

Received 4 December 2014; revised 16 February 2015; accepted 24 February 2015. Date of publication 5 March 2015; date of current version 24 March 2015.

Digital Object Identifier 10.1109/JTEHM.2015.2410286

An Optokinetic Nystagmus Detection Method for Use With Young Children

MEHRDAD SANGI¹, (Student Member, IEEE), BENJAMIN THOMPSON²,
AND JASON TURUWHENUA¹

¹Auckland Bioengineering Institute, University of Auckland, Auckland 1010, New Zealand

²Department of Optometry and Vision Science, University of Auckland, Auckland 1010, New Zealand

CORRESPONDING AUTHOR: J. TURUWHENUA (j.turuwhenua@auckland.ac.nz)

This work was supported in part by the Auckland Bioengineering Institute and in part by the University of Auckland Faculty Research Development Fund under Grant 1165695.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

ABSTRACT The detection of vision problems in early childhood can prevent neurodevelopmental disorders such as amblyopia. However, accurate clinical assessment of visual function in young children is challenging. optokinetic nystagmus (OKN) is a reflexive sawtooth motion of the eye that occurs in response to drifting stimuli, that may allow for objective measurement of visual function in young children if appropriate child-friendly eye tracking techniques are available. In this paper, we present offline tools to detect the presence and direction of the optokinetic reflex in children using consumer grade video equipment. Our methods are tested on video footage of children ($N = 5$ children and 20 trials) taken as they freely observed visual stimuli that induced horizontal OKN. Using results from an experienced observer as a baseline, we found the sensitivity and specificity of our OKN detection method to be 89.13% and 98.54%, respectively, across all trials. Our OKN detection results also compared well (85%) with results obtained from a clinically trained assessor. In conclusion, our results suggest that OKN presence and direction can be measured objectively in children using consumer grade equipment, and readily implementable algorithms.

INDEX TERMS Eye tracking, head tracking, optokinetic nystagmus, pupil/iris detection, video stabilization.

I. INTRODUCTION

The early detection of visual problems is beneficial in both arresting and preventing disorders such as amblyopia. Unfortunately, adult test of visual function are unsuitable as they require cognitive, attentional and language skills that children do not typically possess. Furthermore results of existing pediatric vision tests can be variable [1]. Therefore, tools and methods for the accurate assessment of visual function using objective approaches in this age-group are desirable. The detection of optokinetic nystagmus (OKN) offers a potential solution to this problem. The optokinetic response is an involuntary sawtooth movement of the eye that occurs in response to moving stimuli such as a rotating drum, or drifting bars on a computer screen [2]. The response consists of an alternating sequence of slow phases (SPs) during which the eyes track a feature of the moving stimulus, and quick phases (QPs) where the eyes move rapidly in the opposite direction to the moving stimulus [2], [3] (see Fig. 1). As the response is reflexive, the presence or absence of OKN indicates whether

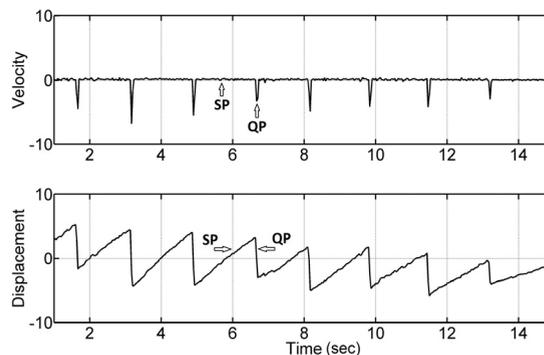


FIGURE 1. Sample of OKN velocity and displacement signal with indicated slow phase (SP) and quick phase (QP).

or not a moving stimulus was visible, without the explicit cooperation of the observer. Determining the presence/absence of the optokinetic response is a recognized technique for the *subjective* assessment of visual function that has been used for both adults [4] and young children [5], [6].

However, in children, the additional challenge is to ensure that measurements are as robust and accurate as possible, whilst being non-invasive, quick, and allowing unrestrained head movement. Computer aided measurement and analysis may facilitate the controlled, accurate and rapid clinical measurement of visual function in this age group [7].

Camera based eye tracking is a non-invasive computerized approach in which the eye signal is extracted from (typically) infrared (IR) video [8]–[13]. Recently, Hyon *et al.* [14] and Han *et al.* [15] used such a system for the *objective* assessment of OKN. Distance visual acuity was measured objectively by this method yielding good reproducibility and significant correlation between distance visual acuity measured using a standard eye chart and objective visual acuity measured using OKN. Shin *et al.* [16] continued this work, describing the relationship between OKN response and visual acuity in adult humans with ocular disease using the same approach. They showed that in subjects with 20/60 or worse visual acuities, objective visual acuity estimated by OKN can predict subjective visual acuity.

Unfortunately, current generation eye tracking systems (such as the setup described by Hyon *et al.* [14]) generally employ chin rests or head-mounted cameras to account for head movement. Furthermore, precise calibration procedures are necessary before eye position data can be acquired. The requirement for high levels of patient cooperation makes such systems unsuitable for use with young children. Companies such as Tobii (Tobii Technology) and SMI (SensoMotoric Instruments) provide remote eye tracking (heads free) solutions, but remain at present largely confined to the research setting as they are expensive and expect careful calibration which is difficult to achieve with young children. Moreover, these systems do not readily compensate for rapid and large head motions that often occur when a child views visual stimuli with an unrestrained head [8], [9]. Therefore, subjective assessment of eye movement remains the current standard for OKN detection in children [1].

A number of related systems for measuring eye movements (with free head movement) have also been described. Eizenman [17], [18] and Model [19] proposed point gaze estimation using Purkinje images including methods for calibration with infants (6-, 8-, 9-, 15-, and 16 month-old). Zhu and Ji [20] used a method to evaluate gaze with a table mounted stereoscopic system without using any user-dependent eyeball parameters. The system compensated head movements and gave an estimation of the gaze direction with respect to a reference head position. To estimate 3D configuration of cameras, lights, screen, head and eyes, the stereo vision system was calibrated a priori. However, these methods require a customized setup and are not readily available in the clinical setting.

In this study, we report offline methods that were developed to *retrospectively* analyze videos collected from children as they underwent an assessment in which they viewed drifting stimulus patterns. Our aim in doing this was to test whether a low-cost recording setup, using

semi-automated approaches could replicate the assessment of an experienced clinical assessor who judged the direction of the quick phase eye movements in the video footage. The authors are not aware of similar studies to date, in which attempts have been made to automate the measurement of the optokinetic response in children using consumer grade video equipment in particular.

