A Head Tracking Method for Improved Eye Movement Detection in Children

M. Sangi[1], B. Thompson[1,2], E. Vaghefi[1,2] and J. Turuwhenua[1,2]

[1] University of Auckland, Auckland Bioengineering Institute, Auckland, New Zealand
[2] University of Auckland, Department of Optometry and Vision Science, Auckland, New Zealand

Abstract— **The presence of untreated visual disorders in early childhood can result in abnormal visual cortex development (amblyopia). However, accurate clinical assessment of visual function in young children is highly challenging. Reflexive eye movements may allow for precise measurement of visual functions such as resolution acuity in young children if age appropriate, clinically acceptable, quantitative eye tracking techniques can be developed. Children do not tolerate chin-rests or head mounted eye-tracking equipment, therefore we have developed a method to measure and compensate for unrestrained head motion that may facilitate detection of eye movements. We implemented an automatic feature-based algorithm to track features on the face in pre-recorded videos. These data were used to "lock" the head to its initial position. Secondly, we implemented a single un-calibrated camera method to estimate the 3D movements of the head. The method was tested using video footage from five children who observed visual stimuli designed to induce horizontal optokinetic nystagmus (a reflexive sawtooth motion of the eye consisting of pursuit and saccadic eye movements). The children's heads were unrestrained, thereby exhibiting natural movement within the video. Markers placed on participants' faces were manually segmented to yield ground truth data. The standard deviation of head movement improved from (18.6676, 8.9088) to (1.8828, 1.4282) pixels after stabilization. The average mean square error (MSE) between the manual and automatic stabilization methods was 7.7494 pixels. The percentage error for 3D pose estimation was 0.2428 %. Stabilization of the eyes (relative to the head) was achieved. In conclusion, our initial results suggest that head movement stabilization is possible as a post processing step which could significantly facilitate the monitoring of eye movements in children. Furthermore automated methods could improve the monitoring of neurodevelopmental disorders that manifest through head movement.**

Keywords— **3D Head Tracking, Eye Tracking, Head Movement Disorders, Saccadic Eye Movement, Optokinetic Nystagmus.**

I. INTRODUCTION

The early detection and treatment of visual disorders can prevent developmental disorders of vision such as amblyopia. However young children typically do not possess the cognitive, attentional and language skills required by adult tests of visual function and results from pediatric vision tests can vary [1]. Optokinetic nystagmus (OKN, a reflexive stereotyped movement of the eyes in response to moving stimuli) can be quantified using a commercial eye tracker to objectively measure visual acuity in adults [2]. This technique has the potential to allow for the accurate and rapid clinical measurement of visual function in young children, however there are barriers that need to be overcome. Current eye tracking methods remain largely confined to the laboratory setting as standard systems are expensive, need careful calibration and require stabilization of the head relative to the tracking camera. The latter issue is particularly problematic in a pediatric context as children do not tolerate chin-rests or head mounted equipment. Measuring 3D head motion in world-space and removing it from eye movement in post-processing is a potential alternate solution to this problem. Furthermore, the measurement of head motion is of potential clinical value in its own right. Disorders of the neural mechanisms responsible for head stabilization and/or eye movement control may cause head tremors or involuntary, abnormal head movements [3, 4]. In addition, assessment of combined eye–head gaze shifts may assist in the diagnosis of attention-based disorders [5, 6].

In this paper we describe an automatic 3D head tracking method that utilizes a low-cost consumer grade video camera to capture and analyze head movement. We compare the outcomes of our automated algorithm with the results obtained by manually identified markers applied to participant's faces (referred to henceforth as manual tracking).

II. MATERIALS AND METHODS

An automatic video stabilization method was developed to detect and fix the position of the head (thereby stabilizing the location of the eyes) in every frame. Videos were recorded using a single uncalibrated off the shelf camera.

