Development of a low-cost, portable, tablet-based eye tracking system for children with impairments

Harshita Karamchandani
Medical Device Research Institute
Flinders University
GPO Box 2100 Adelaide SA 5001
+618 8201 3167
harshita.karamchandani@flinders.edu.au

David Hobbs
Medical Device Research Institute
Flinders University
GPO Box 2100 Adelaide SA 5001
+618 8201 3167
david.hobbs@flinders.edu.au

Dr. Tom Chau
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Rd, Toronto ON M4G1R8
+1 416 425 6220 x3515
tom.chau@utoronto.ca

Leslie Mumford
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Rd, Toronto ON M4G1R8
+1 416 425 6220
leslie.mumford@utoronto.ca

ABSTRACT

Eye tracking technology can enable children with severe speech and motor impairment to communicate. Eye tracking systems for the use of human computer interaction have long been an area of interest in the assistive technology field. However, a number of factors have prevented eye tracking from being an accessible technology, including the invasiveness, robustness, availability, and cost of eye tracking systems. Moreover, a common drawback of some commercial eye tracking systems is that head motion is not typically considered, and many systems are not portable or mobile.

This work describes the design and development of an eye tracking system for children with disabilities. The system is an economical alternative to commercially available devices. It does not require any sophisticated hardware, and is tablet-based. It uses raw images from a webcam and relies on distinct features of the eye that can be detected and tracked using image processing functions. The simple eye tracking system can differentiate between 16 different points of gaze enabling the user to have access to 16 options on a 4 by 4 display.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – Tracking. H.5.2 [Information Interfaces and Presentation]: User Interface – Eye Tracking System.

B.4.2 **[Input/Output and Data Communications]**: Input/Output Devices.

General Terms

Algorithms, Design, Human Factors.

Keywords

Eye tracking, impairment, children, low cost, webcam, feature based, access, communication, computer vision

1. INTRODUCTION

Communication is an essential part of every individual's life. It is the means by which we interact, share experiences, and express ourselves. Children with conditions such as cerebral palsy (CP) or spinal muscular atrophy (SMA) commonly experience significant difficulty with communication due to limited or absent speech. Moreover, motor impairment in these individuals can restrict the use of movement to communicate [1].

Eye tracking technology can be a useful alternative where good control of the eyes and head is retained. Most commercial eye trackers are accurate and provide a range of features. However, full eye tracking systems are often very complex, not mobile, and expensive, ranging from $5,000 to $40,000 [2].

A simple, portable, non-invasive and inexpensive eye tracking system intended for use by children with severe disabilities was developed. It does not require any external hardware, unlike many systems that require sophisticated cameras and high performance hardware.

2. BACKGROUND

There are two methods that are commonly used to track the eyes: infrared illumination, which is used in commercial systems, and computer vision, which can be used with basic off-the-shelf or inbuilt webcams in laptops and tablets. There are three main types of computer vision methods: template-, appearance- and feature-based methods [3]. A hybrid of the template- based and feature-based methods was used in the implementation of the system reported here. As head movement is an inherent problem in eye tracking systems, this was also considered in the design of the system.

2.1. The human eye

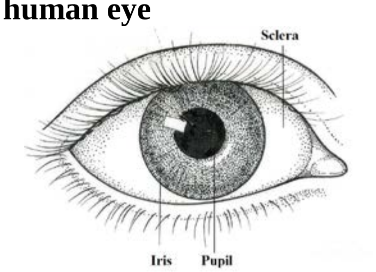

Figure 1. Features of the eye

Computer vision methods rely on the presence of distinct features that can be easily tracked by a camera. The most visible and distinguishable part of the human eye is the iris as shown in Figure 1. Due to its distinctive presence in an eye image, the iris is a commonly tracked feature to determine the direction of gaze [4]. The limbus is the boundary between the sclera and the iris. Due to the contrast of these two regions, the limbus is easily detected and tracked, especially in individuals with dark coloured irises.

