09-08

15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Initial Response Time Measurement in Eye Movement for Dementia Screening Test

Tamotsu Endo, NAIST
Norimichi Ukita, NAIST
Hiroki Tanaka, NAIST
Norihiro Hagita, NAIST
Satoshi Nakamura, NAIST
Hiroyoshi Adachi, Health Care Center, Osaka University
Manabu Ikeda, Graduate School of Medicine, Osaka University
Hiroaki Kazui, Graduate School of Medicine, Osaka University
Takashi Kudo, Health Care Center, Osaka University

Abstract

This paper proposes a method for detecting eye movement by using a generic RGB camera such as the one in a laptop computer or a smartphone. This method is designed for diagnosis of brain dysfunction such as dementia. For such a diagnostic purpose, precise detection of the eye movement is required, e.g., ± 1 video frame. However, noisy iris localization makes it difficult. The proposed method detects the moments when the iris begins and finishes moving by binary classification of eye movement velocity. This classification is achieved by discriminant analysis for optimization of these moments. We developed two diagnostic systems using the aforementioned function. The first one is designed just for diagnosis in which a subject gazes at a target point that appears on a white screen. The second one is merged with a generic web browser in which the gaze of a subject is measured while he/she reads texts displayed on the browser. Experimental results with these two systems demonstrated that the proposed method can detect the moments when the iris begins and finishes moving more accurately than required for a dementia screening test.

1 Introduction

It is well known that the number of dementia patients has been increasing in the aging society. The symptoms of dementia are caused by organic brain lesion(s). Since it is now impossible to heal these lesions but possible to slow down the progress, it is important to find early symptoms of dementia.

However, the early symptoms of dementia are almost identical to those appeared in a generic aging process. This makes it difficult to find the early symptoms even using a diagnostics procedure such as MMES [3].

One possible solution for finding the early symptoms of dementia is to always/often measure physical signs appeared on an elderly person for stochastic diagnosis. Such a daily measurement can be realized by information technologies.

While there are a variety of symptoms of dementia, in this paper, we focus on dysfunction in eye movement. It can be measured by a generic camera. In addition, it has been reported that the reticulotegmental nucleus of pons that controls eye movement can be damaged in the early stage of dementia. These facts motivate us to develop a method for measuring subtle eye movement for diagnosis of dementia.

However, prior eye-measurement technologies for the screening diagnosis of brain dysfunctions such as dementia cannot be used for home and daily use. This is because they require an operation by medical doctors, must be used with medical systems in hospitals, and/or are very expensive. Based on the discussion above, this paper proposes a method that detects eye movement in a video captured even by a generic camera for an easy-to-use dementia screening system.

2 Related Work

Not only for dementia but also for various brain dysfunctions, human gaze patterns provide several signs. For example, in [1, 8], atypical gaze patterns in patients of neurodegenerative disorders and Alzheimer's disease are reported, respectively. Such atypical patterns are also observed in dementia patients [9]. In these works, a gaze measurement device was used (e.g., search coil, EOG measurement devices, and Tobii). Since these devices are expensive, noisy, and/or difficult to use, they are not suitable for daily use at home. For daily use, the proposed method is designed to be applicable to a generic RGB camera such as the one in a laptop computer or a smartphone.

A number of gaze measurement methods have been proposed. Recently, appearance-learning-based methods [6, 15] allow us to accurately estimate a 3D gaze direction. While their performance is promising, they requires more or less a calibration process for estimating the geometric configuration between a screen and the user's head/eye. Since such a 3D gaze direction is not needed in the proposed method in which only eye movement is required, a simpler iris localization method is sufficient for our purpose. In contrast to conventional model-fitting-based methods such as [12, 13], the state-of-the-art [16] enables calibration-free robust iris localization in images/videos. This method [16] fits the goal of our work.

3 Overview and Our Contributions

We develop a HCI system for dementia screening [10]. While this system evaluates several features for the screening (e.g., acoustic, linguistic, dialogue, and facial features), this paper focuses on a gaze feature.

