Detecting Concentration of Students Using Kinect in E-learning

Takashi Ito^{a,*}, Shuhei Kotake^a, Kenichi Takahashi^a, Michimasa Inaba^a

^a Hiroshima City University, 3-4-1 Ozukahigashi, Asaminami-ku, Hiroshima 731-3194, Japan

*Corresponding Author: takahasi@hiroshima-cu.ac.jp

Abstract

In some lectures using e-learning systems, there are some cases where students do not concentrate properly on learning, and teachers cannot grasp whether students properly and intensively studied or not. In response to this problem, an e-learning system that judges whether or not students study appropriately with concentration using Kinect and warns them to concentrate when they do not. Through experiments, the usefulness of the system is examined. Additionally, a questionnaire is employed to see the feeling of students.

Keywords: e-learning, face tracking, sys direction, Kinect.

1. Introduction

In recent years, along with the spread of personal computers and the Internet, e-learning systems, with which students study on his/her own initiative, have become widespread^(1,2). Actually, e-learning systems are used at many educational sites such as universities and enterprises. Advantage of e-learning includes cost aspect, ease of data management by teachers, and learning without time or place being limited as well as without gathering at the same time, which give advantages to both students and teachers^(3,4). On the other hand, there are drawbacks: learners cannot ask questions to the teacher on the spot, learners are difficult in maintaining the motivation and concentration, and teachers cannot grasp the attitude of learners to tackle the learning.

In general, e-learning systems are aimed at voluntary studying such as the study for qualification acquisition. However, in some lectures using e-learning systems, there are some cases where learners do not concentrate properly on learning, and answer questions inappropriately, sleep in the middle of learning, touch smartphones during the lecture. In response to this problem, we have conducted research on learning systems which improve the degree of concentration

and motivation for learning. We have implemented those functions such as a ranking function competing among learners, a function of displaying scolding and praising images to encourage students, and encouragement in images, a function a function presenting problems each other on the web-based e-learning system aiming at improving motivation for learning. Those functions are shown to be effective^(5,6). Additionally, we have shown that an e-learning system which introduces the pleasure of the game refreshes learners during the break time by doing games and is effective in maintaining of concentration^(7,8).

Currently, systems called LMS (Learning Management System) or CMS (Course Management System) have been developed for teachers to grasp the attitude of students during e-learning; the systems can record as the learning log the access history including log-in and log-out times, and the answer history in which answers to questions of each learner are recorded⁽⁹⁾. With this system, the teacher can grasp information corresponding to attendance or performance like in the face-to-face lecture. However, those data do not provide visual information such as facial expression, eye directions, posture of the learner that the teacher can observe in the face-to-face lecture environment. We cannot grasp whether learners intensively learn or not just by using the access history and the answer history. Therefore, in the previous study, we developed an e-learning system that captures images of learners in real time and judges whether learners study appropriately and with concentration by using template matching, Adaboost, and watching mouse move⁽¹⁰⁾. This system checks whether the background screen is appropriate, whether the learner's face is facing forward, whether the learner is not moving compared with the image one second ago, and whether the mouse has not stopped for a certain period of time or more. It displays a warning image if the learner is judged that he does not concentrate. Furthermore, in order to ascertain the usefulness of this system, we examined the correct answer rate of each

judgment and investigated whether there is a difference in the duration of concentration and the score rate depending on whether or not there is a warning when they are not in concentration. The system had a problem that the accuracy was low and the learner had difficulty to concentrate due to many warnings.

In recent years, Kinect has attracted attention as a motion sensor device that can be purchased at low cost⁽¹¹⁾. Kinect is equipped with an RGB camera, a depth sensor, a multi-array microphone, etc. It can relatively easily and accurately detect and recognize information such as voice, human body movement, human face movement and expression, just by programming⁽¹²⁾. Therefore, in recent years, research on applying Kinect to media art, new interface system, and optical systems has been actively in progress. In this research, we introduce "Face Tracking" (face recognition function), one of the functions of Kinect, to the e-learning system, and check whether the learner's face and eye direction are facing forward and whether the eyes are open. In addition to them, we check whether the mouse is not stopped for a predetermined time or more. We construct a system to determine whether the learner is properly and intensively studying using those information, and warn the learner to concentrate if the system determine that the learner does not concentrate. Furthermore, in order to verify the usefulness of the system, we examine the accuracy of detecting concentration and compare the growth rate of the grading points with warning with that without warning. Additionally, we conducted the questionnaire about the system after the experiment is finished, and we discuss the usefulness of the system.

