EPiC Series in Computing
Volume 77, 2021, Pages 52–59
Proceedings of ISCA 30th International Conference on Software Engineering and Data Engineering

Estimating Attention Level from Blinks and Head Movement

Masaaki Goto, Tetsuo Tanaka, and Kazunori Matsumoto
Kanagawa Institute of Technology, Japan

{t-tanaka,matumoto}@ic.kanagawa-it.ac.jp

Abstract

In this paper, we develop a real-time method for estimating the level of attention while performing a task. This method uses only a low frame rate video from a standard camera so that it can be available even on a small computer. Eye blinks and head movements are calculated from a video by using landmarks. Existing blink detection methods use standard frame rate videos, making them difficult to process on a computer with low processing performance. One obvious solution is to use videos with a reduced frame rate. We investigate the error caused by reducing the frame rate, and then to overcome the error, we further develop a new method that uses the head movements calculated from the reduced frame rate videos. Then we demonstrate the error is within acceptable ranges by using the method, and show it is effective to estimate the attention level. Since this method uses only landmark information obtained from facial images, it reduces the mental burden on the user, and also partially protects personal information. In this paper, we explain the details of the method and report the experimental results.

1 Introduction

When performing a task, the mental state of the worker, i.e., the degree of interest in the content of the task, likes or dislikes, and the degree of motivation, etc. affect the efficiency of the task. The physical condition of the worker and the surrounding environments, such as temperature and humidity, also affect the degree of the above factors. A skilled manager carefully observes the workers themselves, and the surrounding conditions to estimate the mental and physical states of them, and to ensure that the efficiency of work is maintained. If we can automate these observations and estimations of the manager, it can be applied to various fields and can bring a lot of benefits. We assume in this study that the mental and physical states or conditions are exposed to the outside and can be observed. This is not only empirically accepted, but has already been demonstrated by several studies [2,8]. These states or conditions are, in this study, collectively referred to as the level of attention. In other words, we use the term of attention as a comprehensive indicator of work efficiency. In order to measure the level of attention, methods using fluctuations in respiration, heart rate, electroencephalogram, and skin potential have been studied [4,6,15]. The equipment required for these measurements is not common, and there is a physical burden from wearing the equipment, they thus are not widely available in the standard situation. We need a more applicable method using only commonplace devices such as Web cameras.

Blinking is closely related to the attention level [2,12,13], and the development of technology to automatically detect it is increasing. The development of systems that apply blink detection is also increasing [7,8,9]. In particular, it is often used to detect drowsiness when driving a car [1]. It is also beginning [4,8,12] to be used as a technology to monitor the status of students in online classes. As the application of blink detection becomes more widespread, it is also expected to be used on small computers such as Raspberry Pi. Therefore, there is a growing need to develop methods that can be used on less capable computers. In this paper, we attempt to improve the efficiency of the conventional method using video images by reducing the frame rate, and examine the accuracy drop. In addition, we show that the combination of several methods can be used for practical applications.

2 Blink Detection from Low Frame Rate Video

Different approaches to blink detection have been proposed, a compact survey of the studies is available in [6]. Methods using video images are developed: in some methods, images are used as is, while in others, landmarks are extracted from the images and used in the successive processes. The landmarks are the points attached to characteristic parts of a face image, such as the periphery of the eyes, nose, and mouth. The method for extracting landmarks and their evaluation are given in [10]. Procedures for extracting landmarks are readily available as the common libraries, we use one of them in this study. Figure 1 shows the landmarks overall face, and the figure also shows the landmarks around the eyes. The eye aspect ratio EAR is often used, which is calculated from the landmarks with the following equation. This definition of EAR is taken from [10]. Although it is possible to determine the blink rate based on the threshold EAR value alone, it has been found that there are some accuracy problems, such as false recognition. To solve this problem, the method in [10] use the successive EAR values of the six frames before and after a certain point, and by using these values they train the SVM that detects blinking. Experiments have shown that this method gives better results than the simple method using only the EAR.