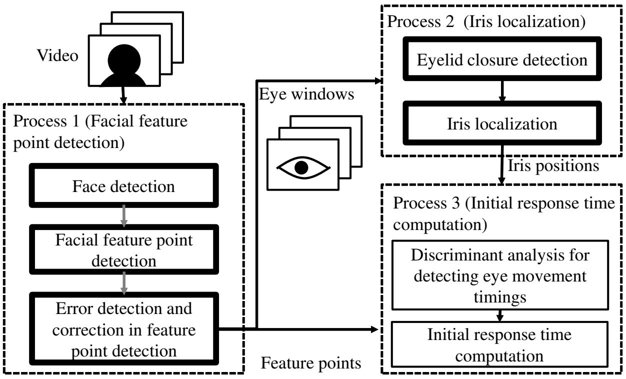


Figure 1. Overview of the proposed method. Processes 1, 2, and 3 are explained in Sections 4, 5, and 6, respectively.

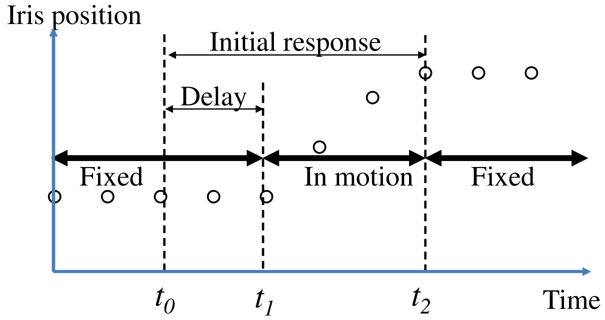


Figure 2. Initial response time in eye movement. The horizontal and vertical axes denote the time and the position of an iris captured in a video. Initially a subject is requested to gaze at a fixed target point. When the target simultaneously disappears and appears in a different location at t_0, the iris starts moving fast in order to gaze at the target at t_1 after a short delay in response. The iris is fixed again at t_2 when the gaze is directed towards the target. Thick arrows indicate whether the eye is fixed or in-motion. Delay in eye movement (i.e., $t_1 - t_0$) gets longer as the symptom of dementia gets worse. $(t_2 - t_0)$ is called an initial response time in eye movement.

Figure 1 illustrates the overview of the proposed method. A user is requested to gaze at a target point that continues to appear, move, and disappear on the screen of a computer, Microsoft Surface 3 in our experiments. A built-in camera of Surface 3 recorded a 2 Mbps video with 1920x1080 pixels at 30 fps.

Based on the results of extracting facial feature points and an eye window at each frame in preprocesses (Process 1 in Fig. 1), an iris is localized (Process 2). Then eye movement is captured around when the target point appears (Process 3). The aforementioned processes are achieved only for an eye (i.e., not for both eyes) just for simplicity.

Figure 2 shows the history of iris positions during eye movement. After the target point appears at t_0, the iris of a user starts moving at t_1. It stops when the user's gaze is directed towards the target point at t_2. Since various dysfunctions in eye movement are known as criteria for dementia screening [14, 5], we also focus on an initial response time defined to be $(t_2 - t_0)$ as such a criterion.

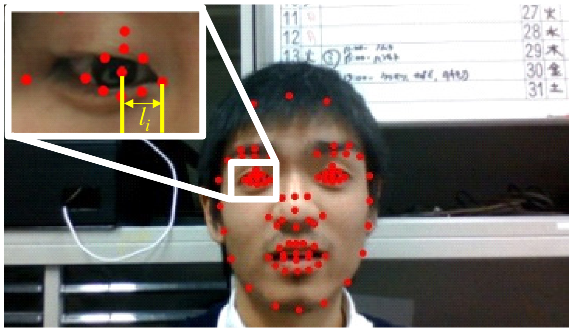


Figure 3. Facial points detected by ASM [7] in an image captured by a built-in camera of Surface 3.

From a technical point of view, the contribution of this work lies in how to pinpoint the timing of eye movement in a noisy trajectory of the iris.

4 Facial Point Detection for Eye Window Extraction

Process 1 extracts facial feature points at each frame. After face detection using the Viola-Jones detector [11], the active shape model (ASM) [2] is used for extracting facial feature points, as shown in Fig. 3. A variant of the ASM, STASM [7], is used in our implementation.

