A Feature-Based Color Threshold Selection Approach to Pupil Tracking

Vijay Pillai
Bard College at Simon's Rock
Great Barrington, Massachusetts 01230
Email: vijay@vijaypillai.com

Harold Hastings
Bard College at Simon's Rock
Great Barrington, Massachusetts 01230
Email: hhastings@simons-rock.edu

***Abstract*—Pupil tracking can be achieved by intrusive methods (where equipment is on or inside the user) and non-intrusive methods. Intrusive methods are not preferred because they are frequently cumbersome and require a large investment in equipment. One non-intrusive approach to eye-gaze tracking involves active infrared illumination. This research investigates a novel approach to bright-pupil tracking that takes advantage of the data resolution provided by color. This method is shown to be functional and potentially robust.**

I. Introduction

Pupil tracking has been researched heavily in the past, and there are many established methods for eye and gaze tracking. However, common users have yet to see a reasonable, effective, or readily available application of eye tracking to the user interface. This lack of consumer solutions may be attributed to the fact that, until very recently, most commercially available solutions for eye tracking were expensive and relied on proprietary software or complex infrastructure. In addition, solutions discovered in academia frequently remain undeveloped, inaccessible, and complex in design.

Both intrusive and non-intrusive approaches to pupil tracking have been proposed. Intrusive methods are frequently complex; some require electrodes to measure skin potentials, or inductive contact lenses and electromagnetic field generators [1]. Most consumers are unlikely to learn complex procedures for manipulating a computer, so intrusive methods are impractical for the consumer market. This study will focus solely on non-intrusive approaches.

There are two classes of commonly used non-intrusive approaches: eye tracking by entirely mathematical approaches, and eye tracking enhanced by active infrared (IR) illumination. Much research has been done on purely mathematical approaches. While some of these methods have proven to be accurate and robust, many are not suited for eye tracking under variable conditions (especially those where the device is not worn and body movement is not restricted). Active illumination offers an increase in signal to noise ratio, reducing the complexity of these methods.

Eye tracking with active IR illumination uses Purkinje images, which are retroreflections from the eye; put simply, light projected into the eye is reflected back. The first and fourth Purkinje images (also respectively denoted as dark-pupil and bright-pupil images) have been found to be particularly useful for eye tracking. Lighting on the axis from the camera lens to pupil might be used to create a bright-pupil image (Figure 1). Similarly, an off-axis light source might be used to create a dark-pupil image.

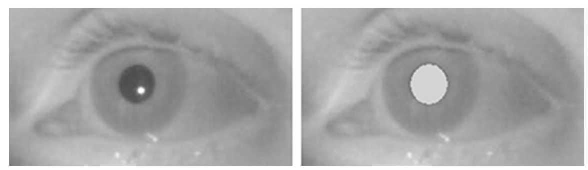

Fig. 1. Dark Pupil (left), Bright Pupil (right), reprinted from [2] in JNER under CC BY 2.0.

Many researchers have used Purkinje images to approach pupil tracking. Some use only bright pupils, which create a distinct color contrast between the iris and the pupil for pupil detection. Similarly, other methods use only dark pupils. For example, Wang et al. developed a wearable head-mounted system that isolates the pupil by using off-axis lighting to create interference points in the pupil [3]. Gneo et al. also takes a dark-pupil only approach, using a triangular pattern of lights and evaluating their use for pupil center estimation by checking the geometry of the reflections [2].

An increasingly popular approach combines both on-axis and off-axis illumination from two separate arrays of lights, one off-axis and the other on-axis. Alternately exciting each array of lights yields a composite video in which frames alternate between dark pupil and bright pupil images. Aside from some noise, the notable difference in these images is the presence of the bright pupil. These images are then subtracted such that the bright pupil image can be isolated. This approach has proved to be rather robust and, unlike methods that use only bright or dark-pupil images, frequently functions even when the user is wearing glasses [4]. Furthermore, there have been many attempts to improve this approach [5]–[7].

