Behavior Detection and Analysis for Learning Process in Classroom Environment

Mengling Yu, Jing Xu, Jinrong Zhong, Wei Liu, Wenqing Cheng
School of Electronic Info. and Comm., Huazhong University of Science and Technology, Wuhan 430074, China
E-mail:{yu_mengling, xujing, zhongjr, liuwei, chengwq}@hust.edu.cn

Abstract—**Classroom observations have been widely used in education over the past couple of decades to measure effective teaching practice. The traditional observation methods rely on human observers, which are short of scalability and objectivity. In this paper, we implement a kind of automatic behavior measurement system, which utilizes the Microsoft Kinect devices to record the students' performance in classroom. Several Kinect devices are installed under the ceiling of one classroom. The facial images of attended students are collected and recognized. The typical gestures of students (such as sitting, raising hand, standing, sleeping and whispering) are also detected and recorded. A queue-based analysis engine is proposed to distinguish the meaningful learning behaviors from those pointless actions. Experiment results show that this system can be utilized to measure the students' active behaviors in typical learning processes, which will be helpful for the analysis of behavioral engagement in classroom teaching.**

I. Introduction

Classroom observations have been widely used in education over the past couple of decades to measure effective teaching practice. For example, [1] studied the off-task behavior of elementary students; [2] proposed a protocol, namely BERI (Behavioral Engagement Related to Instruction) to describe the in-classroom behaviors of college students. It is notable that, these traditional observation methods rely on human observers, which is difficult to extend to large-scale or long-term measurement.

With the development of intelligent teaching system, it became popular to analyze the in-system behavior of students. For instance, [3] reported the results of deploying Cliker in classroom interactions; [4] studied the students' off-task behavior in one intelligent tutoring system. However, these approaches rely on the deployment of intelligent teaching system and can not directly adopted in the natural classroom.

In order to measure the students in the school classroom environment, the researchers proposed to utilize image processing and computer vision technologies. For example, a facial expression recognition method was proposed to monitor the students' attendance in [5]. Considering the dense sitting positions of the students, it is not easy for the classical two-dimensional image processing techniques to recognize the faces and the body gestures.

Recently, infrared sensors are proposed in the motion sensing technology, which provides depth images. The most representative device is the Microsoft Kinect. It is able to process the depth images and even report the human skeleton data. It can be used to detect the gestures of the hand [6], the limb [7] and the body [8]. In our previous work [9], we proposed to use Kinect to detect the head pose of single student in the course.

In this paper, we focus on the behavior measurement problem in classroom. We implement a kind of automatic behavior measurement system, which utilizes the Microsoft Kinect devices to record the behavior of a number of students in one classroom. Multiple Kinect devices are installed under the ceiling of one classroom. The facial images of attended students are collected and recognized. The typical gestures of students (such as sitting, raising hand, standing, sleeping and whispering) are also detected and recorded. A queue-based analysis engine is proposed to distinguish the meaningful learning behaviors from those pointless actions. The experiment results show that it is feasible to implement face recognition and gesture detection for a group of students with multiple Kinect devices.

The rest of the paper is organized as follows. Section II introduces the measurement system framework. The experiment results are provided in Section III. Section IV discusses the problems and limitations of implementing Kinect in classroom environment, followed by the conclusion in Section V.

II. Learning Behavior Measurement System

A. System Overview

In order to enlarge the detection range of each Kinect, four Kinect 2.0 devices are installed under the ceiling of the classroom. The height of installed Kinect is 2.5 meters to the ground. Each Kinect is responsible to monitor two to three rows of students, which is around 3 meters in distance. The illustration of the deployed system is shown in Fig.1.

We divides the measurement system into three modules, including the data acquisition module, the behavior detection module, and the behavior analysis module. The main framework of this system is given in Fig.2.

Firstly, the data acquisition module controls the Kinect devices and collects the sensing data, including the depth image, the skeletal data and etc. (One sample measurement result is provided in Fig.3.) Then these data is fed into the behavior detection module. The facial images of attended students are collected and recognized. The typical gestures of students (such as sitting, raising hand, standing, sleeping

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

Fig. 1. Deployment of Multiple Kinect in one classroom

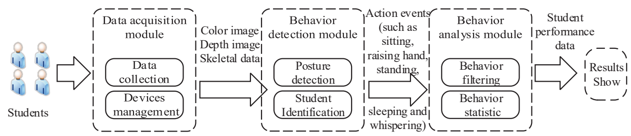

Fig. 2. Framework of the Behavior Measurement System

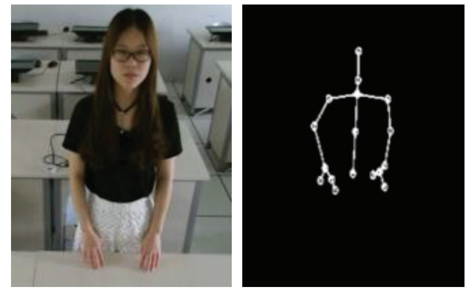

Fig. 3. Examples of Kinect measurement (color and skeleton images)

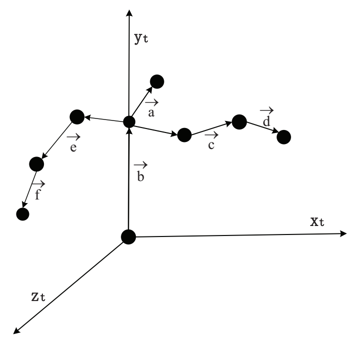

Fig. 4. Skeleton feature vector of Upper body

and whispering) are also detected and recorded. The outputs of behavior detection module are the detected action events, which may be meaningful learning behaviors but also may be pointless actions. Finally, the behavior analysis module checks the temporal relations of these actions, and filters out the useful learning behaviors in the context of one learning process.