In this work we present simple to implement techniques for detecting the absence/presence of OKN, using footage obtained from a consumer grade video recorder. The work includes the novel application of known techniques to the issue of OKN detection in young children, as well as the implementation of a new feature based stabilization technique which allows for precise head stabilization. In particular, we describe methods that extract the motion of the eyes with respect to the head, and processing of that signal in order to estimate the presence or absence of OKN. Two methods for extracting the eye movement signal from video are proposed: (1) a method based on stabilization of the head and subsequent eye motion analysis (the *stabilization method*), and (2) direct subtraction of the head signal from the combined head and eye signal (the *direct subtraction method*). We compare these methods with (3) Visage SDK, a commercial grade head and eye tracking system. The performance of the head stabilization process is further assessed by comparison with manual tracking of markers, in the form of stickers that the participants wore on their faces.

II. METHOD

We developed readily usable tools that would facilitate eye tracking (in the frame of the head) in RGB video. Fig. 2 shows the entire procedure that we adopted. The steps involved are head tracking, head stabilization and pupil center and limbus edge tracking. These are explained in the following sub-sections.

A. HEAD TRACKING AND STABILIZATION METHOD

Video stabilization is the process of removing camera or unwanted movement within the scene [21]. Recently Bai *et al.* [22], [23] described a method that tracked general features (or specific features within a region of interest) and warped the video frames to remove unwanted movement. Kinsman *et al.* [24] reported a technique using head mounted cameras, which compensated head motion, in order to improve eye fixation detection in a wearable device. That algorithm was based on maximizing cross correlation between two consecutive frames, and was used to estimate the global motion of the scene resulting from camera movement. Rothkopf and Pelz developed a method that compensated rotational motion of the head by using an omnidirectional camera in a wearable device [25]. We implemented an approach to stabilize head movement in which Shi-Tomasi Features [26] were detected in a mask, and then tracked (using the Kanade Lucas Tomasi (KLT) point tracking [27]) to obtain the transformation required to stabilize the video.

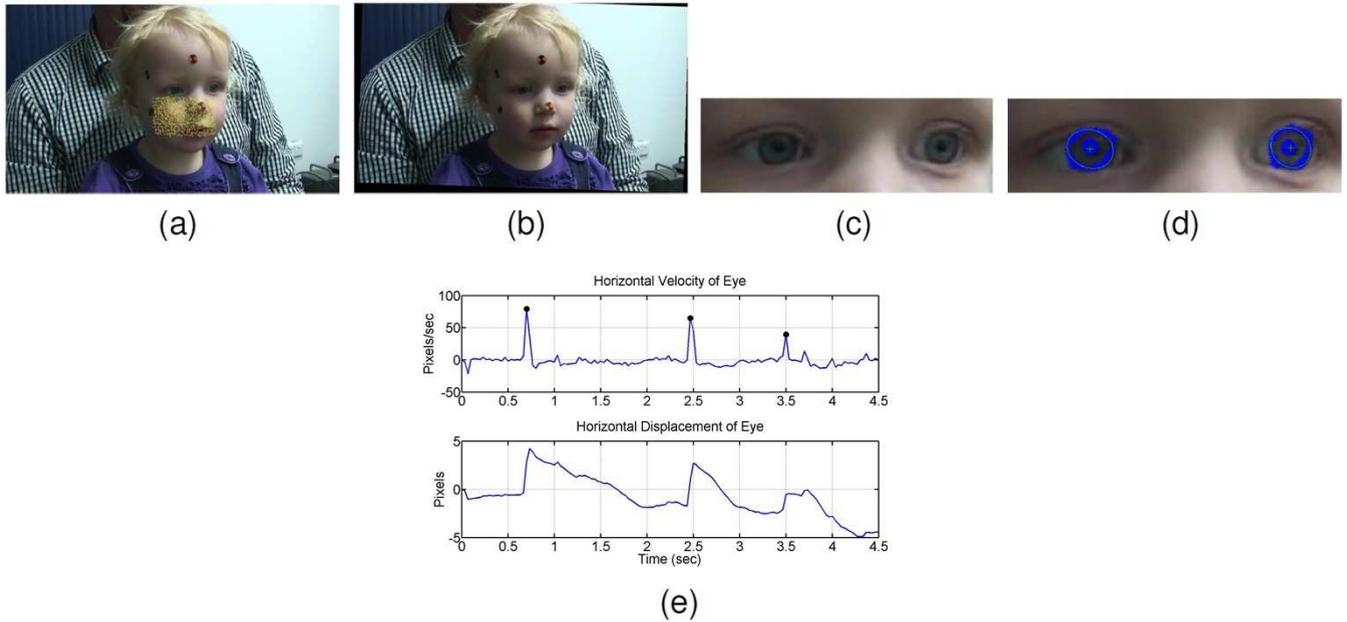


FIGURE 2. (a) Shi-Tomasi features detected and tracked on the face region. Features in the region of markers are removed. (b) Stabilized face in which the video is now transformed with the head fixed. (c) Eye region cropped. (d) Detected pupil center, limbus edge and Starburst features. (e) Resulted eye velocity and displacement signal.

1) HEAD TRACKING

A semi-automated method was implemented to enable head motion tracking. An initial face mask was manually selected within the facial region in a way that excluded eye regions. By doing this we sought to avoid contaminating the head signal with eye movement information. Shi-Tomasi features were detected in the initial frame, within the selected face region only. These features were tracked across frames with the Kanade Lucas Tomasi (KLT) point tracker. During our testing we observed that a proportion of features (around 6%) were detected in the region of the manually applied markers (i.e., the stickers on the participant's faces). As we did not intend to use these markers as a tool for facilitating face feature detection, we removed these features from the analysis. We found that sufficient features remained for the tracking step of the algorithm to be effective. Fig. 2(a) shows the remaining features.

