Enhanced Eye Gaze Direction Classification Using a Combination of Face Detection, CHT and SVM

Amer Al-Rahayfiah
Department of Computer Science and Engineering
University of Bridgeport
Bridgeport, the United States
aalarhay@my.ridgeport.edu

Miad Faezipour
Department of Computer Science and Engineering
University of Bridgeport
Bridgeport, the United States
mfaezipo@bridgeport.edu

Abstract—Automatic estimation of eye gaze direction is an interesting research area in the field of computer vision that is growing rapidly with its wide range of potential applications. However, it is still a very challenging task to implement a robust eye gaze classification system. This paper proposes a robust eye detection system that uses face detection for finding the eyes region. The Circular Hough Transform (CHT) is used for locating the center of the iris. The parameters of the Circular Hough Transform are dynamically calculated based on the detected face information. A new method for eye gaze direction classification using Support Vector Machine (SVM) is introduced and combined with Circular Hough Transform to complete the task required. The experiments were performed on a database containing 4000 images of 40 subjects from different ages and genders. The algorithm achieved a classification accuracy of up to 92.1%.

Keywords — Eye direction detection, face detection, Circular Hough Transform, SVM, Viola-Jones.

I. INTRODUCTION

Eye is the most direct way of interaction with the surrounding world and therefore, eye detection and gaze estimation has become a significantly active research area. Different approaches are used to estimate the direction of eye gaze. Being a convenient and natural way of interaction, the field of eye gaze estimation is growing and it is used in various applications, such as Human Computer Interaction (HCI), virtual reality, driver assistance systems and assistance for people with disabilities.

Eye gaze classification systems aim to locate the eye in the image and then use the obtained information about the eye region and the head pose to estimate the eye gaze direction to be used in applications such as cursor pointing on the screen.

Eye gaze can reveal the attention and express the interest of a person. The eye movements also contribute in expressing the person’s emotional states [1]. Eye gaze and movement are required to identify and describe the visual world. Hence, it becomes very demanding to develop robust eye detection and gaze estimation methods to improve human-computer interaction and take it to the next level.

Many applications which are based on eye gaze direction classification technology have been developed in different fields. The first step in many applications in the field of computer vision is eye detection. These applications include iris detection and recognition, facial recognition, facial feature tracking and expression analysis. The next processing steps in these applications depend on the accuracy of the eye detection phase. Although it is not investigated as much, another application that employs eye gaze estimation is to obtain insight about the perception animations and synthesized images for quality optimization and development of efficient algorithms.

Donder [2] stated that there is only a subset of positions to which the position of the eye is restricted. According to Donder’s law, the eye orientation can be uniquely determined based on the eye gaze direction. The law also states that the eye orientation is independent of the previous eye positions.

This paper introduces a new method for eye direction classification, evaluates the performance of a real-time eye gaze direction classification system and suggests some performance enhancements. The proposed enhancements focus on required CPU processing time.

The rest of the paper is outlined as the following: Section II introduces different eye detection and gaze direction classification methods available in literature. Section III describes the methods used in the proposed system. Section IV explains the eye gaze direction classification procedure. In Section V, the performance of the proposed system is evaluated and the results of the experiments are discussed. Finally, Section VI draws the conclusion of this paper.

II. RELATED WORK

There has been much research working to develop robust eye detection and gaze estimation methods. Eye trackers differ in the degrees of freedom which they can track. Simple eye trackers report only the direction of the gaze relatively to the head.

Many eye detection and gaze estimation algorithms have been presented in literature. Chern-Shen Lin et al. [3] presented a method based on mapping polar coordinates to detect the eye gaze. Morinoto et al. [4] proposed an eye gaze...
estimation scheme that determines the eye gaze direction based on the glint-pupil vector which is defined as the vector between the centers of the corneal reflection (CR) and the eye pupil. A second-order polynomial equation is used for the computation of gaze point coordinates.

Beymer and Flickner [5] introduced a 3D eye tracking approach using the 3D eye anatomy model. This approach has a drawback as it needs four cameras and two lighting points.

