eye tracker [1]. Tracking eye gaze of an individual has numerous applications in many domains. Gaze tracking is a powerful tool for the study of real time cognitive processing and information transfer-as human attention can be deduced in many cases by following gaze to the object of interest. Latest human-computer interacting devices use person’s gaze as input to the computer in the same way as using a mouse. Gaze estimation has also found usage in the auto industry for monitoring driver vigilance, based on driver’s gaze patterns [2]. Eye gaze also has applications in the medical field, for example studying the behaviors of patients suffering from neurological, communication and vision disorders like pervasive developmental disorder and autism spectrum disorder. Studies show that during a typical interaction with another person, rather than focusing on the eyes, children with ASD (Autism Spectrum Disorder) tend to focus more on mouth, bodies and other objects at the scene [3-5]. Thus the gaze pattern could serve as a useful tool for early detection of ASD.

1.2. Problems with existing Technique:

Most of the accurate eye tracking devices available today are expensive because they require high quality cameras and specially designed equipments like wearing custom designed glasses, high speed machines etc. Some of the devices require trained individuals in collecting the raw data and for conducting experiments. The techniques they use mostly concentrate on the reflection of the IR light on the pupil and track its position using mathematical analysis. In our thesis, we tried to show that the iris and sclera can also be used for gaze estimation and they can be detected easily. The objective of this thesis is to devise a simple web-cam based eye tracking algorithm that can provide reasonably good accuracy. The basic design of the proposed algorithm in shown in fig1, 2:
Select calibration points

Select Tracking points

Points are tracked using LK Algorithm

Embed a rectangular mask on the eye using calibrated and tracking points

Apply adaptive thresholding technique for extracting features.

Using Sclera and Iris Areas

Gaze Estimation

Using Sclera Area and distance between tracking points along iris axis

Fig1: Basic design model

Fig2: Embedding a rectangular mask on an eye (Blue points indicate calibration points and green points indicate tracking points).
1.3. Proposed Methods:-

First by selecting one point on each eye (basically a point on pupil of each eye) from the camera image, these points are tracked for the entire duration of video sequence. We use Lucas – Kanade optical flow estimation algorithm [22] for tracking these points. Lucas-Kanade algorithm is a two-frame differential algorithm in which the optical flow is assumed to be constant in a small neighborhood around the center point under consideration at any given point of time. The principal advantage of Lucas-Kanade algorithm over other tracking algorithms is its robustness to noise. Furthermore it does not yield a very high density of motion vectors unlike other algorithms.

Assuming that the head is stationary, gaze pattern is estimated by extracting sclera and iris regions from the tracked eyes. There are a number of technical difficulties in extracting the sclera region including varying illumination, a large variety of shapes and sizes as well as blinking of eyes. In this dissertation, the sclera and iris of an eye are extracted by using a novel adaptive thresholding technique. In the proposed adaptive thresholding technique, we use a small spatial mask covering the entire eye. We then calculate the mean of all the pixels in the image/video in the neighborhood of the center pixel encompassed by the mask and compare it with the actual intensity of the center pixel. If the center pixel or desired pixel intensity is greater than the mean then that pixel is determined to be a pixel in sclera region and the intensity of that pixel is changed to white (say equal to 255), otherwise it is considered as an iris pixel and intensity is made equal to zero (i.e. to black). The sclera region can also be extracted by using thresholding in different color space, by using color-based Bayes decision thresholds [10] and by using a modified time adaptive self-organizing map (TASOM) based active contour models (ACMs) [11]. However the results obtained using our adaptive thresholding technique are superior to other methods since it is less affected by illumination and also it is computationally efficient.