2) HEAD STABILIZATION

A non-reflective similarity transformation allowing rotation, scaling, and translation was estimated by Maximum Likelihood Estimation Sample Consensus (MLESAC) [28] by using the Shi-Tomasi feature points generated in the current and previous frames. MLESAC is a generalized version of the Random Sample Consensus (RANSAC) algorithm [29]. This algorithm extends upon the RANSAC method by maximizing the likelihood of a solution based on the outliers' and inliers' distribution.

$$\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = H_j \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (1)$$

$$H_j = \begin{bmatrix} s_j r_{1j} & -s_j r_{2j} & T_{xj} \\ s_j r_{2j} & s_j r_{1j} & T_{yj} \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Here (x_i, y_i) and (x'_i, y'_i) are feature points in the previous and current frame respectively, s_j is the scaling factor, T_{xj} , T_{yj} and r_{1j} , r_{2j} are translation and rotation components between the previous and current frames respectively.

$$H_{Total} = H_j \times \dots \times H_1 \quad (3)$$

Using (3) the total transformation H_{Total} between the current frame and the first frame was calculated, thereby allowing registration of the face to its position in the initial frame by way of the inverse transformation (H^{-1}_{Total}). The processing pipeline for the stabilization method is shown in Fig. 3(a). Our semi-automatic stabilization method was validated using two approaches. Firstly, the results were compared to the stabilization achieved by replacing the feature selection and tracking with manual selection (across frames) of the four markers placed on the participant's face during video recording (Fig. 2(a)). For each frame, four corner points and the center of each marker were selected manually. The order in which the features were selected was consistent across frames to maintain correspondence of features from frame to frame [30]. Secondly, the results of the stabilization method were compared to those obtained from a commercial grade head and eye tracking product (Visage SDK, Visage Technologies). In this case Visage SDK detected facial landmarks that were tracked by the SDK.

3) PUPIL CENTER AND LIMBAL EDGE TRACKING

Given the head transformation, we next sought to determine the motion of the eye, with respect to the estimated head motion. Eye region bounding box was selected manually in

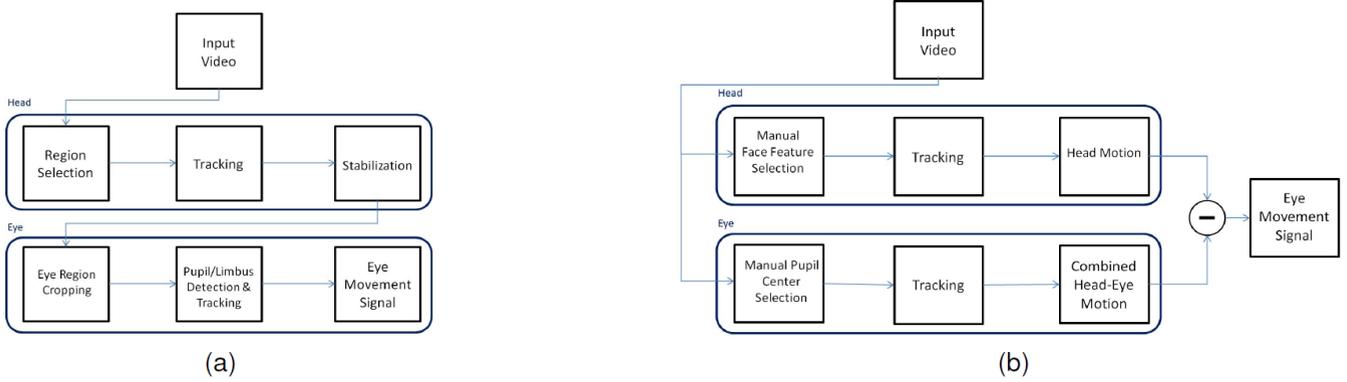


FIGURE 3. (a) Work flow of stabilization method. (b) Work flow of direct subtraction method.

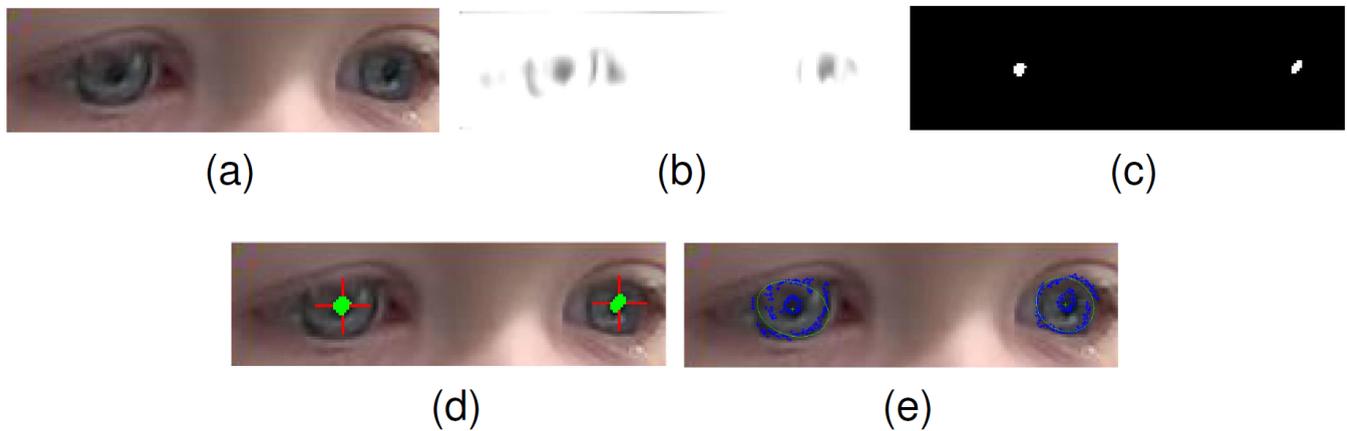


FIGURE 4. (a) Original Frame. (b) Processed by Gabor filter. (c) After thresholding and morphological operation. (d) Detected Pupil region and center. (e) Detected limbal edge by Starburst method.

the first frame of the stabilized video. This bounding box was then applied to crop the eye region across all subsequent (stabilized) frames. In this way the eye region was isolated from the head movement thereby serving a simplified analysis.

A number of methods have been described for pupil/iris detection and tracking including edge detection and gradient based methods [31]–[33], deformable models based methods [34]–[36], and machine learning algorithms [37]–[40]. Most of these methods are not suitable for use with standard RGB video footage and they are often computationally intensive, particularly those based on deformable models and machine learning algorithms. In this work we used a method utilizing simple Gabor filters to locate the pupil centroid, which was followed by the Starburst algorithm [41], [42] to locate limbus edge features.

The pupil was located using multiple oriented (θ) Gabor filters [43] applied to the image.

$$g(x, y; \lambda, \theta, \phi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \quad (4)$$

where $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$. Here λ is the number of cycles/pixel (wavelength), θ is the angle of the normal to the sinusoid (orientation), ϕ is the offset of the sinusoid (phase), and γ is the spatial aspect ratio. The spatial envelope of the Gaussian is controlled by the bandwidth. After thresholding, standard morphological operations were used to find the centroid of the pupil region (see Fig. 4(b)-(d)).