The ASM localizes an iris as well as a number facial feature points including inner and outer corners of the eye, as shown in a magnified image in Fig. 3. Since the head may move during the video capturing for a screening test, the proposed method computes the length between the iris and the eye corner and regards this length as the position of the iris, which is invariant against the head motion. More specifically, the length between the iris and the inner corner of the eye (i.e., l_i in Fig. 3) is computed because we found that the inner corner can be localized well by STASM [7] rather than the outer corner; 7 pixels in the inner corner vs 8 pixels in the outer corner on average, and 58 pixels in the inner vs 86 pixels in the outer at most.

While the ASM provides the position of an iris, its localization accuracy is not sufficient for eye movement detection because the iris is partially occluded by the eyelid. More accurate iris localization is achieved in a window around the feature points of the eye (i.e., the magnified image in Fig. 3). This window is called an eye window.

5 Iris Localization

Before iris localization, the proposed method detects whether or not the eyelid is opened at each frame. If not, the iris is not localized at that frame, and a dementia screening test is not performed using frames before and after that frame.

The proposed method localizes the iris by [16]. This algorithm [16] localizes the pupil as well as the iris, as shown in Fig. 4. Our proposed method employs the center position of the iris because of higher accuracy of the iris localization in our tests. Examples of the

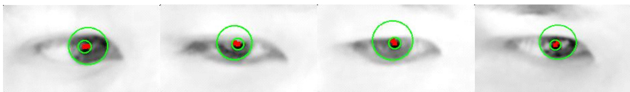


Figure 4. Examples of iris localization results [16]. Inside and outside circles indicate a pupil and an iris, respectively.

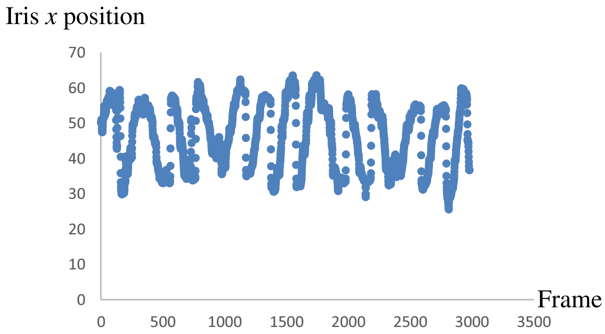


Figure 5. Measurement results of temporal iris positions. The x coordinate of an iris in a video is measured while a target point continues to disappear, appear, and move.

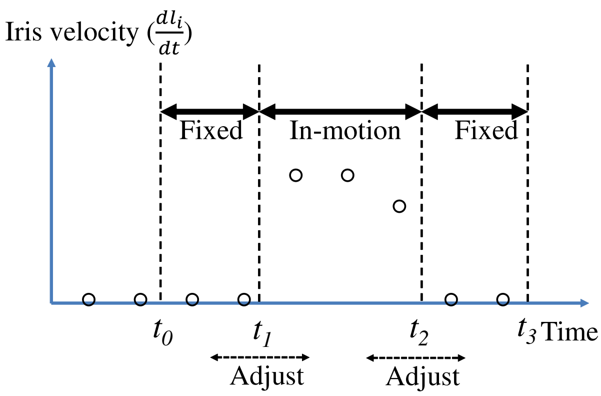


Figure 6. Eye movement velocities are classified into two classes, namely "fixed" and "in-motion" by discriminant analysis. t_1 and t_2 are adjusted in a bruteforce manner for the best discrimination between fixed and in-motion frames.

temporal history of the iris positions, which were captured while a user gazes at a target point on a screen, are shown in Fig. 5.

With this iris localization algorithm, l_i is computed at each frame. Its localization error in a validation data was 3.7 pixels on average. With the temporal history of l_i, eye movement is detected as described in Sec. 6.

6 Eye Movement Detection by Timing Optimization using Discriminant Analysis

While the iris can be localized as shown in Fig. 5, noise in localizing the iris and the eye corner makes it difficult to accurately detect eye movement timings. This is because it is difficult to discriminate between such noises and movement of the iris at t_1 and t_2.