In this paper, we describe the details of detection method of concentration in Section 2. Experiments of judging concentration using Kinect is discussed in Section 3, and results of the questionnaire are shown in Section 4. Finally, in Section 4 we summarizes concluding remarks and future task.

2. Detection of Concentration

2.1 Overview

This system monitors the learning situation of learners using a set of "Kinect for Windows v2" as shown in Fig. 1 and estimates the degree of concentration of a learner by using Face Tracking⁽¹³⁾, one of the functions of Kinect.

The flow of this system is shown below.

S1) First, we initialize Timer and Count to 0. Timer is a

variable that controls the interval at which the degree of concentration is checked, and Count is a variable that counts the number of times of observing that the mouse does not move.

S2) After that, Kinect is started and a page for face recognition is displayed on the screen. The learner let Kinect recognize his / her face watching the screen. If the face recognition succeeds, the concentration degree judgment using Face Tracking and e-learning starts.

S3) Next, data obtained from Face Tracking is updated, and the degree of concentration is measured using that data. If $\text{Timer} < \alpha$, Timer is incremented by 1. If $\text{Timer} \geq \alpha$, the concentration degree judgement and the mouse-move judgement are performed. Here, α represents the interval at which the degree of concentration is determined, and in this experiment it is set to 10 seconds ($\alpha=10$). In the concentration degree determination, the degree of concentration is calculated using measurement data for a seconds.

S4) Thereafter, if the value of the concentration degree is lower than the threshold value, a warning is issued prompting concentration. Timer is initialized to 0 after α seconds.

S5) Finally, after 90 minutes learning, the system is terminated.

We also determine whether or not the position of the mouse is moving during β seconds.

S1) Initialize count to 0 if there is movement in the mouse. If there is no movement of the mouse, Count is incremented by 1.

Fig. 1. The learning situation with Kinect.

S2) If $\text{Count} \geq \beta$, a warning prompting learning is issued, and then, Count is initialized to 0. Here, β corresponds to the duration that there is no mouse motion to issue a warning representing tens of seconds, and in this experiment β was set to 3 (30 seconds). The system updates data obtained from Face Tracking at every one second.

If the learner's face is not recognized by Kinect, the system assume that during that time the learner is not concentrated. "Face Tracking" used in this system and the calculation method of concentration is described in detail in the following subsections.

2.2 Kinect

Kinect is a generic term for RGB cameras developed by Microsoft Corporation, depth sensors, motion sensor devices with microphones and SDK⁽¹¹⁾. "Kinect for Windows v1" was released in 2010, "Kinect for Windows v2" was released in 2014. Since the number of skeletons that "Kinect for Windows v2" can detect increased, it can detect the movement of the neck, and it can recognize the expression of the detected person⁽¹²⁾. Thus, using these functions, we employ "Kinect for Windows v2" to construct a system for judging concentration level that focuses on the movement of the neck and facial expression of a learner while learning.

The functions provided by "Kinect for Windows SDK v2" are listed below.

- Acquisition of color images
- Obtaining depth data
- Acquisition of infrared images
- People detection
- Detection of human joints
- Acquisition of voice
- Acquisition of voice direction
- Acquisition of face position and expression
- Obtaining modeled information of faces
- 3D scan
- Non-contact UI
- Gesture recognition
- Voice recognition

2.3 Detection Method

"Face Tracking" is used as a method of determining the concentration level of learners. "Face Tracking" is one of the functions provided by "Kinect for Windows SDK", and it can acquire various information on the face of a person detected by Kinect.

Information that can be obtained is as follows.