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||}$$

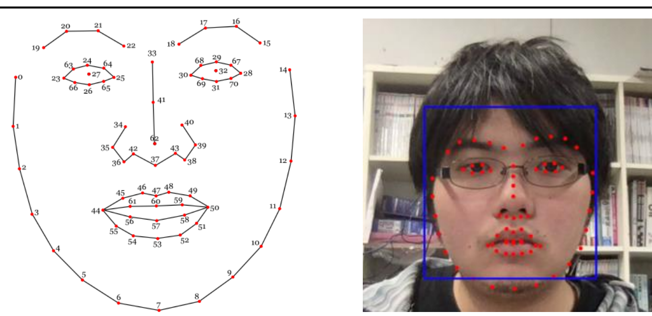

Figure1. Landmarks

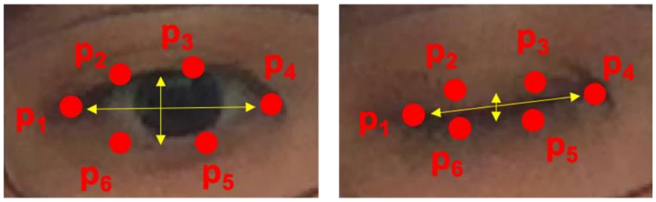

Figure 2. Landmark around Eye

Since the SVM methods in [10] deal with 30 fps videos, processing speed may be a problem on low performance computers such as Raspberry Pi and Arduino. Then real-time processing on such computers becomes hard. In order to achieve the high efficiency and realtime-ness, we adopted a method of dropping the frame rate. We use a simple EAR threshold value instead of the SVM. When the EAR value is less than the threshold value, which is set to around 0.2 usually, the frame is judged to be blinking. The threshold of the EAR can be adjusted for each individual, taking into account the fact that the aspect ration of eyelids varies from person to person.

If the average duration of a blink is m milliseconds, m is around 500, and the frame rate is x fps, each blink will span m/x frames. Therefore, when counting the number of blinks, the number of frames determined to be blinks is divided by m/x for compensation. For example, if the average blink time is 200 milliseconds and the number of frames counted as blinks at 30 fps is 70, the number of blinks is

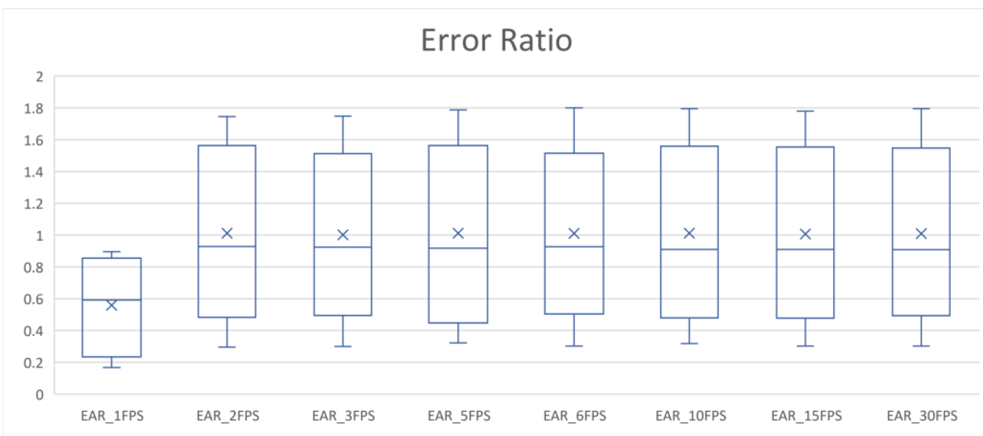

Figure 3. Error Ratio and Frame Rate

70/6.67 = 10.5. To confirm the effectiveness of this method, we carry out experiments using a standard video database [14].

The graph in Figure 3 shows the ratio w/b in the box charts, where b is the true number of blinks and w determined by our method. The horizontal axis is the frame rate, which is varied from 1 fps to 30 fps. The true number of blinks was determined manually by the authors while referring to the results of the method with SVM [6] . The SVM method judges blinks based on the change of the EAR values in successive frames, it has higher accuracy than the methods using only the fixed threshold EAR value. As can be seen from the graph, the results did not change in magnitude when the frame rate was changed.

At all frame rates, the median value was almost 1.0, which is almost the same as the true number of blinks. The detected blinks with our method spread in the range of half to twice the true count. From the results of this experiment, it is confirmed that dropping the frame rate to 1 fps does not cause much problem, so we set to 1 fps in the method. This means that our method can stably run even on a slow computer.