As mentioned, once this was determined, we used the readily available, Starburst algorithm [41] to find limbus edge features. We found that it was fast and worked well on our RGB video footage. This approach tests derivative along radial rays extending away from pupil center until a threshold is exceeded. A distance filter is used to remove outlier features. The starting point is then replaced with the geometric center of the remaining features and the filtering repeated again. An ellipse can be fitted to the limbus feature points using RANSAC algorithm [29]. (see Fig. 4(e)) KLT tracked the detected pupil centers and limbus edge features in every frame. This procedure was restarted when tracking was lost or a blink occurred. A simple automatic blink detection was developed based on pupil area measurement. When a

blink occurred, the pupil area decreased to zero (when the eye lid closed completely) and then started increasing to full size again. Analysis of the resulting pupil area signal was used to identify the entire blink duration [44], [45]. Horizontal velocity of the eye $V_x(t)$ in the frame of the face was estimated by finding the velocity of the tracked pupil centers and limbus edge in the stabilized image.

B. DIRECT SUBTRACTION METHOD

Fig. 3(b) shows the work flow for the *direct subtraction method*. For this method an interactive tool was developed in MATLAB (Natick, MA: The Math Works Inc.), to initiate and track head and eye movements. Points were selected on the face and pupil centers to initialize tracking. Pupil tracking was used as it is a standard method of eye tracking. These points were then tracked (using KLT), from which the head and eye velocities could be calculated separately. In particular, the eye velocity was the combination of head and eye motion (eye in space) because head movement (head in space) had not been compensated for [46].

$$V_{eye} = V_{head-eye} - V_{head} \quad (5)$$

Pure eye motion (eye in head) was calculated from (5). There was no need for frame transformation in this method, therefore yielding a less computationally expensive method. On the other hand, it did not provide a convenient visualization of the eye region as per the *stabilization method*.

We processed video frames again with Visage SDK, using the SDK's ability to detect facial landmarks (including pupil centers) and track movements of the head and eye robustly. Direct measurement of the eye movement (eye in head) was determined by subtracting the head motion (head in space) from the eye signal (eye in space), which comprised of mixed head and eye movement. Fig. 5 shows landmarks identified by Visage SDK.



FIGURE 5. Facial landmarks detected by Visage SDK.

C. OKN DETECTION

In this work we used the eye velocity signal to estimate the presence of OKN [46]. After obtaining the velocity signal $V_x(t)$ by three different methods (*stabilization*, *direct subtraction* and Visage SDK), we used a recently proposed

OKN detection approach proposed by Turuwhenua et al. [47] in which quick phases fitting heuristic criteria were averaged, and then result used to identify the presence and direction of OKN. All peaks within the eye velocity signal were detected and thresholded by an empirical value. Peaks which were less than a given number of frames apart (obtained empirically based on stimulus velocity) were rejected. We assumed that an isolated peak was not enough evidence to indicate the presence of OKN. Therefore peaks with no other maxima of the same sign in the trial were rejected. The resulting peaks were averaged and scaled using (6) to calculate an OKN consistency value \bar{K} .

$$\bar{K} = \begin{cases} \frac{1}{N\sigma_v} \sum_{j=1}^N Q(j) & N > 0 \\ 0 & N = 0 \end{cases} \quad (6)$$

Where $Q(j)$ are peak velocities, N is the number of peaks and σ_v is an empirical threshold. All peaks were classified to positive and negative peaks. Values of \bar{K} of greater than 1 indicated the presence of OKN and less than 1 meant no consistent OKN was detected. The sign of \bar{K} indicated the OKN direction. It is noted that in Turuwhenua et al.'s work, the velocity signal was determined using a method based on optical flow of the limbus. In this application we applied the algorithm to the velocity signal obtained by our pupil center and limbus edge feature tracking (using KLT) approach. For validation purposes, the quick phase periods of OKN were identified by an experienced observer viewing the same videos but masked to the results of the detection algorithm. The observer annotated a range of frames in which quick phases appeared to be present (using QuickTime 7.6.6 on a *MacBook Pro (15-inch)*). This observer had three years of experience in viewing and interpreting OKN in adults and children. From this analysis we sought to determine the method's sensitivity, specificity and predictive values (positive/negative). Secondly a clinically trained assessor observed the videos and decided the direction (right/left) and presence of OKN (yes/no) in each trial. The assessor was an optometrist with over three years of experience in recording and assessing OKN responses in young children and was also masked to the results of the detection algorithm. Rather than focusing only on the quick phase eye movements (which was what the experienced observer was instructed to do). The clinically trained assessor made a holistic judgment of OKN presence and direction based on both quick and slow phase eye movements and clinical experience.

D. EXPERIMENT

1) ETHICS STATEMENT

The study was approved by The University of Auckland Human Participants Ethics Committee (reference no. 2011 066).

2) PARTICIPANTS

Five children (two female and three male) with normal vision (ages = 21-25 months) participated in this study.

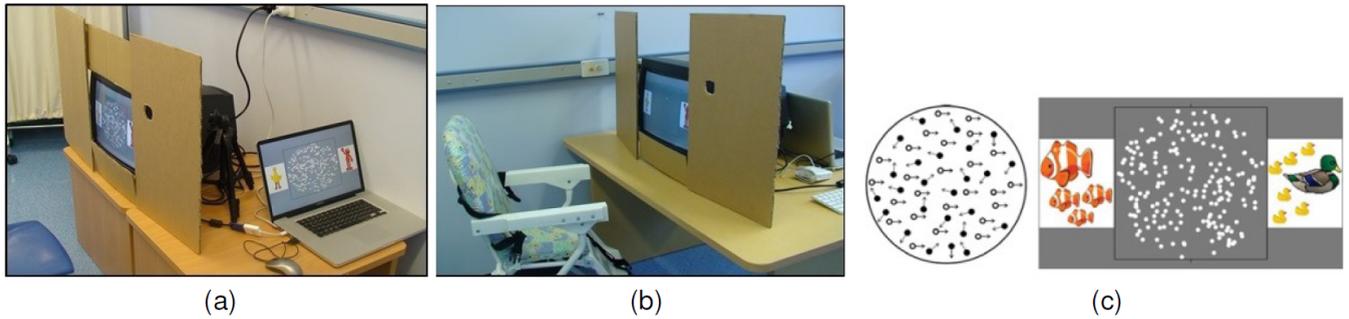


FIGURE 6. (a) and (b) Experiment setup. (c) A random dot kinematogram stimulus (white dots are signal dots which all moving in same direction, black dots are noise dots which moving in random direction).