However, these previous approaches have several flaws. The method proposed in Gneo et al. discards frames where the reflections move onto the sclera, and wearable mechanisms like the one proposed by Wang et al. are unwieldy, bulky, and limit use modes. The need for both dark and bright pupil images to detect a pupil limits the frame rate of pupil tracking to half of the camera's maximum frame rate. Additionally, multiplexing lights and using tilt-pan mechanisms to track the face introduce circuital complexity, cost, and causes for error. Dark-pupil methods require the camera to be too close to the face or to have a zoom lense, which makes adjusting for head movement difficult, and these methods tend to be single-eye, which limits data resolution. Finally, multi-camera solutions increase cost and setup-complexity.

Most aforementioned studies use black and white imaging for analysis of Purkinje images, and there is comparatively little research on active illumination tracking that takes advantage of the data resolution provided by color. For this reason, we investigated a novel bright-pupil only approach to eye tracking inspired by the aforementioned studies and which attempts to solve some of their problems. This study was approved by Bard College at Simon's Rock Human Research Review Committee (IRB). Our device is safe for use and has irradiance $\leq 1mW/cm^2$, well below the $10mW/cm^2$ recommended for long-term exposure (IEC 62471). Consent was obtained for testing and image publication from all subjects.

II. Setup

Our device was based on the Logitech C920 webcam. The C920 has an internal wide bandpass filter, whose removal allows detection of IR light. The first step was disassembly of the camera followed by removal of the bandpass filter. After this, a coaxial illuminator was constructed on perfboard and fitted with eight 940nm 5mm LEDs. This design was inspired by images in articles in which active IR illumination was also used [4], [5] [7], [8]. The camera and illuminator circuit boards were remounted together inside a solid framework to produce consistently reproducible imaging (Figure 2).

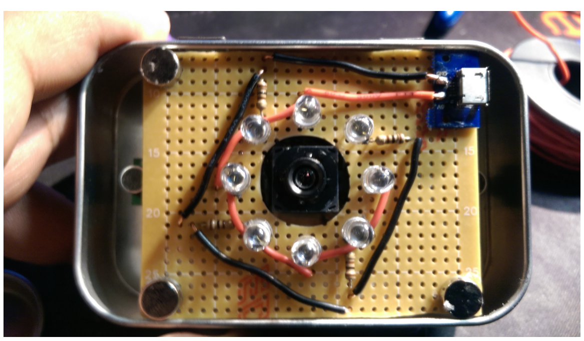

Fig. 2. Final Assembly

III. Pupil Tracking Algorithm

Our hypothesis is that the hue, saturation, and color values of the user's pupils can be thresholded and applied to each frame given by the camera such that all pixels of the same values are isolated. The first step in extracting the position of the pupils is to identify the color thresholds that coincide with the bright-pupil image. This step is the calibration step. The program grabs a frame and uses Viola-Jones Haar Feature-Based Cascade Classification [9] to identify the user's face. This image is then cut based on natural anatomical proportions adapted from Tristain Hume's implementation of methods presented by Timm and Barth (Figure 3) [10].

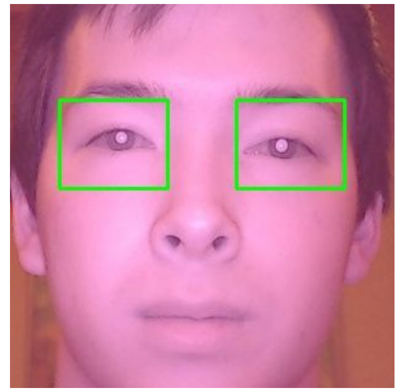

Fig. 3. Eye ROIs (Regions of Interest)

The pupils are further extracted from this image by converting the image to grayscale and then identifying the brightest pixel in the image, which coincides with the first Purkinje image. A circular mask is used to extract the pixels within a constant radius to roughly extract the pupils from the color image (Figure 4).

