Journal of Mobile Multimedia, Vol. 11, No. 3&4 (2015) 205–212

TOWARDS ACTIVITY RECOGNITION OF LEARNERS IN ON-LINE LECTURE

HIROMICHI ABE, TAKUYA KAMIZONO, KAZUYA KINOSHITA,
KENSUKE BABA, SHIGERU TAKANO, and KAZUAKI MURAKAMI
Kyushu University
Motooka 744, Nishi-ku, Fukuoka, 819-0395, Japan
hiromichi.abe@soc.ait.kyushu-u.ac.jp

Understanding the states of learners at a lecture is useful for improving the quality of the lecture. A video camera with an infrared sensor Kinect has been widely studied and proved to be useful for some kinds of activity recognition. However, learners in a lecture usually do not act with large moving. This paper evaluates Kinect for use of activity recognition of learners. The authors considered four activities for detecting states of a learner in an on-line lecture, and collected the data with the activities by a Kinect. They repaired the collected data by padding some lacks, and then applied machine learning methods to the data. As the result, they obtained the accuracy 0.985 of the activity recognition. The result shows that Kinect is applicable also to the activity recognition of learners in an on-line lecture.

Keywords: Activity recognition, Kinect, data mining, e-learning.

1 Introduction

A massive open online course (MOOC) [1] has been attracting attention as an advanced model of learning. Generally, understanding the states of learners in a lecture is useful for improving the quality of the lecture. Some states of learners are estimated by exams conducted after the lecture as a degree of understanding, or some lecturers might try to know them by individual communications to learners at the lecture. However, especially in a massive or an on-line lecture, it is difficult to know the states of each learner at the lecture. Detecting learners' states automatically and in real time is expected to innovate in the current style of learning.

In this paper, we are trying to recognize activities of a learner to detect some states of the learner. Mukunoki et al. [2] found a relation between activities of learners in a lecture and a degree of understanding. They categorized some activities into states of a learner such as "concentrated" and "distracted", which means that the relation between learners' activities and understanding was obtained by considering learners' states. On the assumption that learner's activities correspond to their states, recognizing learners' activities is regarded as understanding the states of learners in a lecture.

Activity recognition has been widely explored. For activity recognition of learners, Mota and Picard [3] intended to recognize learner's states by measuring their posture. Chaouachi and Frasson [4] showed a relation between learners' electroencephalography data and their response time. For general activity recognition using video data, recently, a spread of depth sensors enabled us to analyze the subject in terms of 3-dimensional coordinates [5]. By video data with depth information, some motions and facial expressions can be recognized with

high accuracy. The target of our study is learners who are sitting in front of their respective monitors with small moving. The purpose of this paper is to confirm that video data with depth information can realize activity recognition also for learners in an on-line lecture.

In this paper, we evaluate the data collected by a camera with an infrared sensor Kinect [6] for activity recognition. In our experiment, the subjects in front of a monitor perform the four activities: meditating with the eyes closed, reading texts on the monitor, looking away with moving the face, and sitting at a far point from the monitor. Since a Kinect sometimes fails to capture data, we conduct a preprocessing to repair the data. Then, we apply the two machine learning techniques, K-nearest neighbor algorithm (K-NN) and support vector machine (SVM) [7, 8], to the processed data, and investigate the accuracy of the activity recognition. The result of our experiment shows that Kinect is applicable to the activity recognition of learners in an on-line lecture.

The rest of this paper is organized as follows. Section 2 describes the methods of our experiments of activity recognition and formalizes the criteria for the accuracy. Section 3 reports the results of the experiments. Section 4 shows considerations about the results and future directions of our study.

2 Methods

We collected video data by a Kinect when subjects performed four kinds of activities. Then, we applied a machine learning algorithm to the collected data to recognize the activities, and investigated the accuracy of the recognition.

2.1 The Activities

In this study we considered the following four activities of a learner sitting in front of a monitor:

a. Meditating with the eyes closed,

b. Reading texts on the monitor,

c. Looking away with moving the face,

d. Sitting at a far point from the monitor.

Figure 1 shows the images of the four activities performed by a subject in the experiments. The target of our study is a learner in an on-line lecture. We considered that the four activities correspond to the conditions of a learner:

- Dozing,
- Being concentrated to the lecture,
- Being distracted,
- Doing completely different things from the lecture,

respectively.