3) VISUAL STIMULI

The visual stimuli used to elicit OKN were a random dot kinematograms (RDKs) presented on a cathode ray tube monitor (Dell E772p, 1280 × 1024 resolution, 60Hz refresh rate) placed 60cm from the observer. The stimulus was designed to elicit horizontal OKN. A program written in MATLAB using the Psychtoolbox on a Macbook Pro (15-Inch) was used to generate the stimuli. The RDK was presented within a circular stimulus aperture with a radius of 8.3° and was made up of 250 white dots (138cd/m²) presented on a grey background (42cd/m²) (dot density = 1.16 dots per degree). Each dot had a diameter of 0.5°, a speed of 8°/second (generated by displacing each dot by 0.13° on every frame) and a limited lifetime defined as a 5% chance of disappearing on each frame and being redrawn in a random location. The duration of each RDK presentation was 8 sec. The noise dots had a constant speed and direction. Signal dot direction (left or right) was randomized for each trial. Coherence levels (the proportion of signal motion to noise motion in the stimuli) were in the range of 84% to 100% and leftward vs. rightward motion was randomized across trials. See [48] for further details. These stimuli were designed to measure global motion perception in young children [48], and the parameters were therefore selected to optimize global motion processing, at the expense of an optimal OKN response. These were useful stimuli for our purposes as they provided a conservative test of the performance of our system.

4) PROCEDURE

Video footage was collected and our methods were developed to retrospectively analyze the results. Video format, with a spatial resolution of 1440 × 1080 pixels and temporal resolution of 25 frames/sec. The camera was placed beside the monitor. The children's heads were unrestrained and therefore exhibited a natural range of movements within the cameras field of view. To evaluate accuracy and performance of head tracking method, markers were added to participants face where possible. Fig. 6 shows the experimental setup.

20 trials were collected in total. Prior to the measurement, training was conducted. The child was asked to follow a single dot with a finger. Directly prior to the presentation of

TABLE 1. Standard deviation of tracked feature location before and after stabilization with three different methods.

Method	Feature location before stabilization(std in pixels)	Feature location after stabilization(std in pixels)
Proposed Method	(27.4018, 7.1702)	(0.3072, 0.2766)
Manual	(29.8755, 6.9806)	(0.3670, 0.3451)
Visage SDK	(28.9918, 7.6997)	(3.0691, 1.3302)

an RDK, a flashing fixation point accompanied by a beeping sound was presented in the center of the screen to attract the child's attention. Each trial was then initiated by the experimenter when the child was looking at the screen. The characteristics of OKN vary depending on whether observers are asked to simply look at the moving visual stimulus or stare at a central point within the display. Although we were unable to instruct the children whether to look or stare at the screen due to their young age, it is most likely that our procedure generated "look OKN" rather than "stare OKN" [49].

III. RESULTS

A. STABILIZATION RESULTS

Table 1 compiles the standard deviations (σ_x, σ_y) of a tracked feature location (within the face region) before and after stabilization, using our proposed method, manual selection of marker stickers on the face and Visage SDK. The results indicate a dramatically decreased standard deviation after stabilization for all three methods. The final error was roughly equivalent between the proposed method and the manual selection approach stabilization (the gold standard), but both methods improved upon results obtained by Visage SDK. An example of head trajectory for one of the subjects before and after stabilization is shown in Fig. 7.

B. DIRECT SUBTRACTION RESULTS

Eye velocity signal produced by the *stabilization method* and the *direct subtraction method* had a low error and high correlation (MSE = 0.197 ± 0.0474 pixels/frame,

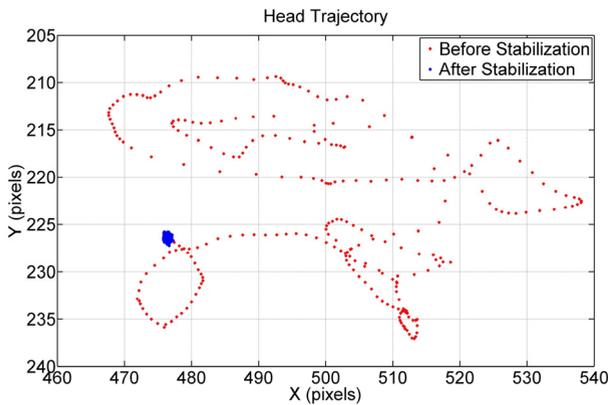


FIGURE 7. Head trajectory before (red-plus) and after (blue-star) stabilization with proposed method.

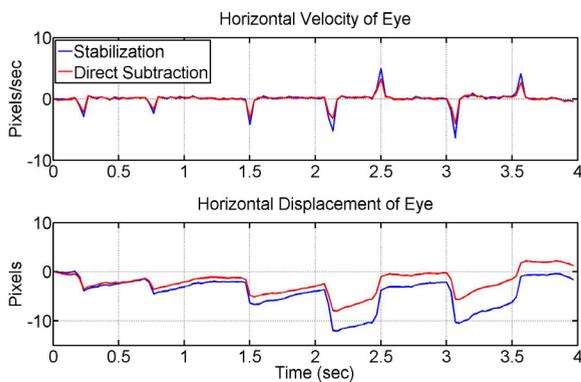


FIGURE 8. Horizontal velocity and displacement of pupil center by *direct subtraction method* (red), *stabilization method* (blue).

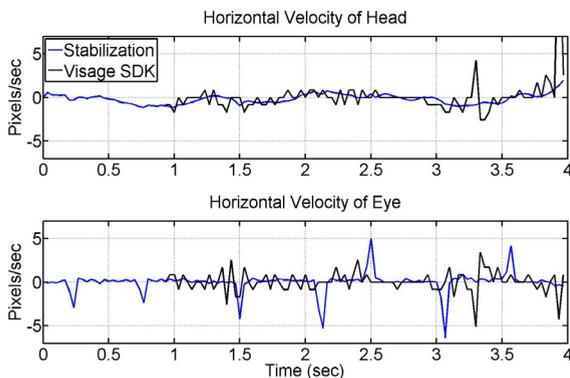


FIGURE 9. Horizontal velocity of head and eye measured by *stabilization method* (blue) and by Visage SDK (black). Visage SDK signal does not start from the beginning because it takes time for Visage SDK to initialize tracking.

correlation = $95.94\% \pm 3.25\%$) over all children. Fig. 8 shows the OKN velocity and displacement signal before (with *direct subtraction method*) and after stabilization. Fig. 9 shows the horizontal velocity of the head and pupil measured by the Visage SDK in comparison with the *stabilization method*. We found that the Visage SDK required a certain duration of video to initialize tracking (12 frames on average), as it attempted to find the best frame

for facial landmark detection. Hence the figures show a period in which no data was collected. The pupil center signal was noisy in comparison to our methods.