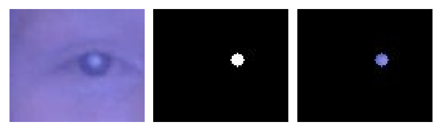

Fig. 4. Pupil Extraction

These images are then used for threshold calibration. The thresholding method was inspired by the Otsu method for black and white images used in Wang et al. [3]. Colour thresholds are adaptive, in that they are identified for each session based on the features of a histogram of color value frequencies for the masked pupil images. Thus, thresholds are not constant across sessions and can be recalculated for settings with different color profiles. We begin our threshold estimation by converting the pupil images to the HSV (Hue, Saturation, Value) color model. This allows us to better isolate the particular color of the pupils while allowing for a broader range of intensities. We then generate frequency histograms for the pupils with the OpenCV calcHist function. The H, S, and V distribution of both images are generated such that all black pixels are ignored (Figure 6).

We select the H (hue) value for our threshold by selecting the brightest and darkest pixels in an array composed of the non-black pixels of the right and left masked pupils. The brightest pixel's H is added to the upper bound for the new threshold, and the darkest H is added to the lower bound. Following this, the upper bounds for S (saturation) and V (value) are identified using an algorithm hereafter referred to as the "cohort isolator function" (Figure 5). This is necessary, for the brightest and darkest pixel HSVs are, in practice, insufficient for successfully thresholding the pupils. The cohort isolator function is designed to identify a set of maxima in the histogram given distance parameters. The maxima with the larger value is used as an upper bound for thresholding. This function was inspired by the observation that appropriate color thresholds for the pupils frequently coincided with prominent features (i.e. sets of maxima) in the histogram. It works by identifying "cohorts", or the values of maxima with the same frequencies. The upper bound is set to the value of the maximum with the greatest frequency. The lower bound is set to the nearest value that belongs to a cohort that has a frequency difference greater than a given number (termed "cohortgap") and a numerical difference greater than a given number (termed "mingap"). If the lower bound has a greater value than the upper bound, then the two are swapped. The upper bound is then returned. The lower bounds for S and V are then determined by subtracting constants from their respective upper bounds.

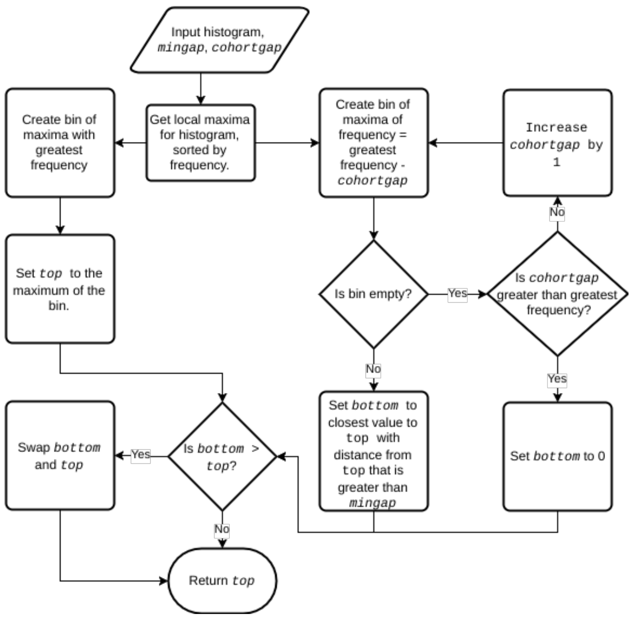

Fig. 5. Cohort Isolator Logic

This threshold is applied to the image and the output is displayed to allow the user to evaluate the efficacy of the calibration. A good calibration is distinguished by solid and complete extraction of the pupils (Figure 7). The calibration step is repeated until the user approves three calibrations, and is completed with computing the average of the three approved color thresholds. This method was found to be much more effective than simply thresholding over the brightest and darkest colors in the pupils. Poor calibration causes the program to fail at tracking the pupils.