The proposed method reduces this difficulty by binary optimization of eye movement. Eye movement is represented by the velocity of the iris and classified into two classes, "fixed" and "in-motion". Frames where the iris velocities are fixed and in-motion are determined by t_1 and t_2, as illustrated in Fig. 6. t_1 and t_2 are optimized by discriminant analysis as follows:

Step1: When a target point simultaneously disappears and appears at t_0 in our dementia screening system, a high-frequency beep sounds. This beep is recorded in a video and allows us to pinpoint t_0 in the video.

Step2: t_3 is determined so that the following two conditions are satisfied: (1) $t_3 > t_2$ and (2) $(t_3 - t_2)$ is not so large; if $(t_3 - t_2)$ is much larger than $(t_2 - t_1)$ discriminant analysis between fixed and in-motion shown in Fig. 6 becomes difficult. While $(t_2 - t_0)$ changes depending on the individual at each test, its variation is small. The difference between the max and min of $(t_2 - t_0)$ is around 20 frames. This allows us to determine t_3 satisfying the aforementioned two conditions. In our experiments, $(t_3 - t_0) = 30$ frames.

Step3: t_1 and t_2 are then optimized so that a discrimination score (5) is maximized by a bruteforce search. Then the initial response time, $(t_2 - t_0)$, is determined.

$$\omega_f = \sum_{k=t_0}^{t_1} k + \sum_{k=t_2+1}^{t_3} k \tag{1}$$

$$\omega_m = \sum_{k=t_1+1}^{t_2} k \tag{2}$$

$$m_f = \frac{\sum_{k=t_0}^{t_1}(x_k - x_{k-1}) + \sum_{k=t_2+1}^{t_3}(x_k - x_{k-1})}{\omega_f} \tag{3}$$

$$m_m = \frac{\sum_{k=t_1+1}^{t_2}(x_k - x_{k-1})}{\omega_m} \tag{4}$$

$$\sigma^2 = \omega_f \omega_m (m_f - m_m)^2 \tag{5}$$

7 Experimental Results

We validate accuracy of initial response time computation. The following three types of experiments were conducted in order to demonstrate the effectiveness of the proposed method in different scenarios:

Ex1: Subjects were 13 males and 3 females. All of them were twenties, grad school students. A black target point was displayed on a white screen of Surface 3. The target disappeared and appeared 14 times in a test for each subject.

Ex2: The set of subjects was same with the one of Ex1. Instead of a target point, a subject was requested to read an article displayed on a web browser. When a popup window appeared, the subject turned his/her gaze towards the popup, which is is regarded as a target point in Ex1. Whereas Ex1 simulates a software used only for a screening test (e.g., screening software used in a hospital), Ex2 simulates a scenario where screening is automatically performed while a user operates a computer in daily life.

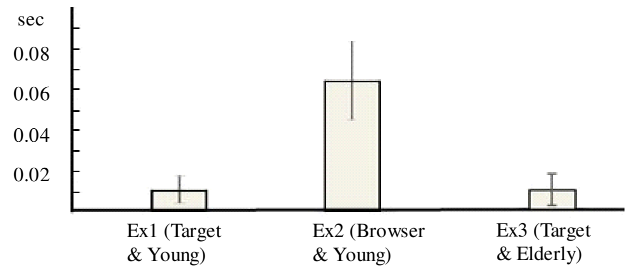
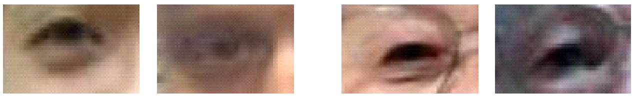


Figure 7. Error in measuring the initial response time (sec). A mean value and a standard deviation are shown for each experiment.



Young Elderly

Figure 8. Eyes of young and elderly people.

Ex3: While the experimental environment is identical to the one of Ex1, subjects were 18 elderly people including 9 very early stage of dementia patients. All data for this experiment were recorded by medical doctors.

All experiments were approved by ethics committees in the authors' institutes.

As described before, the RMSEs of the inner corner of the eye and the iris were 7 pixels and 3.7 pixels, respectively. On the other hand, the iris moves around 3 pixels at each frame when a subject moved his/her gaze towards a target appeared on a screen. That means it is difficult to detect the moments when the iris starts and finishes moving.