C. OKN DETECTION RESULTS

Our experiment in which an experienced observer identified OKN-like movements in video revealed a sensitivity and specificity for our OKN peak detection of 89.13% and 98.54% respectively across all children in *stabilization method* (true positives = 82, true negatives = 2094, false positives = 31, false negatives = 10). The positive predictive value and negative predictive value was 72.57% and 99.52% respectively.

By comparing the presence or absence of OKN indicated by the value with the trial by trial yes/no responses of the clinically trained observer, we found that the OKN detection algorithm agreed with the *subjective* assessment in 17 of 20 trials. The direction of OKN indicated by the \bar{K} value correlated exactly with that identified by both the experienced observer and the clinically trained assessor. OKN and OKN-like detection results were exactly the same for pupil center and limbus edge tracking methods.

Fig. 10 shows examples of the horizontal velocity and displacement of the pupil with periods of OKN. Red regions indicate periods of OKN identified by the experienced observer, whilst the computer detected peaks are shown with black dots.

IV. DISCUSSION

The assessment of pediatric visual function in the clinic poses significant challenges. We have presented, for the first time, offline tools and methods for the purpose of analyzing OKN in children, using readily available algorithms and consumer grade equipment. Our key additional contributions are: (1) a method for stabilizing the head based on random facial feature detection (without using specific facial landmarks such as eye corners, nose tip, lips corners and etc.), and (2) a demonstration of a method able to estimate the presence and direction of the optokinetic reflex from the stabilized video footage of young children with unrestrained heads.

The head *stabilization method* was compared against alternate methods based on: (2) *direct subtraction* of the head signal from the eye signal (combined head and eye motion), and (3) a commercially available head tracking and eye gaze estimation system (Visage SDK). Our head *stabilization method* produced slightly lower error than a gold standard method in which markers placed on the face were manually tracked (the difference in error between the two was sub-pixel). Both of these methods out-performed the Visage SDK which generated errors that would have interfered with the detection of OKN. The inclusion of stabilization in our processing also facilitated the detection of the pupil center, and *subjective* assessment of eye movements, as reported by an experienced observer. We believe that head stabilization is a valuable tool in the analysis of OKN-like movements.

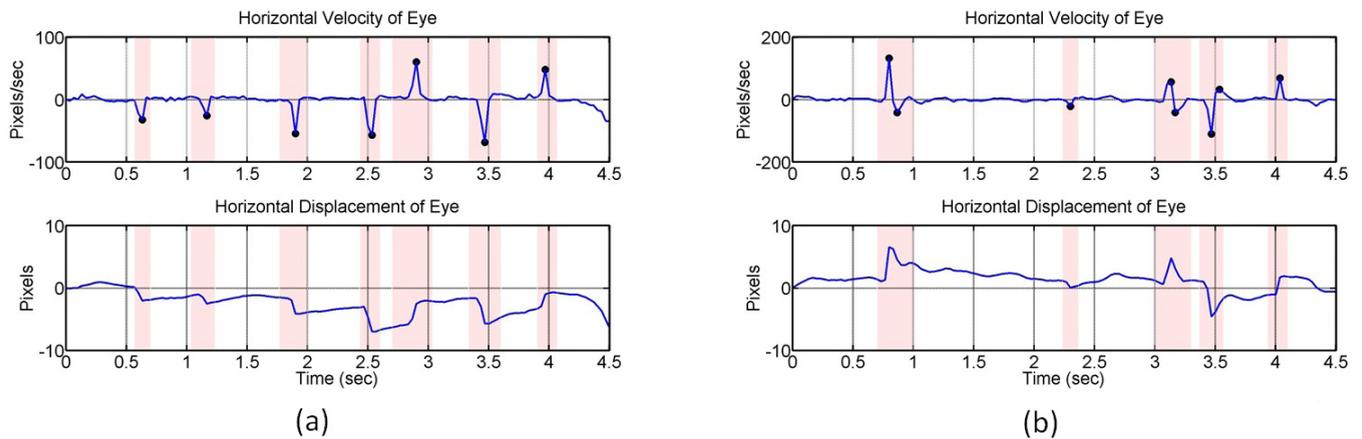


FIGURE 10. Horizontal velocity and displacement of pupil center (pixels/sec) with detected peaks (black-balls) and with OKN presence regions identified by experienced observer (red bars). (a) A left direction OKN, $=-1.38$. (b) Not consistent OKN, $=0.5$.

We found a high correlation ($MSE = 0.099 \pm 0.0574$ pixels/frame, correlation = $95.94\% \pm 3.25\%$) between the eye signals obtained using the *stabilization method* and the *direct subtraction method*. The peaks of the eye velocity signal were matched exactly for both methods and hence there was no difference in OKN detection results obtained using either method. The *stabilization method* was as effective as the *direct subtraction method* for the robust detection of OKN. Our results indicated no meaningful difference between the pupil center and limbus edge approaches ($MSE = 0.0512 \pm 0.009$ pixels/frame, correlation = $96.28\% \pm 2.156\%$).

The OKN detection results correlated highly with an experienced observer (17 of 20 trials, correlation = 85%). The direction of OKN was easily determined by finding the OKN consistency sign and it was identical to the assessment of a clinically trained assessor who viewed the same video footage. The three trials for which the OKN detection results did not correlate with human observers were characterized by very large head and body movements ($N = 2$ trials) and the presence of a care-giver's hand in the frame which temporarily obscured the face and eyes ($N = 1$ trial). Under these conditions, the OKN detection results indicated no OKN whereas the human observers reported the presence of OKN.

Our proof-of-concept analysis suggests that the techniques described here could form the central component of a vision screening tool. Specifically, the OKN detection algorithms could be combined with the presentation of visual stimuli to measure acuity and contrast sensitivity in young children.