- Position of eyes, nose, mouth corners

- Bounding box surrounding the face area
- Face orientation
- Face condition (smile, presence of glasses, opening and closing of eyes, opening and closing of mouth and movement, whether or not the face is a full face).

Fig. 2 shows an example of how information of "Face Tracking" is displayed. The details of the displayed items are shown below.

- Pitch, Yaw, Roll: Acquisition of the direction of the learner's face in quaternion (quaternion). Pitch corresponds to the vertical movement of the face that is converted into the degree of rotation around the rotation axis. Yaw corresponds to the movement of the left or right of the face that is converted into the degree of rotation around the rotation axis. Roll corresponds to the movement of the face position that is converted into the degree of rotation around the rotation axis.
- Happiness: indicates whether or not the learner is smiling by Yes, Maybe (probably), No, Unknown (I do not know).
- Engaged: indicates whether or not the learner is facing the screen by Yes, Maybe (probably), No, Unknown (I do not know), based on the results of RightEyeClosed / LeftEyeClosed and LookingAway.
- Engaged: indicates whether or not the learner is facing the screen by Yes, Maybe (probably), No, Unknown (I do not know), based on the results of RightEyeClosed / LeftEyeClosed and LookingAway.
- WearingGlasses: indicates whether or not the learner wears glasses by Yes, Maybe (probably), No, Unknown (do not know).
- LeftEyeClosed / RightEyeClosed: indicates whether the learner's right eye / left eye is closed or not by Yes, Maybe (probably), No, Unknown (do not know).
- MouthOpen: indicates whether the learner's mouth is open or not by Yes, Maybe (probably), No, Unknown (do not know).

Fig.2 An example of display of "Face Tracking".

know).

- MouthMoved: indicates whether or not the learner's mouth is moving by Yes, Maybe (probably), No, Unknown (I do not know).
- LookingAway: indicates whether or not the learner is looking at the direction of Kinect by Yes, Maybe (probably), No, Unknown (do not know).

In this system, we check whether the learner is not moving by Pitch and Yaw values, whether the learner is facing the screen or not by Engaged, and whether the learner is awake or not by LeftEyeClosed / RightEyeClosed. Concentration level judgment is activated when Kinect recognizes the learner's face after 10 seconds from when Kinect is on, and ends when learning ends (90 minutes have elapsed). Each item of Face Tracking is updated every second, and measurement of concentration level are done every second.

Concentration measurement is performed as follows.

- Pitch, Yaw: When the value is between -20 and +20, we assume that the learner is facing to the screen. The thresholds are determined by preliminary experiments. Let Cp and Cy denote counters corresponding to Pitch and Yaw, respectively. When the value of Pitch (Yaw) of the learner falls within this range, we set 1 to Cp (Cy), otherwise 0.
- Engaged: Let Ce denote a counter corresponding to Engaged. When it is YES/Maybe then we set 1 to Ce, otherwise, 0.
- LeftEyeClosed / RightEyeClosed: Let Cl (Cr) denote a counter corresponding to LeftEyeClosed (RightEye Closed). When LeftEyeClosed is No/Unknown then we set 1 to Cl, otherwise, 0. Same to Cr.

Let Ca denote a variable indicating whether or not the learner is awake. We set 0 to Ca when both of Cl and Cr are 1, otherwise 0. The above judgment is performed every second. Let DC(i) denote the degree of concentration of item i (i=Cp, Cy, Ce, and Ca), and FC denote the final degree of concentration. When 10 seconds elapsed, the degree of concentration of each item is calculated by (1), and the final degree of concentration is calculated by (2).