The results of these three experiments are shown in Fig. 7. The RMSEs in Ex1, Ex2, and Ex3 are 0.010 sec, 0.064 sec, and 0.028 sec, respectively. The RMSE was computed over all trials of all subjects in each experiment. It is reported that the mean value of the initial response time in dementia patients is more than 0.1 sec longer than that of elderly control subjects [4]. The experimental results demonstrate that accuracy of the computed the initial response time is sufficient for discriminating between dementia patients and elderly control subjects. In particular, the RMSE is less than 1 frame (i.e., 1/30 sec) in a test using a target point (i.e., Ex1 and Ex3).

The difference in the RMSE between Ex1 and Ex3 is caused by difficulty in localizing the iris occluded by the eyelid for elderly people rather than young people as shown in Fig. 8.

A higher RMSE in Ex2 was caused because of a wide range of initial response times of subjects. Unlike Ex1 and Ex3 in which a subject were concentrated on a target point, a large delay (i.e., $t_1 - t_0$) was observed in several tests. Such a delay causes an imbalance between frames of fixed and in-motion within the interval between t_0 and t_3. In the worst case, the iris is still in-motion at t_3. While this interval is fixed for simplicity in our experiments, it should be adjusted online depending on the moment when the iris starts moving.

8 Concluding Remarks

This paper proposed a method for measuring an initial response time in eye movement with a generic RGB camera. In the proposed method, discriminant analysis optimizes the moments when the iris begins and finishes moving. Experiments using these two systems were conducted with young adults and elderly people. Experimental results demonstrated that the proposed method can measure the initial response time more accurately than required for a dementia screening test.

This work was supported by Yanmar Lab 2112.

References

[1] T. J. Anderson and M. R. MacAskill. Eye movements in patients with neurodegenerative disorders. *Nature Reviews Neurology*, 9:74–85, 2013.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[3] M. Folstein, S. Folstein, and P. McHugh. "mini-mental state". a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

[4] I. Fukumoto. A basic research of diagnostic method for alzheimer type dementia based on eye movements. Technical report, Nagaoka University of Technology, 2002.

[5] A. Jones, R. P. Friedland, B. Koss, L. Stark, and B. A. Thompkins-Ober. Saccadic intrusions in alzheimer-type dementia. *J Neurol.*, 229:189–194, 1983.

[6] Y. Lin, R. Lin, Y. Lin, and G. C. Lee. Real-time eye-gaze estimation using a low-resolution webcam. *Multimedia Tools Appl.*, 65(3):543–568, 2013.

[7] S. Milborrow and F. Nicolls. Active shape models with SIFT descriptors and MARS. In *VISAPP*, 2014.

[8] M. L. F. Pereira, M. von Zuben A Camargo, I. Aprahamian, and O. V. Forlenza. Eye movement analysis and cognitive processing: detecting indicators of conversion to alzheimers disease. *Neuropsychiatr Dis Treat.*, 10:1273–1285, 2014.

[9] V. E. Sturm and et al. Mutual gaze in alzheimer's disease, frontotemporal and semantic dementia couples. *Soc Cogn Affect Neurosci.*, 6:359–367, 2011.

[10] H. Tanaka, H. Adachi, N. Ukita, T. Kudo, and S. Nakamura. Automatic detection of very early stage of dementia through multimodal interaction with computer avatars. In *ICMI*, 2016.

[11] P. A. Viola and M. J. Jones. Robust real-time face detection. In *ICCV*, 2001.

[12] J. Wang, E. Sung, and R. Venkateswarlu. Eye gaze estimation from a single image of one eye. In *ICCV*, 2003.

[13] E. Wood and A. Bulling. Eyetab: model-based gaze estimation on unmodified tablet computers. In *ETRA*, 2014.

[14] G. Zaccara, P. F. Gangemi, G. C. Muscas, M. Paganini, S. Pallanti, A. Parigi, A. Messori, and G. Arnetoli. Smooth-pursuit eye movements: alterations in alzheimer's disease. *J Neurol Sci.*, 112:81–89, 1992.

[15] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Appearance-based gaze estimation in the wild. In *CVPR*, 2015.

[16] Z. Zhao and A. Kumar. An accurate iris segmentation framework under relaxed imaging constraints using total variation model. In *ICCV*, 2015.