V. FUTURE WORK AND IMPROVEMENTS

The *stabilization method* provides a stable visualization of the eye, which is advantageous to the operator. On the other hand, whilst effective, it is a more computationally expensive way to process eye velocity (as opposed to *direct subtraction method*). In some instances, we found that the head signal

could not be measured, leading to baseline shifts in the observed eye displacement signal.

Here, our methods were presented with manual selection of the face region, and an interactive process to initiate the eye velocity estimation. We found that the interactive nature of the process was not particularly onerous. Our subsequent experimentation with automated face detection methods such as Zhu and Ramanan face detection algorithm [50] suggests that automated methods will provide further improvements in terms of robustness and usability of the algorithms. In this work our main focus was on head stabilization based on fast and simple approaches, and therefore the inclusion of automated face detection methods will be the subject of future work.

An obvious improvement would be the incorporation of the Visage SDK head signals into our methods. With additional work, it is likely that our methods could be combined to improve the robustness of the approaches we have developed. Visage SDK allows real-time processing, that could be used to develop a real-time system. Recently, free options such as Face++ [51] have also been proposed, that in our experience are likely to be effective as well. Notably, a face tracking option is available for mobile devices with limited frame rate and face size options.

There were some failures in face and pupil detection and tracking. In those cases the software attempted to find a next valid frame. Parallax error was present in our videos, due to the fact that our camera was placed to the side of the head in some cases. According to our results, it did not affect detection of OKN presence and direction but for precise analysis of OKN parameters such as gain and quick phase velocity, parallax free, 3D measurement would be useful. As the head rotates, the number of features tended to decrease, thereby resulting in a less effective stabilization. We note that yaw and pitch rotations were not well compensated for, due to the nature of these motions. In the future, we will investigate how multiple

cameras could be used to enable continuous monitoring of the 3D eye position during rotation. Advanced signal processing methods for analyzing the OKN waveform and assessment of the impact of camera frame rate and resolution role on detailed analysis of the eye movement signal will also be the subject of future work. A fully automated real-time implementation of the method is a long-term goal.

VI. CONCLUSION

We developed tools that improved the visualization of the eye, and facilitated the measurement of OKN related eye movements from videos of children. There was no need for special hardware, multiple cameras or infrared lights for this purpose. Our results in five participants indicated that our simple semi-automated method based on general face region features was sufficient for this purpose. The presence and direction of OKN was assessed *objectively* and validated by an experienced observer and a clinically trained assessor viewing the same video footage. The results here suggest that consumer grade equipment can be used for the assessment of eye movements in young children which may help in the diagnosis of visual disorders.

ACKNOWLEDGMENT

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

REFERENCES

- [1] N. S. Anstice and B. Thompson, "The measurement of visual acuity in children: An evidence-based update," *Clin. Experim. Optometry*, vol. 97, no. 1, pp. 3–11, 2014.
- [2] L. M. Chalupa, J. S. Werner, and C. J. Barnstable, *The Visual Neurosciences*, vol. 1. Cambridge, MA, USA: MIT Press, 2004.
- [3] J. Waddington and C. M. Harris, "Human optokinetic nystagmus: A stochastic analysis," *J. Vis.*, vol. 12, no. 12, 2012, Art. ID 5.
- [4] K. W. Wright, "Clinical optokinetic nystagmus asymmetry in treated esotropes," *J. Pediatric Ophthalmol. Strabismus*, vol. 33, no. 3, pp. 153–155, 1995.
- [5] P. Verwey, "Measuring vision in children," *Community Eye Health*, vol. 17, no. 50, pp. 27–29, 2004.
- [6] S. Garbutt and C. M. Harris, "Abnormal vertical optokinetic nystagmus in infants and children," *Brit. J. Ophthalmol.*, vol. 84, no. 5, pp. 451–455, 2000.
- [7] A. Hathibelagal, "Objective assessment of visual acuity in infants," M.S. thesis, School Optometry, Univ. Waterloo, Waterloo, ON, Canada, 2013.
- [8] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.
- [9] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," *IEEE J. Transl. Eng. Health Med.*, vol. 1, 2013, Art. ID 2100212.
- [10] S. M. H. Jansen, H. Kingma, and R. L. M. Peeters, "A confidence measure for real-time eye movement detection in video-oculography," in *Proc. 13th Int. Conf. Biomed. Eng.*, 2009, pp. 335–339.
- [11] B. Zhu, P. Y. Zhang, J. N. Chi, and T. X. Zhang, "Gaze estimation based on single camera," *Adv. Mater. Res.*, vol. 655, no. 2, pp. 1066–1076, 2013.
- [12] H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe, "Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions," in *Proc. Symp. Eye Tracking Res. Appl.*, 2008, pp. 245–250.
- [13] R. Valenti, J. Staiano, N. Sebe, and T. Gevers, "Webcam-based visual gaze estimation," in *Proc. Image Anal. Process. (ICIAP)*, 2009, pp. 662–671.
- [14] J. Y. Hyon, H. E. Yeo, J.-M. Seo, I. B. Lee, J. H. Lee, and J. M. Hwang, "Objective measurement of distance visual acuity determined by computerized optokinetic nystagmus test," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 2, pp. 752–757, 2010.
- [15] S. B. Han, E. R. Han, J. Y. Hyon, J.-M. Seo, J. H. Lee, and J.-M. Hwang, "Measurement of distance objective visual acuity with the computerized optokinetic nystagmus test in patients with ocular diseases," *Graefes Archive Clin. Experim. Ophthalmol.*, vol. 249, no. 9, pp. 1379–1385, 2011.
- [16] Y. J. Shin, K. H. Park, J.-M. Hwang, W. R. Wee, J. H. Lee, and I. B. Lee, "Objective measurement of visual acuity by optokinetic response determination in patients with ocular diseases," *Amer. J. Ophthalmol.*, vol. 141, no. 2, pp. 327–332, 2006.
- [17] D. Model and M. Eizenman, "An automatic personal calibration procedure for advanced gaze estimation systems," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1031–1039, May 2010.
- [18] D. Model and M. Eizenman, "An automated Hirschberg test for infants," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 103–109, Jan. 2011.
- [19] D. Model, "A calibration free estimation of the point of gaze and objective measurement of ocular alignment in adults and infants," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2011.
- [20] Z. Zhu and Q. Ji, "Novel eye gaze tracking techniques under natural head movement," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2246–2260, Dec. 2007.
- [21] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum, "Full-frame video stabilization with motion inpainting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1150–1163, Jul. 2006.
- [22] J. Bai, A. Agarwala, M. Agrawala, and R. Ramamoorthi, "Selectively de-animating video," *ACM Trans. Graph.*, vol. 31, no. 4, 2012, Art. ID 66.
- [23] J. Bai, A. Agarwala, M. Agrawala, and R. Ramamoorthi, "User-assisted video stabilization," *Comput. Graph. Forum*, vol. 33, no. 4, pp. 61–70, 2014.
- [24] T. Kinsman, K. Evans, G. Sweeney, T. Keane, and J. Pelz, "Ego-motion compensation improves fixation detection in wearable eye tracking," in *Proc. Symp. Eye Tracking Res. Appl.*, 2012, pp. 221–224.
- [25] C. A. Rothkopf and J. B. Pelz, "Head movement estimation for wearable eye tracker," in *Proc. Symp. Eye Tracking Res. Appl.*, 2004, pp. 123–130.
- [26] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1994, pp. 593–600.
- [27] C. Tomasi and T. Kanade, "Detection and tracking of point features," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, 1991.
- [28] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.
- [29] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [30] M. Sangi, B. Thompson, E. Vaghefi, and J. Turuwhenua, "A head tracking method for improved eye movement detection in children," in *Proc. 15th Int. Conf. Biomed. Eng.*, 2014, pp. 508–511.
- [31] T. M. Khan, M. A. Khan, S. A. Malik, S. A. Khan, T. Bashir, and A. H. Dar, "Automatic localization of pupil using eccentricity and iris using gradient based method," *Opt. Lasers Eng.*, vol. 49, no. 2, pp. 177–187, 2011.
- [32] F. Jan, I. Usman, S. A. Khan, and S. A. Malik, "A dynamic non-circular iris localization technique for non-ideal data," *Comput. Elect. Eng.*, vol. 40, no. 8, pp. 215–226, Nov. 2014.
- [33] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," in *Proc. VISAPP*, 2011, pp. 125–130.
- [34] Y. Li, S. Wang, and X. Ding, "Eye/eyes tracking based on a unified deformable template and particle filtering," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1377–1387, 2010.
- [35] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, Sep. 2009.
- [36] S. Shah and A. Ross, "Iris segmentation using geodesic active contours," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 4, pp. 824–836, Dec. 2009.
- [37] M. Leo, D. Cazzato, T. De Marco, and C. Distanto, "Unsupervised eye pupil localization through differential geometry and local self-similarity matching," *PLoS ONE*, vol. 9, no. 8, p. e102829, 2014.
- [38] M. Leo, D. Cazzato, T. De Marco, and C. Distanto, "Unsupervised approach for the accurate localization of the pupils in near-frontal facial images," *J. Electron. Imag.*, vol. 22, no. 3, p. 033033, 2013.
- [39] N. Markuš, M. Frljak, I. S. Pandžić, J. Ahlberg, and R. Forchheimer, "Eye pupil localization with an ensemble of randomized trees," *Pattern Recognit.*, vol. 47, no. 2, pp. 578–587, 2014.