$$DC(i)[\%] = (\text{the number of 1s of } i) / 10 * 100, \\ (i = Cp, Cy, Ce, \text{ and } Ca) \quad (1)$$

$$FC[\%] = (DC(Cp) + DC(Cy) + DC(Ce) + DC(Ca)) / 4 \quad (2)$$

For example, when the value of Pitch is 7, the value of Yaw is 7, the value of Engaged is 8, and the value of both eyes is 6, then the degrees of concentration DC(Cp), DC(Cy), DC(Ce), and DC(Ca) are 70%, 70%, 80%, 60%, respectively. The final degree of concentration FC is $(70 + 70 + 80 + 60) / 4 = 70\%$. When the final degree of concentration falls below

the predetermined threshold value, a message prompting concentration to the learner is displayed along with the beep sound. In this research, we set the threshold value to 60% according to the previous research⁽¹⁰⁾. Furthermore, when the learner's mouse does not move for more than 30 seconds, a message urging learning is displayed together with the beep sound. Records such as the degree of concentration and the number of warnings are written to a CSV-format file together with the date and time. By referring to this file, the teacher can grasp whether students focused and worked on learning.

The difference in the previous study⁽¹⁰⁾ from this study is as follows:

- (1) A web camera is used in the previous study, instead of Kinect to detect faces and eyes.
- (2) The degree of concentration is measured in a sliding window of 10 seconds. The sliding window is moved by one second in the previous study, while in this study the window is moved by ten seconds. .

3. Experiments

In experiments, we asked students of our university to take an English word test two weeks before they start e-learning. The test asks 76 problems among those registered in our e-learning system, and the score and correct answer rate are recorded. An example of the test problem is shown in Fig. 3. Next, we introduce a concentration judgment system using Kinect into the e-learning system, and have the students learn English words with the system. In order to have the students learn with competitive attitude and impatience, we show the score of the pre-study test to them. Additionally, in order for the students to study under equal conditions, we instruct the students to learn for 90 minutes in 3 days. When each student finishes learning on the third day, we have him take the English word test again. The test comprised the same 76 questions as the pre-study test, and the score and the correct answer rate are recorded. From the score before the test and the score after the test, we calculate

Fig.3 An example of the test problem asking English words corresponding to Japanese words.

the growth rate of the score defined by (3).

$$\text{Growth rate}[\%] = (S2 - S1) / (76 - S1) * 100, \quad (3)$$

where S1 and S2 is scores of the pre-study and post-study tests, respectively.

In the previous study, 14 students of our university are divided into two groups: one that warnings are issued when students are judged not concentrated and another that warnings are not issued. However, in this study we get only three students for experiments. Therefore, we compare the test results obtained in this study with the test results in the previous research. In order to distinguish these groups from one another, the group of students in this study is referred to as group A, the student group in the experiment with warning in the previous study is referred to as group B, and the student group in the experiment without warning is referred to as group C. Moreover, after completing learning of 3 days, we conducted a questionnaire asking about the feeling of use of the system and discuss from the questionnaire results how the system was felt by the students. The questionnaire consists of 16 questions in 5 grades (1 No ← ... → Yes 5) and one question with free description.

The average scores and the growth rates of group A, groups B, and C are summarized in table 1 for comparison. The full score point of the pre-study test and the post-study is 76. Looking at Table 1 for Group A, we observe that the average score of the post-study test is 65.67, the average growth point is 41.67, and the average growth rate is 80.26, which are higher than the results of the previous study. We can also see in Table 1 that the average growth point of the students in Group B is 40.43, and the growth rate is 71.25. Those of both group A and group B with warning are higher than the average growth point 27.29 and the growth rate 49.49 of group C without warning. From this result, the authors consider that warning during learning contributed to the continuation of concentration of learners and improved the performance of them as a result.

Next, we compare the data obtained from this system with the data of the previous study to examine the effect of warning. Table 2 summarizes the average degree of concentration for each item in this system. Table 3 summarizes the judgment data of group A, and those of group B and group C from the previous research. Table 2 shows high concentration of learners, while only “Engaged” item falls below the threshold of 60%. Item “Engaged” uses two factors to evaluate whether the eyes are closed or not and whether the eyes are facing the screen. Therefore, the judgment is more severe than other items, and the judgment result indicates that if a learner wears glasses, the system

erroneously recognizes the edges of the glasses as eyes. Additionally, in Table 3, we see that the proportion of the time during which a learner did not concentrate on learning in 90 minutes was 10.6%, and the learners were able to learn by concentrating nearly 90% of the learning time. Moreover, the number of warnings regarding the movement of the

Table 1. The average scores, the average growth points, and the average growth rates of Group A, B, and C.