3 Estimating Attention Level

It has been reported that the level of attention is expressed in blinks and body movements [2,4]. The number of blinks of a person decreases when the attention level is high and he/she is concentrated. On the other hand, when the level of attention is low and not concentrated, the number of blinks tends to increase. In addition, not only the number of blinks, but also the pattern of blink occurrences depend on the attention level. In this study, we focused only on the number of blinks in order to reduce the processing complexities. Similarly, body movements are also related to the attention level. The larger the movement, the lower the attention level tends to be, while the smaller the movement, the higher the attention level tends to be. Of course, there are personal differences in these relationships, and it is difficult to directly compare the number of blinks and degree of movements with others. This section explains a method for estimating the attention level using blinks and movements.

The larger the number of blinks, the higher the attention level tends to be. We normalize the number to absorb personal differences. For a given fixed length of time interval t, let F_{max} be the maximum number of blinks and F_{min}be the minimum number of blinks within this interval. These values are determined for each person. If the number of blinks observed now is f for the same length of interval t, we define $a_b$as follows:

$$a_b = \frac{F_{max} - f}{F_{max} - F_{min}}$$

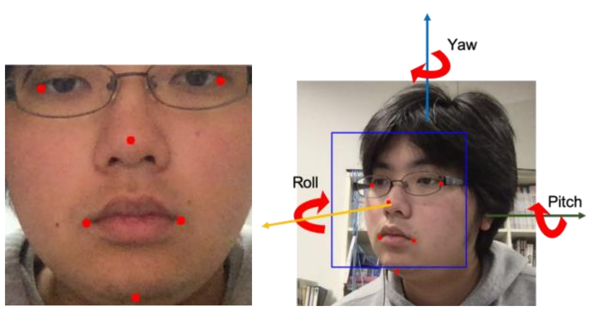

Figure4. Landmark and Rotation

This is normalized by the feature values of each person, and takes real values between 0 and 1. Intuitively, the value corresponds to the level of attention. The value of a_b becomes when the maximum number of blinks is observed, and becomes 1 when the minimum number of blinks is observed. The length of t is determined experimentally, but in the experiments of this paper, we set to $t = 5$ seconds.

As discussed in the first section, our method emphasizes computational efficiency, we use low frame rate videos so that the error in counting the number of blinks cannot be ignored. We thus cannot ensure the reliability of the a_b value, which is calculated from the blink count alone. In order to increase the reliability, we also use an index calculated from head movements. In this study, we deal with the head movement from two different views: the total distance of movements and the rotation. We use six characteristic landmark points to measure both of them.

The move of the head is calculated based on the distance between the landmark points of the previous frame and those of the previous frame. Similar to the above formula, we define M_{max} as the maximum movement and M_{min} as the minimum movement during the interval length t for each person. If the motion observed during the interval of length t is m, it is normalized by the following equation, which intuitively becomes the attention level to focusing on the motion. As in the previous case, it takes real values from 0 to 1, and becomes 0 when the motion is maximum, indicating that the attention level is minimum, and becomes 1 when the motion is minimum, indicating that the attention level is maximum.

$$a_m = \frac{M_{max} - m}{M_{max} - M_{min}}$$

For movements in terms of rotation, we evaluate Roll, Yaw, and Pitch, as shown in Figure 4. As before, the maximum values of $R_{max}, Y_{max}, P_{max}$ are obtained for each person. If the face is tilted too much, we cannot detect landmarks, so the value at the limit of detection is the maximum value. Let r, y and p be the observed roll, yaw, and pitch, respectively, for time interval of length t, and w is the average the ratios of rotation to the largest motion.

$$w = \frac{1}{3}\left(\frac{r}{R_{max}} + \frac{y}{Y_{max}} + \frac{p}{P_{max}}\right)$$

The attention level a finally is obtained by combining a_b and a_m using weight w. This value also ranges from 0 to 1, where 0 is the lowest attention level and 1 is the highest attention level.

$$a = (1 - w) \cdot a_b + w \cdot a_m$$

Errors in the EAR values increase as the rotation of the head increases, and then the accuracy of blink detection decreases. In the above equation, the weight w is used to control this situation.