We practically observed some actual lectures in our university, and found that most learners were taking one of the three activities a, b, and c, or having notes with hanging down

a b c d

Fig. 1. The four activities, (a) meditating with the eyes closed, (b) reading texts on the monitor, (c) looking away with moving the face, and (d) sitting at a far point.

the head. In a preliminary experiment, the subject with the activity "having notes on the lecture" got out of the scope of the camera. The activity d is considered as a possible activity in an on-line lecture. Therefore, we considered the four activities as the target activities of learners in our experiments.

2.2 Data Collection

A Kinect can be set on a monitor as the left-hand in Figure 2, hence it is suitable for capturing video data of a learner at an on-line lecture.

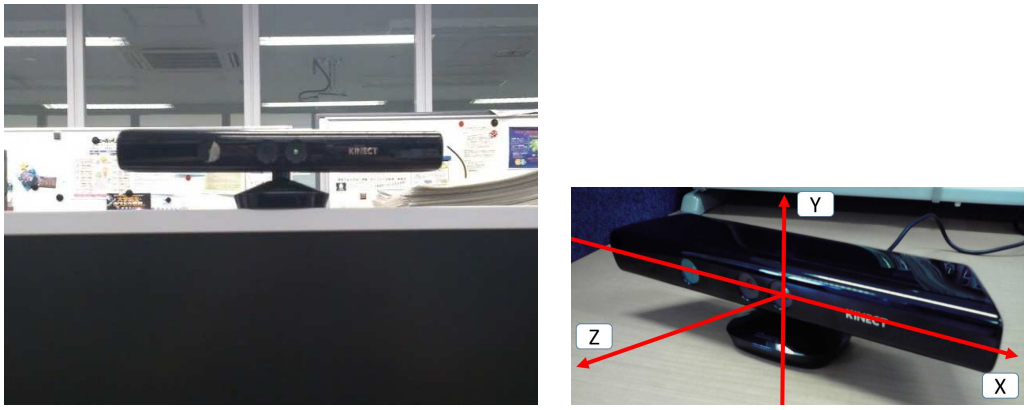

Fig. 2. A Kinect set on a monitor (left) and the 3-dimensional coordinates defined by a Kinect (right).

A Kinect can measure the 3-dimensional data of the subject using depth data captured by the infrared sensor. In our experiments, we used the following 17 kinds of data Kinect can measure:

- The 3-dimensional coordinates of the face of the subject,
- The changes of the 3-dimensional coordinates of the face of the subject,
- The 3-dimensional angular coordinates of the face of the subject,

- The 2-dimensional coordinates of the upper and lower parts of the eyes of the subject (8 dimensions).

The first data are captured on the coordinates defined as the right-hand in Figure 2. The second data are changes on the first data. The third data and fourth data capture the motions of the face and the eyes as Figure 3.

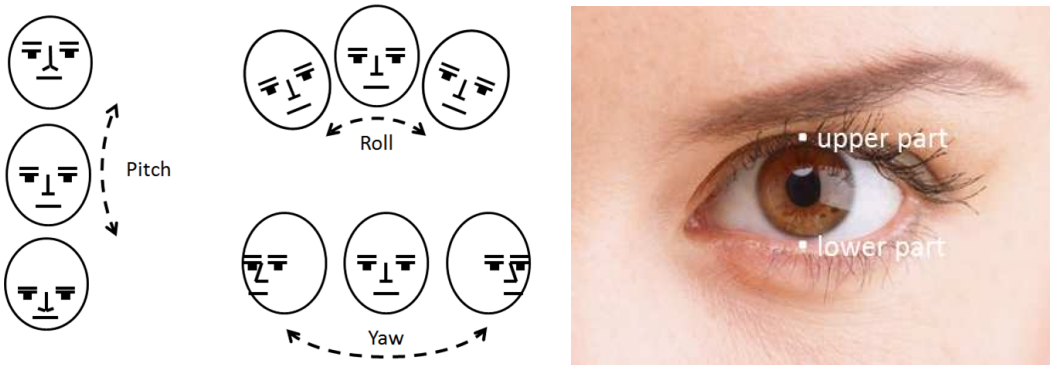

Fig. 3. The 3-dimensional angular coordinates for a face (left) [9] and the upper part and the lower part of an eye (right).

In the experiments, two subjects performed each activity for about 1 minute as a single trial. One of the subjects repeated the activities twice. Generally, a Kinect can capture data in about 6 times per a second at unfixed intervals. Therefore, we collected about 300 vectors for each trial.

2.3 Data Analysis

We tried to recognize the four activities by applying K-nearest neighbor algorithm (K-NN) [7] from the collected data.