Following calibration, Haar-cascade feature detection is applied to each frame. The ROIs for both eyes are then blurred to further reduce noise, converted to HSV, thresholded, converted to grayscale, and blurred again to reduce non-pupil artifacts. Lists of non-black pixels are compiled for both ROIs. These lists are then put through Nimrod Meggido's algorithm,

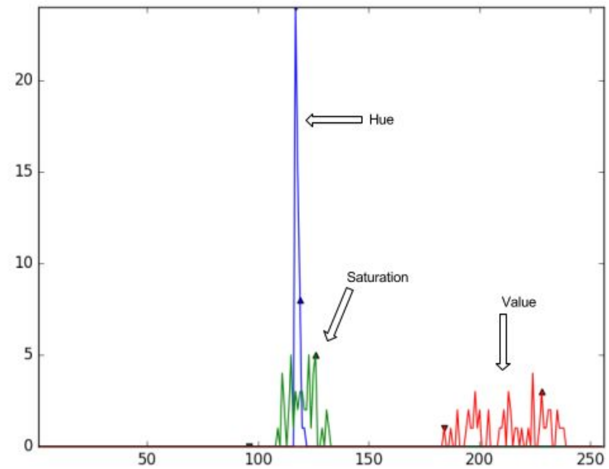

Fig. 6. HSV Histogram With Thresholds Marked

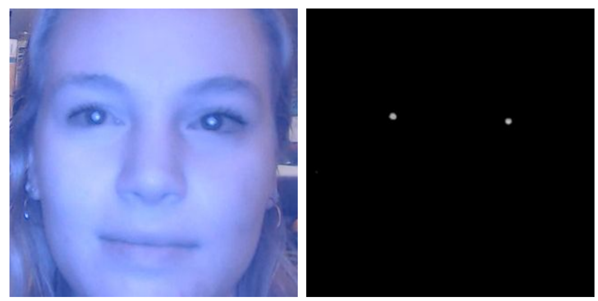

Fig. 7. Example: Good Calibration

a linear-time algorithm for calculating the minimum enclosing circle for a set of points in Cartesian space. The algorithm returns the center and radius of the minimum circle that fits the given collection of points. Thus, we obtain the center of the pupil for both eyes. The dimensions of the face and eyes can be used to translate these coordinates for the original frame. The algorithm finally yields a set of pixels that coincide with the region that contains the user's eyes and are within the bounds of the preprocessed color thresholds. A minimum-fitting circle is applied to these pixels, and the center of the pupil is found (Figure 8).

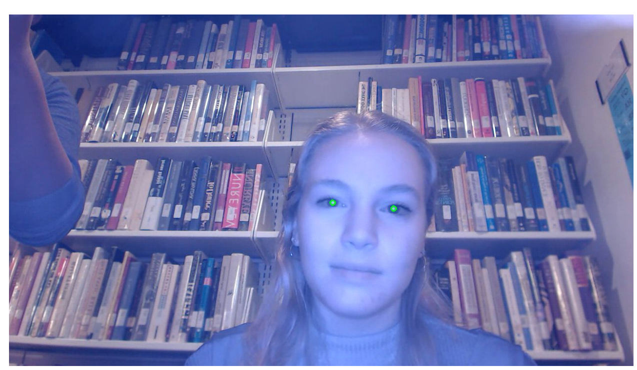

Fig. 8. Program output with accurate pupil detection.

Statistic	%Success $\geq$ One Eye	%Failure $\geq$ One Eye	%Success Both Eyes	%Success Both Eyes
Average [a]	88.5	11.5	60.2	39.8
Std Dev	11.3	11.3	27	27

TABLE I
TEST RESULTS

[a]Program could not be calibrated for subject #12

IV. DISCUSSION

The method of pupil tracking proposed in has been observed to function with fairly good accuracy and rate of success. In 900 sequential images of 12 subjects each, there was 88.5% success for accurate detection of at least one eye, 60% success for accurate detection of both eyes, and one failure to calibrate (Table I). Additionally, it does not suffer from certain common issues that have plagued pupil tracking systems in the past. The camera need not be too close to the user, allowing for more freedom of motion. Furthermore, head motion, often not taken into consideration, is accounted for using feature detection. In other words, the user's pupils can be tracked no matter where they are in the frame if they are facing the camera. Finally, the entire system was built upon a low-cost webcam and an easily replicated, non-multiplexed illuminator circuit with a total cost of approximately $120. As such, this system is much more accessible than previously existing eye tracking systems, and if mass produced, potentially more accessible than some newly introduced solutions.