4 Experimental Results

We carried out two experiments to evaluate the effectiveness of the developed method. The experiments are roughly investigated at this time, we have not yet obtain precise evolutional results. In the first experiment, seven university students were asked to read an easy short novel, seven university students were asked to read a short novel, which took about three minutes to read. Our method measured continuously the attention levels with t=5 until the students finished reading. The level was outputted every t=5 seconds, the average was used as the attention level of the reading. At the same time as the estimation by the method, the authors observed the students, manually estimated the attention levels at each point in time, recorded them, and finally calculated the average.

In Figure 3, the horizontal axis is the attention level obtained by the method, and the vertical axis is the manually obtained attention level. The regression line is also shown in the figure, and the result is roughly consistent with manual judgment, while R^2=0.56.

In the next experiment, we used two different short movies, one is interesting and preferred by students, the other is not interesting and disliked by students. We showed the students these movies, and obtained estimated attention level with t=5. Figure 6 shows the result for one student, the horizontal axis is the time, and the vertical axis is the level of attention. This figure only shows the results for one student, but we have obtained similar results for others. The evaluation of this experiment is yet to be done, but the results are almost consistent with the self-evaluation of the subjects themselves.

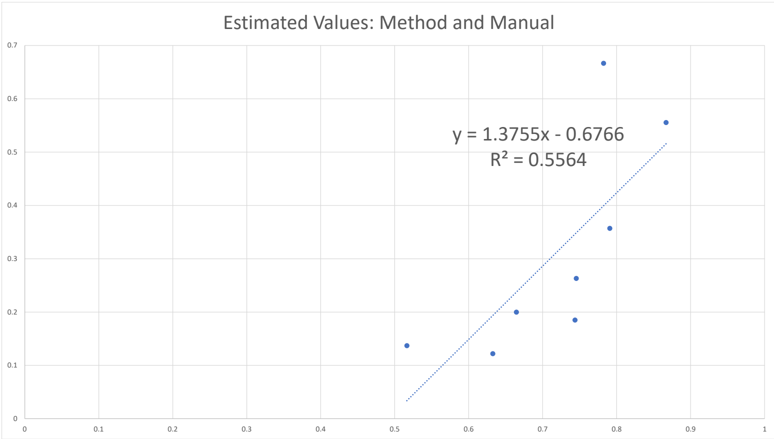

Figure 5. Attention Level: Method and Manual

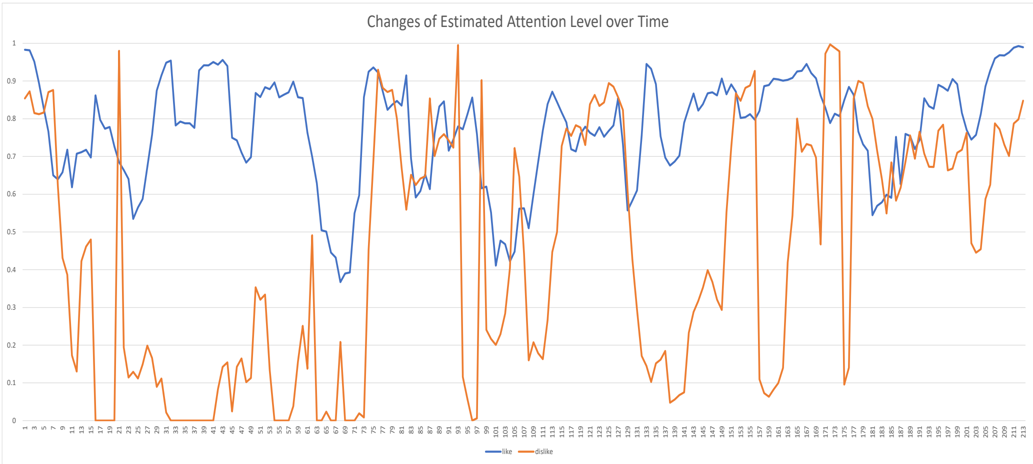

Figure 6. Change of Estimated Attention Level for Two Videos

5 Conclusions

We proposed a method for estimating the level of attention using low frame rate videos that runs even on low-speed computers with cameras. As shown in the experimental results, we roughly verified the effectiveness of the proposed method. The accuracy of the method needs to be further improved, while scientific and rigorous evaluation of the method is a remaining issue.