	<i>Score of Pre-Study</i>	<i>Score of Post-Study</i>	<i>Growth Point</i>	<i>Growth Rate</i>
Group A	23.67	65.67	41.67	80.26
Group B	19.86	59.86	40.43	71.25
Group C	20.00	47.71	27.29	49.49

Table 2. The percentage, maximum, minimum, and average duration of no concentration in 90 minutes.

	<i>Data for No Concentration</i>	
Group A	Percentage of no concentration [%]	10.6
	Maximum continuous duration[sec]	112.2
	Minimum continuous duration[sec]	10.0
	Average continuous duration[sec]	20.0
Group B	Percentage of no concentration [%]	20.1
	Maximum continuous duration[sec]	340.8
	Minimum continuous duration[sec]	4.0
	Average continuous duration[sec]	47.3
Group C	Percentage of no concentration [%]	50.5
	Maximum continuous duration[sec]	460.1
	Minimum continuous duration[sec]	3.2
	Average continuous duration[sec]	41.0

Table 3. The average degree of concentration for each item in the system.

	<i>Pitch Cp</i>	<i>Yaw Cy</i>	<i>Engaged Ce</i>	<i>Closed Eyes Ca</i>	<i>Final Judgement FC [%]</i>
average concentra- -tion degree	0.90	0.87	0.52	0.81	78

mouse was 3.6, and the authors consider that there are no one who did not tackle the learning for a long time. Regarding the maximum duration in which learners did not concentrate, the maximum duration of Group A is 112.2 and the average duration is 20.0, which are lower than those of any group. Thus, we see that learners did not take long time to start concentrating again after out of concentration. Regarding the minimum duration time, we see that the numbers of groups B and C in the previous study are inferior to those of group A. This is because the concentration degree judgment was performed every second in the previous study, while this method is performed every 10 seconds. Thus, the inferiority does not necessarily mean bad results.

4. Questionnaire

We employed a questionnaire to see the feeling of our e-learning system with warning. Results of the questionnaire on the feeling of use of the system after the experiments for groups A and B are shown in Table 4. Firstly, looking at the average of items 1 in Table 4, we see that the values regarding the stress caused by monitoring with Kinect and Web camera are 3.00 for group A and 2.29 for group B, respectively, and group A felt more stress than group B. However, the scores of items 2, 4, and 6 are 4.67, 4.00, 4.00 for group A and 2.57, 3.57, 2.29, for group B, respectively, which indicates that the system with Kinect was more effective for keeping concentration than the previous system with a web camera. In addition, in the free description field of the questionnaire, there is a comment saying that the system with Kinect's monitoring was quite effective for sustaining the concentration. Additionally, regarding items 3 and 5, the group B is 3.43 and 4.43, whereas the group A shows the lower points of 2.67 and 2.00, and thus the warning message and the warning sound of the system with Kinect are much less stressed. As we see from the average of items 7 and 8 in Table 5 that the scores are 3.33, 2.67 for Group A, and 3.71, 2.00 for Group B, respectively, which indicates that the frequency that the system wrongly warn concentrated learners is decreased in the system with Kinect, while the frequency that the system wrongly does not warn un-concentrated learners is increased. In the free description field of the questionnaire, there was a comment saying that the stress of learners by monitoring and warning is reduced but there is room for improvement. Finally, the average scores of items 9 and 10 are 2.29, 3.57 in the previous study, while those are 3.00, 4.00 in this study. We see that the concentration level judgment system of this

Table 4. The average score points obtained from the questionnaire for Groups A and B.