2.2. Design requirements

The aim of this project was to design an eye tracking system that met the following requirements:

1. Portable, so as to allow for use in both the home and school environments;
2. Intuitive and easy to use by children with disabilities;
3. Non-intrusive, so that it does not cause any harm or discomfort to the user;
4. Accurate so as to provide 16 different options in a 4 by 4 grid at any given time;
5. Reliable so as to ensure a consistent output to allow ease in communication;
6. Able to provide the user with enough options to enable functional communication;
7. Able to run in real time so that the user can express themselves without delays.

2.3. Target population

This eye tracking system was intended for children who are unable to use other forms of access technology for communication. Given the 'proof of concept' nature of the project, the eye tracking system that was eventually produced offered a limited number of on-screen options, meaning the number of words or letters available to the user was limited and would be best suited to a child with a small vocabulary set. Hence, the system would be more appropriate for young children with a limited vocabulary compared to older teenage children.

3. SYSTEM DEVELOPMENT

The eye tracking system was designed and developed collaboratively by Holland Bloorview Kids Rehabilitation Hospital in Toronto, Canada, and Flinders University in Adelaide, Australia. It was developed on a HP Pavillion 2000-2D23TU 15.6-inch laptop with an Intel Core i5 processor running a Microsoft Windows 7 operating system. No external hardware was required as it uses the in-built webcam of the laptop. It provides the same functionality as a tablet front-facing camera, hence it was a suitable alternative for prototyping purposes. The algorithm was developed on MATLAB r2012a with the image processing and computer vision toolboxes along with the open source computer vision library, OpenCV [6].

3.1. Eye tracking algorithm

The eye gaze estimation algorithm was based on a combination of the template-based and feature-based methods. The main steps of the eye gaze estimation program are shown in Figure 2. Each of these steps is described in detail in the following sections.

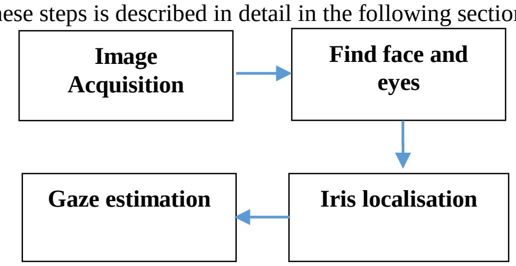

Figure 2. Eye tracking algorithm

3.2.1. Image acquisition and set-up

Images were acquired using the in-built laptop webcam that outputted 1280x720 pixel images. The webcam was positioned at chin level, with the face 30 cm away and between the vertical green lines displayed on the screen. This positioning ensured that the user's face occupied the majority of the image space. The images acquired from the webcam were resized to a 640x480 pixels image. The image was then pre-processed by cropping it to the region between the green lines, converting it to grayscale, and applying histogram equalization (Figure 3).

Figure 3. Pre-processing steps

3.2.2. Face and eye region detection

The Viola and Jones (V&J) [7] algorithm is a commonly used method for robust object detection in computer vision applications. This algorithm was used to detect both the presence of the face and eyes. Once the algorithm had detected a face, the top half of the face region of interest (ROI) was cropped and passed through the algorithm once again to detect the eyes (Figure 4). The output of the V&J detector was noisy and not appropriate for accurate tracking of the iris, hence a Kalman filter [8] was used to smooth the output of the detector.

Figure 4. Detection of face and eye region

3.2.4. Iris centre localization

The distinctiveness of the iris present in the eye region is due to the large contrast between the sclera and the iris as well as its distinct circular shape. The iris region can be found using these distinct features. The process of finding the iris region and locating the centre of the iris involved: thresholding the image based on an image histogram, combining regions of similar intensity, extracting the location and size of each region, eliminating any irrelevant regions using a binary mask, and determining the centre of the iris based on the location of the circular region identified. Samples of images of the eye region obtained from the Viola and Jones eye detector are shown in Figure 5.

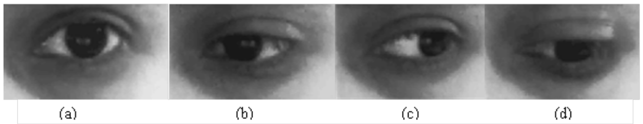

Figure 5. Grayscale images of the eye

The image was divided into six sections by multi-level thresholding using Otsu's method [9]. Figure 6 shows the dark region of the image, including the iris region, highlighted in dark blue, the skin-coloured region in light blue and yellow, and the lighter regions of the image in orange and red.