1.4. Outline of Thesis:-

Finally, we give a brief tour of the material in this dissertation. The dissertation starts in Chapter 2 with explanation of gaze estimation by providing literature review of related
work in this field. Chapter 3 covers Feature Extraction, which is an important part of the dissertation. This includes topics such as calibration (3.1), selection and tracking of points (3.2), adaptive thresholding technique (3.3), thresholding in different color spaces (3.4) and why adaptive thresholding is used for feature extraction (3.5). In Chapter 4, we cover gaze estimation techniques using (i) sclera area with distance between tracking points along iris axis (4.1) and (ii) sclera and iris areas (4.2). In Chapter 5 we discuss about experimental procedure (5.1) and results (5.2). Chapter 6 covers the conclusion, explaining the things which we have done better using our technique and also mentions the steps for improving the accuracy and efficiency of the system.
2. RELATED WORK

2.1. Introduction:

Even though significant research has been done on estimating the gaze pattern, eye tracking remains a challenging task because of the individuality of eyes, occlusion and light conditions [2]. The ideal eye tracking device should satisfy the following usability requirements [8, 9]:

a) Offer an unobstructed field of view with good access to the face and head.

b) Make no contact with the subject.

c) Meet the practical challenge of being capable of artificially stabilizing the retinal image if necessary.

d) Possess an accuracy of at least one percent or few minutes of arc.

e) Offer a resolution of 1 arc.sec\(^{-1}\), and thus be capable of detecting the smallest changes in eye position; resolution is limited only by instrumental noise.

f) Possess a real time response to allow physiological maneuvers.

g) Measure all three degrees of angular rotation and be insensitive to ocular translation.

h) Be easily extended to binocular recording.

i) Be compatible with head and body recordings.

j) Be easy to use on a variety of subjects.

All the above requirements are desirable but not all of them are prerequisites for acceptable eye tracking interfaces. Requirement (j) would be nice, as it would enable a greater freedom of movement for the user, allow one eye to be closed and generally yield more reliable tracking data, it is not essential for the average usage. The current
techniques of today can be classified into three types based on the way they make contact with the subject. They are:

(i) Measuring the reflection of light that is shone onto the eye.
(ii) Measuring the electric potential of the skin around the eyes.
(iii) Applying a special contact lens that facilitates tracking of its position.

Techniques belonging to type 1 make least contact with the user while type 3 includes techniques that make the most contact. However, all the techniques require some form of calibration before usage and also frequent recalibration during use.

2.2. Literature Review:

Modern gaze estimation techniques consist of the following main categories:

i. **Electro-Oculography**: Method in which we measure skin’s electric potential differences of electrodes placed around the eye.

ii. **Scleral Contact Lens/Search coil**: Method involves attaching a mechanical or optical reference object mounted on contact lens which is then worn directly on the eye and employs a wire coil embedded in a Scleral contact lens which is then measured moving through an electro-magnetic field.

iii. **Photo-Oculography (POG) or Video-Oculography (VOG)**: This category groups together a wide variety of eye movement recording techniques involving the measurement of distinguishable features of the eyes under rotation/translation and corneal reflections of a closely situated directed light source (often infra-red).

iv. **Video-based Combined Pupil/Corneal Reflection**: In this method, corneal reflection of light source (typically infra-red) is measured relative to the location of pupil center and corneal reflections are called as Purkinje reflections or images [6]. The gaze estimation methods also include feature based estimation methods and appearance based methods [2]. Feature based gaze estimation methods include extracting the local features such as contours, eye corners and reflections from eye images. Gaze tracking can be also estimated using appearance based methods as described in [7].
The first three categories of techniques are usually quite intrusive, requiring physical contact with the subject and explicit intervention from the operator. The last category of techniques determines where a user is looking from the appearance of the user’s eye. Due to the ambiguity of appearance of the eyes due to head movement, it is important to also obtain a reliable estimation of the head pose. Head pose estimation is nothing but the ability to infer the orientation of person’s head with respect to the view of the camera. The authors of [12] give a detailed survey report of all the head pose estimation techniques and also the difficulties in head pose estimation. In challenging environments where head pose is difficult to estimate, an alternative solution is to rely on head-mounted devices. Head-mounted devices have at least two advantages over most techniques that track distance, they track the light rays that actually enter the eye and one tracker can cover the entire room. The dual purkinje–image tracking technique can in fact record the user’s accommodation of focus i.e., how far away the user is looking. However, this technique also requires the user’s head to stay still in relation to the measuring equipment and one way of doing this is by using a head–mounting device.