A *sample* s is a pair of a vector x and the label y of an activity. Now, we consider the situation that a set of samples $S = \{(x, y)\}$ is given as training data and we estimate the activity of a new vector x_t. Let $d(u, v)$ be the Euclidean distance between vectors u and v. Then, in K-NN,

- A subset S' of S is chosen such that $|S'| = K$ and the largest $d(x_t, x)$ in the K x's in S' is smaller than or equal to the smallest $d(x_t, x)$ in $S \setminus S'$;
- The majority in the K y's in S' is predicted as the activity for x_t.

In the case where there is a tie of the majority, in our experiments, we chose the output randomly from the majorities.

In our experiments, we considered the mean and the variance of each continuous (non-overlap) 5 values for each element in the collected data. Therefore, we obtained 60 34-dimensional vectors from 300 17-dimensional vectors.

With the sample set, we conducted leave-one-out (LOO) cross-validation [7] for the K-NN. We considered the standard criteria for the correctness of a recognition algorithm. For an activity a, let TP_a, FP_a, and FN_a be the numbers of the samples that

Table 1. The numbers of sample data for four activities with three trials.

	Collected	Includes N/A
a. Meditation	168	102
b. Reading texts	157	102
c. Looking away	184	78
d. Sitting at a far point	204	9
Total	713	291

- The predicted activity was a and the actual one was a (TP_a),
- The predicted activity was a and the actual one was not a (FP_a),
- The predicted activity was not a and the actual one was a (FN_a).

Then, the *precision*, the *recall*, and the *F-measure* for a are defined as

$$P_a = \frac{TP_a}{TP_a + FP_a}, \quad R_a = \frac{TP_a}{TP_a + FN_a}, \quad \text{and} \quad F_a = \frac{2P_aR_a}{P_a + R_a},$$

respectively. Let A be the set of the activities for the target activity recognition. Then, the *accuracy* of the recognition algorithm is defined to be

$$\frac{\sum_{x \in A} TP_x}{\sum_{x \in A}(TP_x + FP_x)}.$$

Note that the numerator is the number of the correct predictions and the denominator is the total number of the tests.

3 Results

Table 1 shows the numbers of the obtained vectors in this experiments. The total number of the sample data was 713. The collected data included some "not available (N/A)" values. When we calculated the mean and the variance of 5 values, the value was treated as a "N/A" if all the 5 values were "N/A". There were 291 vectors that included "N/A" in at least one element.

Table 2 shows the result of LOO cross-validation of the activity recognition based on K-NN with the sample set. In the table, a, b, c, and d refer the activities meditation, reading texts, looking away, and sitting at a far point, respectively. The value of K in K-NN was fixed to 3. In the table, the value "N/A" of the predicted activity is the output for a vector that includes at least one "N/A" element. Additionally, + and − refer the cases where the output "N/A" is considered as a fault and ignored, respectively. The bold numbers are the accuracies for the two cases, respectively. The accuracy of the recognition algorithm was 0.578 if we regarded the output "N/A" as a fault, and 0.976 if we ignored them.

4 Discussion

4.1 Major Conclusion

We found that we could recognize the four activities from the data collected by a Kinect.

Table 2. The result of LOO cross-validations for K-NN with the collected data by a Kinect.

		Actural				Precision	F-measure	
		a	b	c	d		+	−
Predicted	a	61	0	2	0	0.968	0.528	0.946
	b	0	55	1	0	0.982	0.516	0.991
	c	4	0	101	0	0.962	0.699	0.957
	d	1	0	2	195	0.985	0.970	0.992
	N/A	102	102	78	9	-	-	-
Recall	+	0.363	0.350	0.549	0.956	**0.578**		
	−	0.924	1.000	0.953	1.000	**0.976**		

Table 3. The result of LOO cross-validations for K-NN with the data with padding N/A by the nearest values.

		Actural				Precision	F-measure
		a	b	c	d		
Predicted	a	163	1	4	2	0.959	0.964
	b	0	155	3	0	0.981	0.984
	c	5	1	176	2	0.957	0.957
	d	0	0	1	200	0.995	0.988
Recall		0.970	0.987	0.957	0.980	**0.973**	

The accuracy of the activity recognition was 0.578 when we treated vectors that contains "N/A" as failures. However, in some applications of the activity recognition, we can suppose the situation that we ignore the output "N/A" and wait the next data. For this situation, we obtained the high accuracy 0.976. Since a sample vector corresponds to 5 pieces of data that collected 6 pieces per second, we have only to wait about 0.8 seconds for a single "N/A" output.