A. Subjects

The study was approved by The University of Auckland Human Participants Ethics Committee (reference no. 2011 066). Five children were presented a random dot kinematogram [7] stimulus (250 white dots within a 8.3° stimulus aperture moving coherently to the left or right at 8°/sec) on a cathode ray tube display. The stimulus was designed to elicit horizontal optokinetic nystagmus, a predictable saw-

J. Goh (ed.), *The 15th International Conference on Biomedical Engineering*, IFMBE Proceedings 43,
DOI: 10.1007/978-3-319-02913-9_129,

tooth movement of the eye. Video footage of children's responses to the stimuli was collected using a SONY digital high definition camera (HDR-CX7EK, Sony Corporation, Tokyo, Japan) in RGB format, with a spatial resolution of 1920 × 1080 pixels and temporal resolution of 25 frames/sec. The camera was placed beside the monitor. The children's heads were unrestrained and therefore moved within the camera's field of view.

B. *Face detection and Tracking*

The face region was detected within each video using the Viola-Jones algorithm [8] provided by the Computer Vision ToolBox MATLAB (MathWorks, Natick, VA), as well as the PittPatt SDK version 5.2.2 (acquired by Google). Robust facial features within this area were identified using the Harris corner detection algorithm [9]. These points were tracked across frames using the Kanade Lucas Tomasi (KLT) point tracker [10] available in MATLAB. A non-reflective similarity transformation (allowing rotation, scaling, and translation) was generated from tracked points between subsequent frames. The inverse transformation was then applied to compensate changes in the position of face features between subsequent frames. The face was thereby computationally "locked" in its initial position. The region around each eye was cropped in each frame, thereby isolating the eye movement from the subjects' head movements.

The automated approach was validated by re-processing the videos using manually obtained feature points corresponding to 3 or 4 markers placed on the participant's faces during video recording (fig. 2). For each frame, five corner points of each marker were selected manually, using MATLAB's ginput command. The order in which the features were selected was consistent across frames to maintain correspondence of features from frame to frame. Because the manual point selection was noisy, a Kalman filter [11] was used to smooth the feature tracking. Four references points were inserted into each recording, and then tracked across all video frames. Superimposing these points allowed differences between manual and automated stabilization to be calculated. Coordinates of these points were compared in manually and automatically stabilized frames.

C. *3D Pose Estimation*

The POSIT algorithm [12] was used to estimate the 3D position and pose of the head. At least four 3D model points (world-space coordinates) and corresponding 2D image points, focal length and principal point of camera were needed for this estimation.

The 3D model points' coordinates (the marker corners) were fixed and the corresponding 2D image points were identified. The camera properties were estimated using equations (1) - (3).

$$f_x = f_y = \frac{\frac{w}{2}}{\tan\left(\frac{120}{2} \times \frac{\pi}{180}\right)} \tag{1}$$

$$c_x = \frac{w}{2} \tag{2}$$

$$c_y = \frac{h}{2} \tag{3}$$

Here f_x and f_y are the focal lengths in *x* and *y* directions respectively (assumed to be equal), expressed in pixel units. The image dimensions are given by (*w*, *h*), the principal point of the camera by (c_x, c_y). Rodrigues' rotation formula [13] was used to retrieve rotations for X, Y, Z directions respectively.

The POSIT error assessment was performed by re-projecting 3D model points to the 2D image plane by the transformation presented in equation (4) [14].

$$s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

where *s* is the pixel size, *x* and *y* are 2D image points X, Y and Z are 3D model coordinates, r_{ij} are rotation parameters and T_x, T_y and T_z are translations. We used the percentage error between the original and reprojected image points to assess errors due to POSIT.

III. Results

The faces were localized to (523.8037, 253.3558) ± (18.6676, 8.9088) pixels prior to stabilization and (500.4114, 260.0309) ± (1.8828, 1.4282) after stabilization (fig. 1). The images were able to be consistently cropped as a result. The average MSE between automatic and manual stabilization was 7.7494 pixels. Percentage error of 3D pose estimation was 0.2428 %.

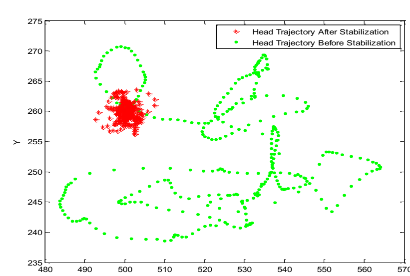

Fig. 1 Head trajectory before (green-dot) and after (red-star) stabilization. Units are in pixels.