Yoo et al. [6] presented an algorithm for eye detection based on cross-ratio-invariance but it achieved very low accuracy. To improve the system accuracy, a virtual plane that is tangent to the eye cornea is proposed. Even with the suggested enhancement, the method complexity issue is still present because two cameras are needed to detect the position of the eye pupil using the difference between dark and bright pupil images captured by the two cameras [7].

Kothari and Mitchell [8] introduced a voting scheme that detects the location of the eyes using spatial and temporal information based on the fact that the gradient along the boundary of the iris goes outward from the center. False pupil candidates are filtered using heuristic rules and a large temporal support.

Gao et al. [9] used Hough Transform to find the coordinates of the pupil center and to determine the direction of the eye gaze. M. S. Devi et al. [10] proposed a system for driver fatigue state detection. They find the face using skin color and then apply Circular Hough Transform to detect the eyes.

Mehrubeoglu et al. [11] used a smart camera for eye detection. It stores the location of previous matches appearing in successive image frames. The system includes pattern matching, edge detection, and computation and recording of eye center coordinates.

Alioua et al. [12] presented a robust eye detection method based on Circular Hough Transform that does not need training data or the use of a special camera.

### III. System Overview

This section discusses the details of the eye gaze direction classification system being investigated. The system basically detects the face in the image and applies Circular Hough Transform to detect the iris of the eye. The gaze direction is determined based on the coordinates of the iris center.

#### A. Viola-Jones Face Detection

The Viola-Jones face detector included three main concepts which led to successful real-time face detection. These three concepts are: the Integral Image, AdaBoost classifier and Cascading classifiers [13].

The Viola-Jones method does not use image intensities directly but calculates features that are reminiscent of Haar Basis functions. Using an intermediate image presentation; the integral image; it is possible to achieve fast evaluation of the features.

The integral image is computed by performing by a few operations per pixel on the processed image. At a location \( x, y \) in the integral image, the point has a value equals to the sum of the pixels above and to the left of the specified location \( x, y \) [13]:

\[
ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')
\]

where \( ii(x, y) \) is the integral image point and \( i(x, y) \) is the pixel at location \((x, y)\) in the original image. The integral image is the double integral of the image. Fig. 1 shows how the integral image points are calculated.

![Fig. 1. Integral Image calculation example.](image)

Three types of features are used by Viola-Jones face detector [13]. The two-rectangle feature is calculated by defining two rectangular regions which are of same size and shape and are horizontally or vertically adjacent and finding the difference between the sums of the values of pixels in them. Three rectangular regions, two outside rectangles and a center rectangle, are defined and the sum of the outside rectangles is subtracted from the center rectangle to calculate the three-rectangle feature. A four-rectangle feature computes the difference between diagonal pairs of rectangles. Fig. 2 illustrates these features.

![Fig. 2. Rectangular features: (A) and (B) The two-rectangle feature, (C) The three-rectangle feature, (D) The four-rectangle feature.](image)
used in an effective classifier. The Viola-Jones algorithm uses variant of AdaBoost to select these features and to train the classifier.

The third concept Viola-Jones algorithm is The Attentional Cascade. It is a technique used to cascade classifiers to reduce computation time and improve the accuracy of detection. Cascade stages are constructed by using AdaBoost for training classifiers [13].

B. Circular Hough Transform

To locate the eye in the captured images, the center of the iris is detected using Circular Hough Transform. There are several versions of modified Hough transform which are considered an effective method used for curve detection in images. The Circular Hough Transform (CHT) is a modified version of the Hough Transform. It was first introduced by Duda et al. [14]. The Circular Hough Transform recognizes circular patterns in the processed image. These patterns may be a complete circle or a circular curve. It transforms the sets of feature points in the image into sets of accumulated votes in parameter space. For each feature point, votes are accumulated in an array which spans all parameter combinations. The highest number of votes indicates that there is a shape. The circle pattern which the Circular Hough Transform detects is defined by Equation 2.

$$\left( x_p - x_0 \right)^2 + \left( y_p - y_0 \right)^2 = r^2$$  \hspace{1cm} (2)

where $x_0$ and $y_0$ are the coordinates of the center and $r$ is the radius. An example demonstrating CHT is illustrated in Fig. 3.