A. Causes For Error and Proposed Solutions

As expected for a method that uses bright-pupil tracking, the success of the algorithm depends highly on lighting conditions. There must be no sunlight in the room for the bright pupil to appear at the proper intensity. Furthermore, as mentioned by Gneo et al., 5-10% of people have insufficiently intense bright-pupils to allow reliable extraction [2]. Several trials over an ethnically diverse sample set further confirmed that the most heavily affected population in this regard is Asians, as reported by Amir et al. [11]. Note that the light multiplexing and image subtraction method also suffers from these deficiencies.

The current implementation also suffers from latency issues. There is a loss in frame rate and the algorithm takes a few seconds to catch up upon head movement, despite the fact that the frame capture rate is 30 fps and the face detection thread is queried every 7 frames (implying a refresh of dimensions 4 times per second). It is expected that these latency issues arise from prototyping in Python, a dynamically typed and interpreted language. Thus, the current implementation cannot benefit from the speed of static typing or compiler optimizations. Accordingly, the first improvement would be to rewrite the code in C++; note that the OpenCV functions used to implement the method are also available for C++. As C++ benefits from compiler optimizations and static typing, a significant decrease in latency can be expected upon this change.

The final notable source for error would be the first Purkinje image. At certain angles, lighting conditions, and threshold calibrations, the first Purkinje image can be produced by the illuminator or other light sources in the immediate vicinity. This noise can result in scleral artifacts in the thresholded image. These artifacts can cause the minimum circle algorithm to extract more than the pupil alone, which leads to incorrect estimation of the pupil center. This problem can be remedied by finding a method of thresholding the first Purkinje image alone, with a separate calibration and off-axis lighting. We have already established that color thresholding can extract distinct features with fair accuracy. The color threshold of the first Purkinje image would be leveraged to remove corresponding pixels after applying the threshold of the bright pupil. This would likely resolve the issue of scleral artifacts.

V. CONCLUSION

We have presented a novel and potentially robust method for pupil tracking. The equipment used is low-cost, easily obtained, and calibration of the software is sufficiently simple for this system to be used by the average individual. This work concludes with the acknowledgement of exciting potential for future research in eye-gaze tracking and gaze interface. There is much work to be done, but interface by eye-gaze is an achievable and worthwhile goal for the advancement of human-computer interaction.

ACKNOWLEDGMENT

This work is based upon the author's senior thesis under professors Harold Hastings, Jackson Liscombe, Amanda Landi and Aaron Williams (advisor).

REFERENCES

[1] Z. Zhu, K. Fujimura, and Q. Ji, "Real-time eye detection and tracking under various light conditions," in *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ser. ETRA '02. New York, NY, USA: ACM, 2002, pp. 139–144.

[2] M. Gneo, M. Schmid, S. Conforto, and T. D'Alessio, "A free geometry model-independent neural eye-gaze tracking system," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, p. 82, Nov. 2012.

[3] J. Wang, G. Zhang, and J. Shi, "Pupil and Glint Detection Using Wearable Camera Sensor and Near-Infrared LED Array," *Sensors*, vol. 15, no. 12, pp. 30 126–30 141, Dec. 2015.

[4] C. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil Detection and Tracking Using Multiple Light Sources," *Image and Vision Computing*, vol. 18, no. 4, pp. 331–335, Mar. 2000.

[5] L. Atan, "Multi-Person Infrared Pupil Tracking for 3d TV without Glasses," Ph.D. dissertation, Ume University, Jun. 2011.

[6] B. Noureddin, P. Lawrence, and C. Man, "A non-contact device for tracking gaze in a human computer interface," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 52–82, Apr. 2005.

[7] Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 124–154, Apr. 2005.

[8] C. H. Morimoto, D. Koons, A. Amir, M. Flickner, and S. Zhai, "Keeping an Eye for HCI," in *XII Brazilian Symposium on Computer Graphics and Image Processing*. IEEE, Oct. 1999, pp. 171–177.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I.

[10] F. Timm and E. Barth, "Accurate Eye Centre Localisation by Means of Gradients." *International Conference on Computer Vision Theory and Applications*, vol. 11, pp. 125–130, 2011.

[11] A. Amir, L. Zimet, A. Sangiovanni-Vincentelli, and S. Kao, "An embedded system for an eye-detection sensor," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 104–123, Apr. 2005.