The number of the output "N/A" was large for the activities "meditation" and "reading texts" compared with the other activities. Therefore, we consider that "N/A" would arise for activities with small moving. The number of "N/A" was small for the activity "sitting at a far point". Since the distance between the sensor and the subject in the activity was different from the other activities, we expect that the frequency of the "N/A" can be controlled by tuning the position of the sensor.

4.2 Key Findings

The accuracy is expected to be improved by padding the lacks of data with appropriate values instead of ignoring the lacks. We modified the sample data such that any "N/A" value is replaced by the value of the same element of the previous vector in the order of the captured time. Table 3 shows the result of LOO cross-validation of the activity recognition based on K-NN with the padded sample set. The K in K-NN was also 3. As we expected, the accuracy was better than 0.578 and worse than 0.976.

The accuracy was obtained by the simple algorithm for the activity recognition. We expected higher accuracy by applying other strong algorithms. Table 4 shows the result of LOO cross-validation of the activity recognition based on SVM with the padded sample set. We used the function `ksvm` in R [10] with the Gaussian kernel and the parameter was

Table 4. The result of LOO cross-validations for SVM with the data with padding N/A by the nearest values.

		Actural				Precision	F-measure
		a	b	c	d		
Predicted	a	165	1	1	0	0.988	0.985
	b	0	153	0	0	1.000	0.987
	c	3	3	183	3	0.953	0.973
	d	0	0	0	201	1.000	0.993
Recall		0.982	0.975	0.995	0.985	**0.985**	

optimized in terms of the accuracy. As we expected, we obtained a slight improvement of the accuracy.

4.3 Future Directions

We expect higher accuracy by considering pattern-based and time-series analyses. We have to consider on-line or real-time processing of the algorithms for activity recognition of learners in a lecture.

Another future work is combining the activity recognition by a Kinect with that by other sensor data. For example, we conducted a similar experiment of activity recognition of learners with an electroencephalograph [11].

In the experiments in this paper, the number of the subjects was only two. We are going to conduct the activity recognition in an actual lecture in our university. In a large scale experiment, data may have some differences between individuals. In addition to the activity recognition, we are going to consider identification of learners.

5 Conclusion

We evaluated video data collected by a Kinect for activity recognition of learners in a lecture. We conducted experiments of activity recognition for the four activities: meditating with the eyes closed, reading texts on the monitor, looking away with moving the face, and sitting at a far point from the monitor. We applied the preprocessing and the two machine learning methods to the collected data, and the accuracy was 0.985 when we used SVM to the repaired data. Thus, we confirmed that Kinect is applicable for the activity recognition of learners in an on-line lecture.

Acknowledgment

This work was partially supported by a joint research with Panasonic Corporation from 2014 to 2015.

References

1. "MOOCs Directory," http://www.moocs.co/. [Accessed Oct. 2014].
2. M. Mukunoki, M. Uematsu, and M. Minoh (2013), "Analyzing the relationship between learners' comprehension and behavior based on item response theory," *Japanese Society for Information and Systems in Education*, vol. 30, no. 1, pp. 65–76.

3. S. Mota and R. W. Picard (2003), "Automated posture analysis for detecting learner's interest level," *Proc. Computer Vision and Pattern Recognition Workshop 2003 (CVPRW'03)*, IEEE, p. 49.
4. M. Chaouachi and C. Frasson (2010), "Exploring the relationship between learner EEG mental engagement and affect," *Proc. 10th International Conference on Intelligent Tutoring Systems (ITS 2010), Part II*, Lecture Notes in Computer Science, vol. 6095. Springer-Verlag, pp. 291–293.
5. L. Chen, H. Wei, and J. Ferryman (2013), "A survey of human motion analysis using depth imagery," *Pattern Recognition Letters*, vol. 34, pp. 1995–2006.
6. "Kinect for Windows," http://www.microsof.com/enus/kinectforwindows/. [Accessed Oct. 2014].
7. C. M. Bishop (2006), *Pattern Recognition and Machine Learning*. Springer.
8. T. Hastie, R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
9. "Microsoft Developer Network" http://msdn.microsoft.com/en-us/library/jj130970.aspx. [Accessed Oct. 2014].
10. "The R Project for Statistical Computing," http://www.r-project.org/. [Accessed Oct. 2014].
11. H. Abe, K. Baba, S. Takano, and K. Murakami (2014), "Towards activity recognition of learners by simple electroencephalographs," *Proc. Information Systems and Design of Communication (ISDOC 2014)*, ACM, pp. 161–164.