- [40] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao, "2D cascaded AdaBoost for eye localization," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, 2006, pp. 1216–1219.
- [41] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR)*, Jun. 2005, p. 79.
- [42] X. Xiong, Q. Cai, Z. Liu, and Z. Zhang, "Eye gaze tracking using an RGBD camera: A comparison with a RGB solution," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct Pub.*, 2014, pp. 1113–1121.
- [43] Y.-W. Chen and K. Kubo, "A robust eye detection and tracking technique using Gabor filters," in *Proc. 3rd Int. Conf. Intell. Inf. Hiding Multimedia Signal Process. (IHMSP)*, vol. 1, Nov. 2007, pp. 109–112.
- [44] S. Chen and J. Epps, "Efficient and robust pupil size and blink estimation from near-field video sequences for human-machine interaction," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2356–2367, Dec. 2014.
- [45] X. Jiang, G. Tien, D. Huang, B. Zheng, and M. S. Atkins, "Capturing and evaluating blinks from video-based eyetrackers," *Behavior Res. Methods*, vol. 45, no. 3, pp. 656–663, 2013.
- [46] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, vol. 373. New York, NY, USA: Springer-Verlag, 2007.
- [47] J. Turuwhenua, T.-Y. Yu, Z. Mazharullah, and B. Thompson, "A method for detecting optokinetic nystagmus based on the optic flow of the limbus," *Vis. Res.*, vol. 103, pp. 75–82, Oct. 2014.
- [48] T.-Y. Yu, R. J. Jacobs, N. S. Anstice, N. Paudel, J. E. Harding, and B. Thompson, "Global motion perception in 2-year-old children: A method for psychophysical assessment and relationships with clinical measures of visual function," *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 13, pp. 8408–8419, 2013.
- [49] C. M. Knapp, I. Gottlob, R. J. McLean, and F. A. Proudlock, "Horizontal and vertical look and stare optokinetic nystagmus symmetry in healthy adult volunteers," *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 2, pp. 581–588, 2008.
- [50] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2879–2886.
- [51] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in *Proc. IEEE Int. Conf. IEEE Comput. Vis. Workshops (ICCVW)*, Dec. 2013, pp. 386–391.



MEHRDAD SANGI received the M.S. degree in electrical engineering-bioelectric from the Sharif University of Technology, Tehran, Iran, in 2010. He is currently pursuing the Ph.D. degree with the Auckland Bioengineering Institute, Auckland, New Zealand. His doctoral research is focused on automated optokinetic nystagmus detection methods for use with young children. An overall objective for this research project is the development of a stand-alone system for OKN analysis that could be used clinically or for research purposes.



BENJAMIN THOMPSON received the B.Sc. and Ph.D. degrees from the Department of Experimental Psychology, University of Sussex, Brighton, U.K. He held a post-doctoral fellowship with the Department of Psychology, University of California at Los Angeles, Los Angeles, CA, USA, and the Department of Ophthalmology, McGill University, Montréal, QC, Canada. He currently holds faculty positions with the University of Waterloo, Waterloo, ON, Canada, and the University of Auckland, Auckland, New Zealand.



JASON TURUWHENUA received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Physics, University of Waikato, Hamilton, New Zealand. He currently holds a joint appointment with the Auckland Bioengineering Institute, Auckland, New Zealand, and the Department of Optometry and Vision Science, University of Auckland, Auckland. His current research interests include the application of techniques from image processing and computer vision to eye related problems.