![Circular Hough Transform](image)

Fig. 3. The Circular Hough Transform.

A group of edge-points in an image is represented by the dark blue circles in Fig. 3. A circle of radius $R$ is added to a 3-dimensional parameter space output accumulator for each of the edge-points. These circles are represented by the light blue circles. The 3-dimensional space parameters are the $x$ and $y$ coordinates of the circle center and its radius. The peak value in the accumulator occurs at the point where the circles overlap which is the center of the original circle.

C. Support Vector Machine

Support Vector Machine (SVM) has gained high popularity in the field of object tracking. SVMs are supervised learning models based on some learning algorithms in which data is analyzed and used in classification.

The input data is mapped into an $N$ dimensional space using a kernel function, such as linear, quadratic, polynomial, Gaussian Radial Basis and multilayer perceptron.

The data points existing on the margin are called the support vectors. These support vectors are used by SVM to define the optimal separating hyper-plane between the classes of input data which leaves the maximum margin from both classes [15]. Fig. 4 shows a linear SVM hyper-plane in a 2-dimensional feature space.

![Linear SVM hyper-plane](image)

Fig. 4. Linear SVM hyper-plane [15].

The Figure illustrates two possible hyper-planes. However, the hyper-plane in Direction 2 leaves a larger margin for both classes and thus, it is chosen as a separating hyper-plane.

D. HSI Color Space

HSI is conic-coordinate representation of color points rearranging the geometry of RGB color representation. HSI stands for hue, saturation, and intensity. Hue describes visual sensation that an area appears to be closer to one of the colors: red, yellow, green, and blue, or a combination of two of these colors. Saturation is the colorfulness of the pixel relative to its brightness. Intensity is the total light amount passing through. Fig. 5 shows the representation of the HSI color space.

![HSI color space](image)

Fig. 5. Representation of HSI color space [16].
HSI separates luminance, which is the image intensity, from chrominance which is the color information. This separation is very useful in eliminating the effect of changes in lighting conditions.

E. Algorithm Implementation

The proposed eye gaze direction classification algorithm first detects the face in the image using Viola-Jones face detector [13]. Defining the face region is beneficial in defining the minimum and maximum values for the eye radius along with other values to be used as CHT parameters. The values are calculated based on the detected face height.

In addition, when the region scanned for circles is minimized, the number of detected circles, which are considered eye candidates, is reduced. This increases the accuracy of the Circular Hough Transform in eye detection.

After the face is detected, the algorithm employs the Circular Hough Transform for eye detection. The parameters of the Circular Hough Transform should be assigned appropriate values. These parameters include radius, threshold, and delta. The Circular Hough Transform is flexible in determining the detected circle radius. The minimum and maximum values of the radius are set and circles having radii within the given range are detected by the transform. The threshold is used to create a binary edge map and normalization is performed on the image before thresholding. Delta is defined as the difference between the circles detected by the transform at which they are treated as the same circle. Delta considers the difference in the centers coordinates and the radii of the circles. Only one of these circles is returned and the others are dismissed.

The Circular Hough Transform detects the circles in the provided eye region using the parameters calculated from the detected face information where the radius range is set to be from 3% to 4.5% of the detected face height and delta is set to be 20% of the face height. The threshold parameter is set to 0.15. The selection of these values will be discussed in detail in Section V.

The detected circles are iris candidates. Only two of the detected circles are true irises. To determine them, the two circles having more circular curve edge points that achieve the circle equation above (Equation 2) are considered to be the true irises. To find which of the two true eye irises is clearer and more suitable for use in classification, the one with the darkest color is selected due to the black pupil in the middle of the iris. An example of eye detection using the presented algorithm is shown in Fig. 6.