There are a myriad of video-based techniques in analyzing the gaze pattern. In [13], the authors use eye segmentation, pupil ellipse fitting and sclera detection to extract relevant eye features from an image of a user which will be used as training data for two neural networks to estimate gaze’s horizontal and vertical coordinates. In sub pixel gaze tracking [14], the authors propose algorithms for detecting the inner eye corner and center of an iris in sub-pixel accuracy and construct a 2-D linear mapping from the vector between eye corner and iris center to the gaze angle. Gaze directions in successive frames are calculated by interpolation. The gaze of a user can also be determined by using human sclera of eyes as mentioned in [15], [16]. In [15], the sclera of the eyes is detected by using a statistical model based on Bayes decision rule to detect the color of sclera of an eye. However, this system lacks efficiency. In [16], the sclera is detected by using time-adaptive self-organizing map (TASOM)-based active contour models (ACMs) for detecting the boundaries of human eye sclera and tracking its moments in sequence of images. This method is insensitive to global intensity changes and illumination changes but it is sensitive to occlusion. Also some of the gaze estimation techniques use neural
networks to estimate gaze by giving the images of user’s both eyes as input to the neural networks [17].

Gaze can also be estimated by tracking face, estimating the 3-D pose and determining the 3-D gaze vector as mentioned in [18]. Gaze can also be found by using images of irises [19]. Two Iris contours are approximately modeled as two circles having known radii and the ellipses of their projections are estimated. Eye gaze is estimated from the projection of an iris contour. This method requires high resolution images hence it uses a zoom-in camera. Gaze can also be estimated by using a single eye [20]. This method relies on the fact that outer boundary of an iris is a circle. With a fully calibrated camera, its elliptical image can be back projected onto the 3-D space yielding two possible circles and the solution is found by using anthropomorphic knowledge of the structure of eye ball. Gaze can be found by methods based on facial feature tracking [21], where direction of the gaze is determined by using 3D vectors connecting both the eye ball and iris centers. In this paper, the geometrical relation between the eye ball centers and facial features and eye ball radius are calculated in advance, and then 2D positions of the eye ball centers are determined by tracking the facial feature. It does not require any special calibration actions.

Our proposed technique is also an example of video-based gaze tracking schemes. Comparing with contemporary schemes, our proposed system uses non-expensive webcam, requires little calibration effort and can be executed in real-time with moderate computing hardware. Our focus is on developing novel techniques for tracking eyes, extracting sclera and iris features, and deducing the corresponding gaze patterns. Head pose estimation has not been investigated; as such we require the head to be stationary. While we believe our feature extraction and segmentation schemes are quite robust, the simple features used in our gaze estimation algorithm have put a limitation on the efficiency of gaze patterns.
3. FEATURE EXTRACTION

In this dissertation, we developed a model for extracting the sclera and iris of the eyes. The model is described below:– The model is a 3 - step process:

1) Calibration.
2) Selection and Tracking of Points.
3) Adaptive Thresholding Technique.