The efficiency of work depends on the level of attention while working on it. The technology to automatically estimate the attention level is eagerly awaited in many fields, it can be used to automatic manage of work to achieve high productivity. It is applicable to many fields with various situations. For example, in the current online learning, students are required to manage their own learning process, the teacher cannot monitor them face-to-face, and are responsible for their own learning results. The attention level estimation can improve this situation, the learning systems can assist the students by adequate support in accordance with the estimated attention level. Our method use only video images from cameras, which are easily available. In addition, to ensure stable operation on small computers such as the Raspberry Pi and Arduino, we adopt an extremely low frame rate of 1 fps. The applicability of the method spread widely.

In the method, we do not deal with the face image in the essential parts of the attention level estimation, we use only the extracted landmark information. This reduces the mental burden on the user, and also protects personal information. The landmark detection method determines the performance of this method, detection may not work well when the subject has long hair or is wearing a mask. The system for applying this method is currently under development, and will be reported in a future paper.

References

[1] M. H. Baccour, F. Driewer, E. Kasneci and W. Rosenstiel, Camera-Based Eye Blink Detection Algorithm for Assessing Driver Drowsiness, *2019 IEEE Intelligent Vehicles Symposium (IV), 2019*, pp. 987-993

[2] C. Becchio, A. Koul, C. Ansuini, C. Bertone, and A. Cavallo, Seeing mental states: An experimental strategy for measuring the observability of other minds, *Physics of Life Reviews, Elsevier, Vol. 24, 2018*

[3] S. George, and S. K. Praharaj, Eye Blink Count and Eye Blink Duration Analysis for Deception Detection, 2*017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*

[4] Y.Gong, and S.Xu, Mental State Detection in Classroom based on EEG Brain Signals, *Natural Science, 2019, Vol.11, No.11, pp. 315-322*

[5] V. Gudžiūnas, E. Paulėkas, G. Urniežiūtė, The Validity of eye blink rate by using quantitative parameters of attention level, 2015 7th International Workshop of Data analysis methods for software systems (DAMSS)

[6] P. Govardhane, and M.V. Wyawahare, Survey of Eye Blink Detection Techniques, *International Journal of Advanced Trends in Computer Applications ,2017 Volume 4, Number 2, pp. 1-4*

[7] D. Hoppea,b, S. Helfmannb, and C. A. Rothkopfa, Humans quickly learn to blink strategically in response to environmental task demands, *2018 Proc. of the National Academy of Sciences of the United States of America, Vol.116, No.9, pp. 2246-2251*

[8] A. Islam, N. Rahaman, Md A. R. Ahad, A Study on Tiredness Assessment by using Eye Blink Detection, *2019 Jurnal Kejuruteraan 31(2), pp. 209-214*

[9] C. Ranti, W. Jones, A. Klin, and S.Shultz, Blink Rate Patterns Provide a Reliable Measure of Individual Engagement with Scene Content, *Scientific Reports, (2020) 10:8267, www.nature.com/scientificreports.*

[10] T. Soukupova, and J. Cech, Real-Time Eye Blink Detection using Facial Landmarks, *2016 Proc. of the 21st Computer Vision Workshop, 2016*

[11] A. Vaitkus, V. Gudžiūnas, E. Paulėkas, G. Urniežiūtė, The Validity of Eye Blink Rate by using Evaluation of Attention, *2017 Jouranal of Medical Sciences, pp. 1-10*

[12] N. Vettivel, N. Jeyaratnam, V. Ravindran, S. Sumathipala and S. Amarakecrthi, System for Detecting Student Attention Pertaining and Alerting, *2018 3rd International Conference on Information Technology Research (ICITR), pp. 1-6*

[13] K.Yajima, K.Yonemura, Y.Takeichi, and J.Sato, Study of a Visualization System of the Concentration Degree by Biological Signal: Case of Active Learning, *2019 International Conference Educational Technologies, pp. 181-186*

[14] S. Zafeiriou, G. Tzimiropoulos, and M. Pantic, The 300 videos in the wild facial landmark tracking in-the-wild challenge, *2015 ICCV Workshop*

[15] F. Zennifa, and K. Iraminina, Quantitative Formula of Blink Rates-Pupillometry for Attention Level Detection in Supervised Machine Learning, *2019 IEEE Access, Vol.7*