The Circular Hough Transform has no radius limitations, where the minimum radius allowed is 1 and circles with radii of a positive integer value can be detected. Fig. 7 demonstrates the success of the Circular Hough Transform in eye detection in an image resized to 4 different resolutions. Appropriate values were used for the transform parameters with each size of the image.

The processed image resolution is a factor that affects the required CPU processing time since the method requires pixel-by-pixel processing. Fig. 8 shows the relation between the image resolution and the required CPU time when the used Circular Hough Transform parameters are the same for all image sizes.

As it can be concluded, the required CPU time is larger for images with higher resolution.
IV. EYE DIRECTION DETECTION

Fig. 9 illustrates the proposed eye gaze direction classification system stages. As explained earlier, the Circular Hough Transform is used to locate the eye iris. After the iris is located, the direction of the eye gaze needs to be determined and classified into one of two directions: right and left.

The Circular Hough Transform defines the circles it detects by 3 values: $x$ and $y$ center coordinates and circle radius. These values are used in the preprocessing stage to extract a sub-image from the original video frame after the image is converted to HSI scale. The extracted sub-image, highlighted in yellow, is large to ensure obtaining information about the eye region completely.

Two regions are defined in the extracted sub-image: right and left regions. Each of these regions has a width and height equal to the iris diameter. The left and right regions are used in determining the eye gaze direction. Depending on the eye gaze direction, one of the two regions, the left or the right, will contain the white area of the eye while the other will contain skin area. In HSI color space, the white and black colors are given maximum hue values while the normal colors will have normal hue values (less than maximum). Thus, there is a noticeable difference between the hue of the white area of the eye and the skin area, as shown in Fig. 10, and hence, the average of the $H$ component from HSI scale for each region, the left and the right, can be used as a good feature for eye direction classification.

The average is calculated according to Equation 3:

$$\text{avg} = \frac{1}{M} \sum_{k=1}^{M} X_k$$

where $M$ is the number of pixels in the region and $X$ is the hue value of the pixel. Fig. 11 shows the detailed block diagram that sums up the preprocessing stage for direction detection.

To determine whether the gaze direction is left or right, SVM classifier is used. The input to SVM is the feature vector containing the averages of the hue value of the left and right regions of the subimage. The SVM was trained using a training dataset consisting of images captured for 40 subjects. These subjects are from different ages and genders. The dataset used is accessible online [17]. The dataset contains 4000 images containing left and right eye movements.

V. EXPERIMENTAL RESULTS

This paper presents an eye direction classification system and investigates its performance. The experimental results of this paper are presented in this section. The experiments on the proposed eye gaze classification system were carried out on MATLAB.

In the proposed eye direction classification system, the eye is detected using Circular Hough Transform. As mentioned earlier, the performance of CHT in detecting the eyes can be enhanced by using face detection. After the face is detected, its size can be used to calculate the minimum and maximum values of the radius used by CHT for eye detection. Fig. 12 shows the range of possible iris radius values compared to the face width. This range has been calculated by detecting the face in 40 images and finding the iris radius in pixels in each image. The ratio of iris radius to the detected face height ranges from 3% to 4.5%.

The relation between the face height and the delta parameter of the Circular Hough Transform was investigated as well. The minimum delta found was 30% of the face height. To provide an error margin, 20% of the face height is set as the delta value used in CHT.

In addition, the eye region can be defined from the detected face area. This minimizes the number of detected eye candidates because it is more likely that the original image contains other circles than the eyes, especially with noisy image backgrounds. Fig. 13 compares the number of the circles detected in the image with and without using detected face information to define the eye region.
The eye direction is classified using SVM. The accuracy of the proposed system was tested using the 4000 images database [17]. For training data, 3600 images were used and the remaining 400 images were used as test samples. Fig. 14 shows the accuracy of eye detection using SVM with different kernels. The highest accuracy is achieved when using the linear SVM kernel which is equal to 92.1%. This result shows that it is possible to use a classification function to determine the eye gaze direction.

**REFERENCES**