3.1. Calibration:–

The first step in the model is to calibrate the tracking points on the eye image. On the first frame of the video the user would select 3 points on an eye. The points to be selected are left corner, right corner and bottom of an eye as shown in fig.3:

![Fig3: Calibration and tracking of points (Black points indicate calibration points and green points indicate tracking points).](image)

3.2. Selection and Tracking of Points:–

Once the calibration step is done, next user has to select a point on top of the iris of an eye. These points are tracked for the entire duration of the video using Lucas-Kanade Algorithm. Lucas-Kanade Algorithm is the mostly commonly used sparse optical flow algorithm for tracking selected points. Basically the algorithm is a two-frame differential algorithm which assumes the optical flow to be constant in a small neighborhood along the given point under consideration at any given time. The detailed mathematical description of the algorithm can be found in the [22] [23]. The Lucas-Kanade Algorithm is used because of its robustness to noise (which is a major factor in real time scenarios) and also algorithm yields very few motion vectors. In this dissertation, we used enhanced version of Lucas-Kanade Algorithm which is called as Pyramidal Implementation of Lucas-Kanade Algorithm. The reason is that the original LK Algorithm will not be able
to track the points if there is significant motion. In the enhanced version, the image is divided into layers and points are tracked along each layer [21]. In this dissertation we used 7 pyramidal layers and a window size of about 21x21 to cope with significant pupil movement. A rectangular mask is then embedded on the eye using calibrated and tracking points.

3.3. Thresholding in different color spaces:

Several other techniques can also be used for extracting regions or features such as thresholding in different color spaces, by using flood fill algorithms, by using template matching or by drawing contours. In flood fill algorithm, the user has to select the region of interest and all the neighborhood regions which are similar according to predefined criteria are considered as a single region. The predefined criterion can be color of the pixels or any other statistical property. In Template Matching, the given region of interest is matched with the trained images. For drawing contours we will first extract edges and then connect them according to a predefined criterion. The problem with these kinds of techniques is that the defined criterion is image dependent that is, the criterions has to be changed for every image and also are affected by noise and illumination. For example, as shown in fig4: suppose we need to extract the sclera region which is filled with red color as shown in fig. and will look at the pixel values in different color spaces.

![Fig4. Test Image for performing thresholding in different color spaces.](image)

3.3.1. RGB:-

In the fig. 5, the red colored dots indicates sclera region pixel values in the corresponding channels i.e. for example, in the first figure; red channel is taken along X-Axis and green channel along Y-axis. Hence we can have a threshold value around 200 along Red-Green channel for extracting sclera region i.e. pixels whose value greater than 200 along Red-Green channel are considered to be sclera pixels. However, as we could see from fig.5, we may lose some data if we choose threshold value to be 200. In the following figures,
red dots indicate sclera pixels and green dots indicate other pixels. In the fig.5 x-axis corresponds to the pixel values in the one channel and y-axis correspond to the pixel values in another channel, similarly in the other channels.

![Image](image1.png)

Fig5: Sclera and other pixel values in RGB color space along different channels

3.3.2. HSV:

Similarly, thresholding along 2-3 channels, yields unsatisfactory results as shown in fig.6 where x-axis indicates pixel values in one channel and y-axis indicates pixel values in another channel.

![Image](image2.png)

Fig6: Sclera and other pixel values in HSV color space along different channels.
3.3.3. Lab:

As indicated in fig.7, x-axis indicates pixel values in one channel and y-axis indicates pixel values in another channel. Thresholding can only be performed along 2-3 channels even that do not give desired results.

Fig7: Sclera and other pixel values in Lab color space along different channels

3.3.4. YUV:

As indicated in the fig8 x-axis indicates pixel values in one channel and y-axis pixel values in another channel. We cannot threshold along any channels as none of them give desired results.

Fig8: Sclera and other pixel values in YUV color space along different channels
3.3.5. YCbCr:

Similarly, in fig.9 x-axis indicates pixel values in one channel and y-axis pixel values in another channel. We cannot use this color space for extracting sclera and iris region using thresholding technique as shown in fig.9.