Item No.	Score of Pre-Study	Group A	Group B
1	Did you feel stress by Kinect or a web camera?	3.00	2.29
2	Did Kinect or a web camera make you concentrate?	4.67	2.57
3	Did you feel stress by the warning message?	2.67	3.43
4	Did the warning message make you concentrate?	4.00	3.57
5	Did you feel stress by warning sounds?	2.00	4.43
6	Did warning sound make you concentrate?	4.00	2.29
7	Were there many cases of warnings when you concentrate?	3.33	3.71
8	Were there many cases of no-warning when you don't concentrate?	2.67	2.00
9	Do you want to use this system again?	3.00	2.29
10	Do you think this system is effective?	4.00	3.57

research showed an effect on the sustainment of concentration without becoming a burden on learners as compared with that of the previous research. As a reason for this, the authors consider as follows. In the previous study, once warning comes out, the warning continues for a few seconds, since the previous system estimates the degree of concentration in 10-second window from the present to the past 10 seconds and slides the window by one second. On the other hand, in this research we estimate the degree of concentration only once in 10 seconds, and thus warning comes out once in 10 seconds, which prevent the warning from becoming a large stress to learners. .

5. Conclusions

In this research, we conducted experiments of concentration level judgment using Kinect system in e-learning learning. We showed through experiments that with the concentration level judgment system using the Kinect, the test growth rate of tests was 80.26, which exceeds the growth rate 71.25 of the group with warnings and 61.69 of the group without warning in the previous study. Additionally, with this system the average duration of the state in which

the learner was out of concentration was 20.0 seconds, which exceeds the duration of 47.3 of the group with warning and 41.0 without warning in the previous study. From these results, we conclude that this system is effective for learning; with this system, learners can more intensively learn and can regain concentration more quickly than using the system of the prior research even if concentration is lost.

Future tasks include setting thresholds to reduce wrong recognition, a method to prevent wrong recognition caused by wearing glasses, and conducting experiments with more cooperators.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 16K00485.

References

- (1) Ministry of Internal Affairs and Communications(2014) "White paper Information & Communications in Japan," <http://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2014/2014-index.html>
- (2) A. Saitoh, T. Nishida, M. Nakanishi, S. Yasutome, K. Baba, Y. Shigehiro, A. Harada, N. Yamai, and T. Matsuura, "Classroom support and system administration support for large scale educational computer system," Trans. IEICE Japan, Vol.J84-D-I, No.6, p.956–965, 2001.
- (3) Kansai Business Information Inc., "E-learning," <https://www.i-netschool.com/e-learning/index.html>
- (4) Proseeds Inc., "Merits and demerits of e-learning," <http://www.pro-seeds.com/trend/meritdemerit.html>
- (5) Y. Yamashita, K. Takahashi, H. Ueda, and T. Miyahara, "Construction and analysis of web-based e-learning system for exercises," Proc. Chugoku Branch Conference of Electrical and Information Related Institutes, Yamaguchi, Japan, p.552, 2004.
- (6) K. Shimada, K. Takahashi, and H. Ueda, "Experiments on an e-learning system for keeping the motivation," Conf. Rec. 2011 SSITE Int. Conf. Computers and Advanced Technology in Education, Beijing, China, p.161–168, 2011.
- (7) M. Takeue, K. Shimada, K. Takahashi, and M. Inaba, "Experiments of displaying images to keep the motivation in e-learning," Conf. Rec. IEEE International Conference on Systems, Man, and Cybernetics, Seoul, Korea p.120–125, 2012.
- (8) J. Kawaguchi, T. Ito, K. Takahashi, and M. Inaba, "affect of what to do during the rest time in e-learning," Proc. of the 2016 IEICE Society Conference, p.151, 2016.
- (9) Tetsuharu Kurakake, "A report on e-learning classes in which a LMS/CMS is used," Research Reports of Kagoshima National College of Technology, Vol.44, p.27–30, 2009.
- (10) K. Takahashi and M. Ueno, "Improvement of detection for warning students in e-learning using web cameras," Proc. 18th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, p.747-756, 2014.
- (11) Microsoft Corporation, "Kinect for Windows Product Blog," <https://blogs.msdn.microsoft.com/kinectforwindows>
- (12) K. Nakamura, T. Sugiura, T. Takada, and T. Ueda, "KINECT for Windows SDK programming," Shuwa System Co. Ltd., 2015.
- (13) Microsoft Corporation, "Microsoft Developer Network," <https://msdn.microsoft.com/ja-jp/library/dn782034.aspx>