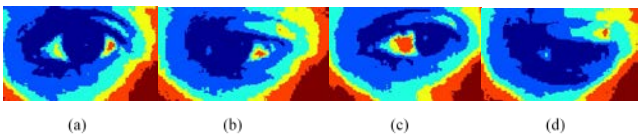

Figure 6. Multi-level Otsu thresholding

The next step involved thresholding the image so that the dark blue pixels that represent the iris were extracted into a binary image. The result was an image containing white pixels that represented the iris region as shown in Figure 7.

Figure 7. Extraction of dark blue region

As can be seen in Figure 7, unwanted pixels associated with the eyelids and dark regions around the eyes were present. These were removed using a circular binary mask with a predefined size of 20 pixels in diameter, based on the approximate size of an iris in the acquired image. This resulted in a white region shown in Figure 8, of which the centre was assumed to be the centre of the iris.

Figure 8. Iris region extracted using binary mask

3.2.5. Gaze direction estimation

The next step involved converting the iris location to a position on the screen. A calibration routine was used to store the iris centre coordinates while the user looked at different points on the screen. As shown in Figure 9, five dots were displayed sequentially on the screen in a 4x4 grid arrangement.

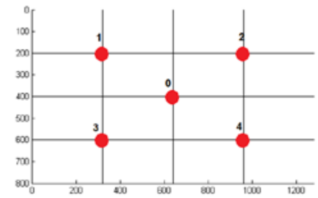

Figure 9. Calibration points

Each calibration point was associated with a mean iris centre coordinate. Every new iris centre coordinate found was compared with each of the stored calibration values to determine which square it belonged to. This was then translated into a location on the screen.

3.1 Eye tracking games

To be able to evaluate the system in an engaging manner, two simple 4x4 grid-based eye tracking games were developed. These were developed to test accuracy and usability of the eye tracking system. The two games were "Match the Tile" and "Follow the Tile" as shown in Figure 10.

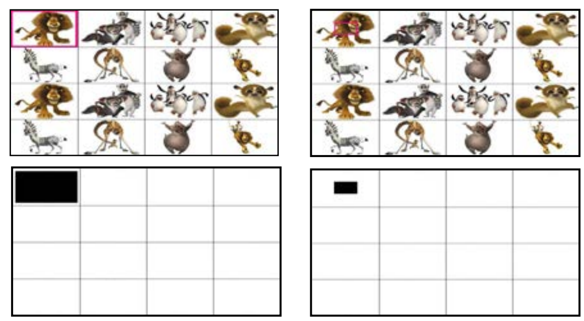

Figure 10. Match the Tile (top), Follow the Tile (bottom)

A tile on the screen could be selected if the user directed their gaze on it for 1 second. The result of a user looking at different tiles on the screen is shown in Figure 11.

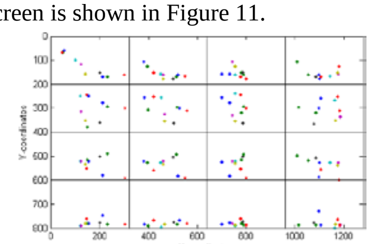

Figure 11. Gaze points on screen

3.2 Head pose estimation

During the design phase of the project, it was assumed that a future user of the system may not necessarily be able to remain still or keep their head upright. Hence, a head pose algorithm was developed to measure the roll angle of the head so that the eye tracking information could be adjusted accordingly.

The steps involved in measuring the head roll are shown in Figure 11. The first two stages of this algorithm are identical to the first two stages in the eye gaze estimation algorithm.

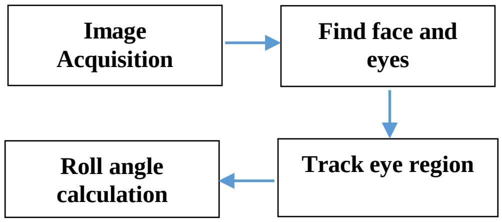

Figure 12. Head pose estimation steps

3.3.1. Eye region tracking

As the Viola and Jones detector has only been trained with frontal upright faces, it fails to detect faces when the head is rolled beyond ±45° [7]. To be able to measure head pose beyond the limits of the Viola-Jones face detector, a separate tracking algorithm using template matching was implemented to track the eye region. OpenCV's implementation of template matching was used to locate the eyes using rotated template images of the eyes as shown in Figure 12.