Fig9: Sclera and other pixel values in YCbCr color space along different channels

3.3.6. NTSC:

In fig.10 x-axis indicates pixel values in one channel and y-axis pixel values in another channel. We can threshold along 2-3 channels which may provide partial results as shown in fig.10:

Fig10: Sclera and other pixel values in NTSC color space along different channels
3.4. Adaptive Thresholding Technique:

After embedding a rectangular mask on an eye using calibration and tracking points, we then apply adaptive thresholding technique along the region of interests. Before discussing the proposed adaptive thresholding technique, firstly we need to discuss about the thresholding techniques and why do we need adaptive thresholding technique, and why adaptive thresholding is preferred over other techniques like thresholding in different color spaces, flood filling, template matching and Haar cascade eye detection for extracting sclera and iris regions of an eye.

The basic idea of a thresholding is to scan the original image pixel by pixel and testing each pixel value against a specific threshold value i.e. let \( F(x, y) \) be an original image and \( T \) be any specific threshold, if \( F(x, y) > T \) then that pixel is classified as a background pixel otherwise as an object pixel as shown in fig.11:

\[
b(x, y) = \begin{cases} 
255, & \text{if } F(x, y) > T \\
0, & \text{if } F(x, y) \leq T 
\end{cases}
\]

![Fig11. Histogram of a gray-level image with dark objects on light background](image)

As shown in above figure if we set the threshold value to, say 100, we can then divide the image into two groups. However, we don’t always get an image where we could
threshold image easily into two groups using a single threshold value. For example, consider the histogram of a gray scale image as shown in fig.12:

For the above case more than one threshold is needed to classify the image’s components. It is called multi-level thresholding. However, in real world scenarios, the images are affected by noise and illumination which may affect the results if global thresholding is used. Hence in order to overcome these problems adaptive thresholding technique is used.

In an Adaptive Thresholding Technique, the threshold value is not fixed i.e. the value changes depending on the local property of the image at any given point. The general definition of a threshold is given as

$$T = T [x, y, p(x, y), F(x, y)]$$

In the above equation $F(x, y)$ is the gray-level of a point $(x, y)$ and $p(x, y)$ is some local property of this point. When $T$ depends only on the $F(x, y)$ i.e. the gray level of image then it becomes a simple global threshold. In order to take the effect of noise or illumination into consideration, the calculation of property $p(x, y)$ is important and is
usually based on environment of the point at hand. Therefore, global thresholding uses a fixed threshold for all pixels in the image and therefore works well only if the gray-level distribution histogram contains distinctively separated peaks and background. Hence we used a combination of local and adaptive thresholding to extract sclera and iris regions of eyes.

In this method we will examine the intensity values of local neighborhood of each pixel. We will then compare the intensity of a pixel with the mean (or any statistic choice of measurement like median or mode) of the pixels in the neighborhood of the pixel under consideration at any given point of time. This method faces two problems: Firstly, the choice of statistic which may vary from one image to another image and is largely dependent on the nature of the image. Commonly used statistics for comparison are mean, median, average between the minimal and maximal gray-level in the neighborhood. The second problem is the choice of neighborhood size. The larger the size of neighborhood the poorer is the result because it is more affected by illumination gradient. And also the larger the environment, the more it is expensive to perform the needed computation. However, if the environment is too small then there is a risk of being exposed to insufficient data which may lead to poor results when noise is introduced. For example, consider an image with noise in the form of sporadic points of extreme intensity like “salt and pepper”. In this case, if we choose a large environment then there is a good chance of moderating the influence of this noise at the cost of less resistance to illumination effects. If we choose environment to be small and the statistic being used is average, then pixels in a noisy environment will be influenced by the noise.

In our dissertation, we kept the neighborhood size to be 3x3 and choice of statistic to be mean, which worked well as illustrated in results section.

Hence by applying adaptive thresholding technique combined with local threshold we would determine the pixel to be a sclera pixel and is set to white color if it is greater than the local mean of the neighboring pixels, or an iris pixel set to black color if it is less than the local mean of neighboring pixels.