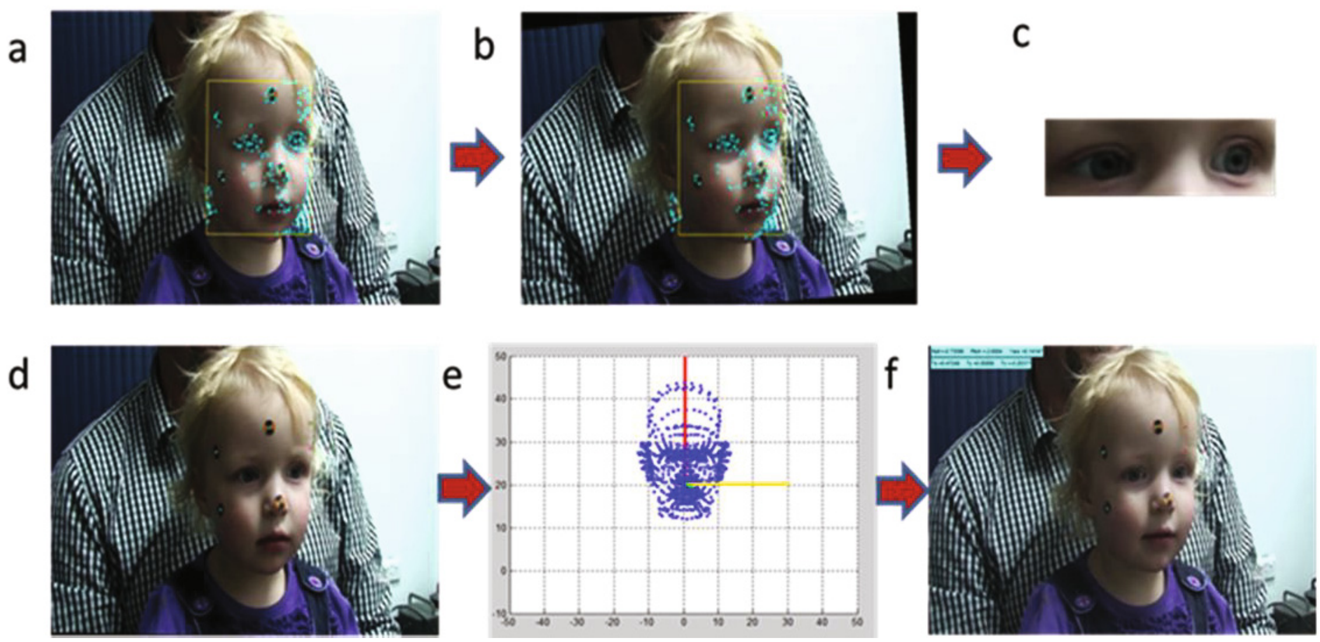

Fig. 2 The head 3D pose estimation and stabilization procedure. (a) Harris corners were located in the automatically detected face region, (b) a stabilized video with the eye region "locked" to the center of each frame, (c) the region around the eyes was cropped, (d) markers with manually selected center points, (e) 3D pose estimation was determined by the POSIT algorithm (units are in centimeters), (f) re-projection of the manually selected marker points using the POSIT algorithm showing estimated head pose. (Four red references points are used to visualize comparison of manual and automatic stabilization). Images reproduced with parental consent.

IV. Discussion

Our preliminary results suggest that simple automated methods are sufficient to keep the eye region centered in video footage of children observing visual stimuli with unrestrained heads. The head motions observed in our study were consistently tracked and compensated for using the methods we developed.

We found that automated feature selection produced similar results to manual feature selection (MSE was 7.7494 pixels), indicating that a more labour intensive approach involving markers, although robust, may not be necessary.