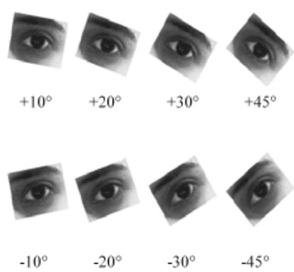

Figure 13. Rotated eye templates

A rough estimation of the location of the eyes was established with this "rotation-invariant" template matching algorithm. Once the location of the eyes was found, the head scale and roll angle was estimated using the distance and angle between both the eyes.

3.3.2. Head scale calculation

The distance between the eyes was used to estimate the scale of the head. The ratio between the initial distance between the eyes and the current distance between the eyes was calculated to estimate the scale of the head. In other words, a larger distance meant that the scale increased, i.e. user is closer to the camera.

3.3.3. Head roll calculation

The roll angle was calculated by finding the arctangent of the slope between the two eyes as illustrated in Figure 13.

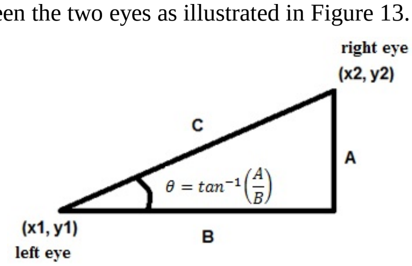

Figure 13. Calculation of head role angle

4. CONCLUSION

A cheap and portable eye tracking system that does not require any external hardware was developed to provide a pathway to communication for children with disabilities. The system is able to differentiate between different gaze points, enabling the user to control a GUI interface on a 4x4 grid. Two grid-based games were developed to test the performance of the eye tracking system. Future study is required to formally test the system performance and robustness.

Preliminary tests were conducted on the system by playing the games developed. It was found that the system was able to track the user's eyes across the 16 gaze points. However, the system was susceptible to noise caused by changes in illumination and motion. Moreover, the iris detection algorithm was susceptible to false detection of the irises caused by blinking.

Key areas for further work include enhancing the eye gaze estimation algorithm to minimize errors testing with a wide range of individuals to factor in variations due to ethnicity, eye colour, and use of spectacles, and extending the head pose estimation algorithm to calculate head yaw and pitch.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr. Simon Williams and Dr. Trent Lewis for their contributions and advice during the course of this project.

6. REFERENCES

[1] Sargent J. et al 2013, "Use of eye-pointing by children with cerebral palsy: what are we looking at?" *International Journal of Language and Communication Disorders*, vol. 48, no. 5, p477-485

[2] Li D., et al 2006, "openEyes: a low cost head mounted eye tracking solution", *ETRA '06 Proceedings of the 2006 symposium on Eye Tracking Research & applications*, p95-100, viewed 5 October 2013, <http://dl.acm.org/citation.cfm?id=1117350>

[3] Santis, A. and Iacoviello, D 2009, "Robust real time eye trackingfor computer interface for disabled people", *Computer methods and programs in biomedicine*, vol.96, issue 1, p1-11.

[4] Morimoto, C and Mimica, M 2005, "Eye gaze tracking techniques for interactive applications", *Computer vision and Image Understanding*, vol. 98, p4-24.

[5] Paul, R. et al 1991, "Communication and Socialization Skills at Ages 2 and 3 in "Late-Talking" Young Children", *Speech-Language Pathology Faculty Publications*, paper 58, Sacred Heart University

[6] Open Source Computer Vision Library (OpenCV), viewed 14 July 2014, <opencv.org>

[7] Viola, P. and Jones, M. 2004, "Robust Real-Time Face Detection", *International journal of computer vision*, vol. 57, no. 2, p137-154.

[8] Welch, G. and Bishop, G. 2006, "An introduction to the Kalman filter", *Department of Computer Science, University of North Carolina*, USA

[9] Otsu, N. 1979, "A threshold selection method from gray-level histograms", *IEEE transactions on sys-tems, man, and cybernetics*, vol. smc-9, no. 1, pp. 62