As mentioned earlier the sclera and iris regions can be extracted by applying thresholding algorithm in any one of the above mentioned color spaces. However, from the above
mentioned results we can deduce that by applying thresholding algorithm along different color spaces will result in non-sclera areas being treated as sclera areas, since we can observe that in all the color spaces the sclera pixel values overlap with other pixel values and also the thresholding algorithm depends on image properties i.e. it is affected by noise and also by illumination. Other techniques such as template matching and contour methods also depend on image properties such as illumination, shape and size of eye, color etc.,

Hence for the above mentioned reasons we are using adaptive thresholding combined with local thresholding for extracting sclera and iris regions.
4. Gaze Estimation

The Gaze Pattern can be estimated by using different parameters and by determining the statistical relationship between those parameters. In this dissertation, we have estimated the gaze by using 2 methods: 1) using the sclera area and distance between tracking point and bottom of an eye. 2) Using the Sclera and Iris areas.

4.1. Using sclera area and distance between tracking points along iris axis:

In this method, eye is divided into two equal halves along the vertical axis at the center of an eye and sclera area is calculated by counting all the white pixels (which are obtained after applying adaptive thresholding technique) on either side of an eye. Let ‘LELScleraArea’ be the sclera area from the center to the left side of left eye, ‘LERSCleraArea’ be the sclera area from the center to the right side of left eye. Similarly ‘RELSceraArea’ be the sclera area from center to the left side of the right eye and ‘RERScleraArea’ be the sclera area from center to the right side of the right eye. The sclera areas of both the eyes as shown in fig.13:

Fig13: Sclera areas of each eye

Fig.14 shows the variation of sclera area of each half of both eyes while looking at 9 different locations on the screen namely top left, top center, top right, center left, center, center right, bottom left, bottom center and bottom right. Fig.15 shows variation of distance between tracking points along iris axis while looking at different locations on screen.
Fig14: Variation of sclera areas with respect to different points on the screen
Fig 15: Variation of distance between tracking points along iris axis with respect to different points on screen

4.2. Using sclera and iris areas:

In this method, eye is divided into two equal halves along vertical and horizontal axis. The sclera area is calculated in the same way as in first method. The iris area is calculated by counting all black pixels above the axis line which connects both the corners of an eye as shown in fig 16:
The Variation of Iris Area with respect to different points on the screen is shown in fig.17:

![Graph showing variation of iris area with respect to different points on the screen.]

Fig17: Variation of iris areas with respect to different points on the screen
5. EXPERIMENTAL RESULTS

5.1. Procedure:

The experimental setup is shown in fig18. below:

Fig18: Experimental set-up.
As shown in fig.18, set-up consists of a webcam placed at the bottom of the screen (of a laptop), a chin rest which restricts the movement of head. Experiment is conducted in a closed room which consists of 6 lights located on the ceiling. Out of the 6 lights, 3 are located closer to the camera and rest of them is located farther. The middle light of first row and the 1st and 3rd light of second row (farther one’s) are turned ON and rest of them are turned OFF. We also used a USB 10 LED laptop light (USB 10-LED Light for Notebook Laptop PC by HDE company) placed at a height of 23cms or 9 inches above the camera. We used a chin rest to keep head static and the distance between two horizontal stands is 18cm or 7 inches. Distance between the camera and the chin rest is 28cms or 11 inches. Height of the camera above the table is 16cms or 6.5 inches. Distance between light and chin rest is 1.2feet or 34cms and monitor size is 14 inches.

First the user is asked to look at the camera and snapshot of the user is taken for calibration. From the calibration image we can get the coordinates of eye corners. Then a test video consisting of 270 frames or 9 sec duration is shown to the user. The test video frame size is 640×480. The test video consists of a randomly moving object and user is asked to follow the object. We will select a point on top of the iris of each eye and these points are tracked for entire video sequence.