We extracted 3D head pose and position using the POSIT algorithm (fig. 3). Our testing indicated that the approach modelled the given data with a low percentage of error. It is noted that the 3D pose and positional information we obtained, whilst naturalistic, was not validated objectively. Our own testing suggests that the approach could yield useful quantitative kinematic information, and this is the subject of further work.

The approach we have developed warrants further investigation. A larger cohort of participants will be required for additional validation, and the limitations of the approach need to be more fully investigated.

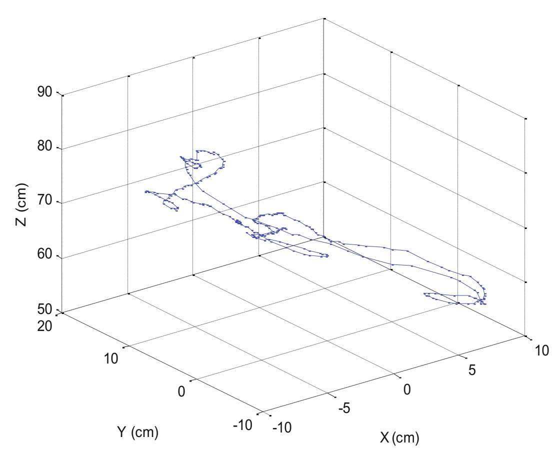

Fig. 3 3D head trajectory estimated by POSIT

V. CONCLUSION

We have developed software to reduce movements of the head in order to improve the visualization of eye movements in videos of children. Our initial results in five participants indicate that simple automated methods are sufficient for this purpose. We applied the POSIT algorithm to extract 3D head position and pose information. Our results in this regard lay a foundation for further investigations regarding the role of head movements in disease, and their interactions with eye movement. Further investigation of these methods is warranted.

ACKNOWLEDGMENTS

The authors would like to thank Sandy Yu for her help in video recording and data collection.

REFERENCES

Anstice, N S, Thompson, B (2013) The measurement of visual acuity in children: an evidence-based update. CLIN EXP OPTOM. In press DOI: 10.1111/cxo.12086

1. Hyon, J Y, Yeo, H E, Seo, J M, Lee, I B, Lee, J H, Hwang, J M (2010) Objective measurement of distance visual acuity determined by computerized optokinetic nystagmus test. INVEST OPHTH VIS SCI 51:752-757
2. Keller, S, Dure, L S (2009) Tremor in childhood. SEMIN PEDIAT NEURO 16:60-70
3. Taylor, D (1980) Disorders of head and eye movements in children. TRANS OPHTH SOC 100:489-494
4. Khan, A Z, Blohm, G, McPeek, R M, Lefèvre, P (2009) Differential influence of attention on gaze and head movements. J NEUROPHYSIOL 101:198-206
5. Cicchini, G. M., Valsecchi, M, De'Sperati, C. (2008). Head movements modulate visual responsiveness in the absence of gaze shifts. Neuroreport, 19(8), 831-834.
6. De Bruyn, B, Orban, G A (1988) Human velocity and direction discrimination measured with random dot patterns. VISION RES 28:1323-1335
7. Viola, P, Jones, M J (2004) Robust real-time face detection. INT J COMPUT VISION 57:137-154
8. Harris, C, Stephens, M. (1988) A combined corner and edge detector. Alvey vision conference 15: 50).
9. Tomasi, C, Kanade, T (1991) Detection and tracking of point features. School of Computer Science, Carnegie Mellon Univ
10. Welch, G, Bishop, G (1995) An introduction to the Kalman filter.
11. DeMenthon, D F, Davis, L S (1992). Model-based object pose in 25 lines of code. Computer Vision—ECCV Springer Berlin Heidelberg 1992, pp 335-343
12. Mebius, J E (2007) Derivation of the Euler-Rodrigues formula for three-dimensional rotations from the general formula for four-dimensional rotations. *arXiv preprint math/0701759*.
13. Hartley, R, Zisserman, A (2000). Multiple view geometry in computer vision. Cambridge University Press, Cambridge

Author: Jason Turuwhenua
Institute: University of Auckland
Street: 70 Symonds
City: Auckland
Country: New Zealand
Email: j.turuwhenua@auckland.ac.nz