5.2. Results:-

Assuming the screen is divided into 9 parts as shown below:

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>P5</td>
<td>P6</td>
</tr>
<tr>
<td>P7</td>
<td>P8</td>
<td>P9</td>
</tr>
</tbody>
</table>

where P1, P2, P3 correspond to top half of the screen, P4, P5, P6 correspond to center and P7, P8, P9 correspond to the bottom of the screen. The experiment was conducted on 9 people.
5.2.1 Advantages and Disadvantages of both methods:

Both methods have their own benefits and in this section we will be discussing them. In both the methods the procedure for calculating sclera area is same i.e., gaze estimation along horizontal axis is same. However, both methods estimate gaze differently along vertical axis. In the first method, distance between the tracking point and bottom of eye along iris axis is taken as the criteria for estimating gaze. The advantage of this technique is that it is computationally faster as we just have to compare the distances between the tracking and calibration points along iris. However, this technique is user-dependent i.e., we cannot have a constant distance between the tracking point and bottom of the eye as some of the users may have smaller eyes and some may have larger eyes. In the second technique, gaze is estimated using iris area along vertical axis. The advantage of this technique is that it is much more efficient since iris area is not user dependent but computationally expensive. However, both the methods fail to estimate gaze accurately while looking at the center of the screen, i.e., switching from top to center or from bottom to center, since there is not much deviation either in the iris position or in the distance between eye lids. The gaze estimation results using both methods are as shown in fig.19:
Fig 19: Experimental results for gaze estimation.

In fig.19 X-axis indicates the points on the screen and Y-axis indicates number of people for which the gaze has been estimated correctly. As you can see from fig.19 that method 2 gives better efficiency while looking at top and bottom of the screen, however its efficiency decreases while looking at the center.
The fig.20 shows extraction of sclera and Iris using adaptive thresholding technique.

Fig20: Experimental results of sclera and iris extraction using adaptive thresholding. (i) Input image (ii) Output where white part indicates that sclera region and black part indicates iris region using adaptive thresholding (iii) Color map of (ii). (iv) Showing sclera regions (v) Showing Iris regions.
As we can see in fig.21, that global threshold has to be changed for images under different conditions which is not the case with adaptive thresholding since it considers only a small region say 3x3 window at any given point of time.
6. CONCLUSION

6.1. Contributions and Limitations:

Gaze is an important tool in many applications such as human-computer interaction, in medical field (helpful in studying neurological, vision and communication disorders) and also to gain some insight about how people view synthesized images and animations with the dual purpose of optimizing perceived quality and developing more efficient algorithms. One of the major drawbacks of gaze estimation is the cost involving equipments such as high performance cameras followed by the burden of wearing different equipment’s by user. In this report, we proposed a new gaze estimation procedure which makes use of a simple web-cam which reduces the cost considerably. Contrast to the present gaze estimation techniques, the procedure is an intrusive one i.e., the user needs to select a point on each eye, usually selecting a point on top of an eye and the selected point will be tracked for the entire duration of video. Because of the time frame we limited our experiments to static head, however, sclera and pupil of an eye have been extracted with good accuracy and gaze is estimated with a good success rate while looking at the corners of screen and moderate success rate while looking at other areas of screen. Even though the gaze estimation is not entirely accurate, but it is right step in the direction of reducing the cost of gaze estimation and also making it simpler to use. In future, the onus will be on increasing the accuracy by using a better mathematical approach for determining gaze by considering the geometric properties of eye, making it more robust to different lighting conditions, and also estimate gaze with head moment.
REFERENCES

1) Eye Tracking Methodology Book by Duchowski, Andrew


23) Gary Bradski, Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library.
Vita

- Date and Place of Birth: 06/06/1985, Hyderabad, India.
- Educational Institutions: Bachelor of Technology in Electronics and Communications Engineering at Jawaharlal Nehru Technological University, Hyderabad, India. Master of Science in Electrical Engineering at University of Kentucky, Lexington, Kentucky, USA.
- Professional Positions: Member of Technical Staff at CloudGrapes Inc, California. Image Processing Engineer at Logovision Inc, Colorado.

Prashanth Rao Periketi.