In the following, we report the details of the behavior detection and analysis modules:

B. Behavior Detection Module

This input is the original data acquired by Kinect. Color image is used for face recognition, head deflection and skeletal position are used for gesture recognition. This module transforms the original measurement data into the basic action events through the methods of threshold identification, classification recognition, etc. The resulted action events include sitting, raising hand, standing, sleeping, and whispering.

1) Facial Recognition: Facial recognition is to identify the behavior object. The recognition of face consists of three steps: face detection, feature extraction and authentication. The procedure of face detection is to detect the face and locate the face area in the image frame extracted from the Kinect video. Feature extraction derives a vector as the face feature, such as LBP (Local Binary Pattern). The final recognition is done by comparing the feature vector extracted with those registered and stored in the database.

2) Gesture Recognition: After collecting the skeletal coordinates of the human body, the skeleton of upper human limbs can be obtained, and then extract the feature vector of the skeleton. The movements of the upper limbs focus on the human torso and both arms. The corresponding structural vectors are expressed as $\vec{a}, \vec{b}, \vec{c}, \vec{d}, \vec{e}, \vec{f}$(shown in Fig.4). According to the predefined discrimination conditions, the gesture at that time can be detected, by discriminating the combination of the vector angles and the modulus ratios obtained at any time.

C. Behavior Analysis Module

This module analyzes the behavioral events of students' learning process and obtains the results of students' learning behavior for exhibition.

After the process of face recognition and gesture recognition, the original data collected by the acquisition module has been transformed into the specific action events. For given event, its data structure contains five attributes, including event type, event generation object, start time, end time, and the event duration. Due to some of these action events are pointless and one single action event is not enough to make decision of the behavior, we extend the detection engine in our previous work [10], and propose a queue-based analysis engine to filter out the useful learning behaviors. The basic implementation of the analysis engine is presented in Fig.5.

According to the time of events occurs, we can regard these events as a discrete time series, represented by e_i ($i = 1, 2, 3, \cdots$),where e_i stands for one of the five typical gestures. The series e_i is the input of the analysis engine. The series are divided into several queues and grouped by event generation object after being fed to the analysis engine. And then filters determine whether these actions can form a meaningful behavior on the basis of the events' duration and frequency. It is notable that behavior determination is queue-based, and the criterion varies with queue.

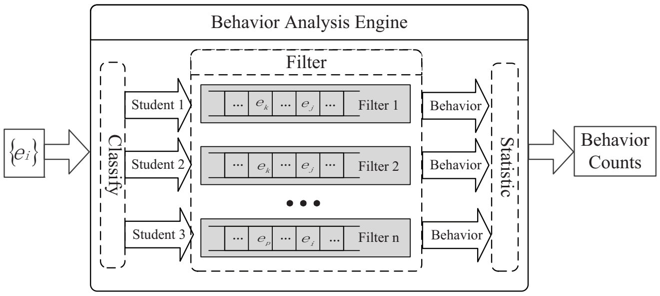

Fig. 5. Event Queue in Behavior Analysis Engine

Fig. 6. Result of face recognition

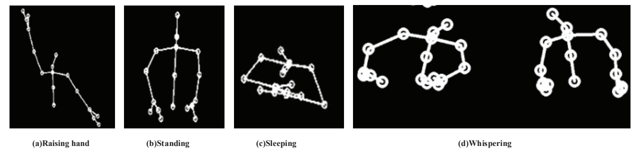

Fig. 7. Result of posture recognition

III. Experiment Results

A. Scenario and Settings

We built a classroom scene and invited 16 student volunteers to participate in a 45-minute course. During the completely teaching process and we used the measurement system to detect and record five specific behaviors (sitting, raising hand, standing, sleeping and whispering) of students and identify the generation object of each event. All results of measurement would be upload to the database in real time.

B. Result of Face Recognition

Fig.6 is a representative result of face recognition. The left image is the truncated picture of a student. After face recognition, a blue rectangular box with the size information is annotated in the face area, and the corresponding student's ID is marked. Finally, the correctness of face recognition is validated by comparing recognition result and the actual picture related to the student's ID.

C. Result of Behavior Detection

Fig.7 illustrates the skeleton map corresponding to five behaviors common in nature classroom. As shown in Fig.7, the results from left to right are successively (a) shows raising hand,(b) shows standing, (c)shows sleeping, (d) shows whispering. The skeleton map of sitting is similar to standing.

The skeleton maps are then detected by predefined discrimination conditions and transferred to the action events. Some event records are shown in Table I.

D. Result of Behavior Analysis

In behavior detection module, we determine action by the gesture of skeleton, but there are some gesture of pointless actions similar to meaningful actions. At this point, we have to analyze the adjacent several action events to determine whether it is a meaningful behavior. For example, when a student turns his head, he may be shaking head, or whispering to others. In order to detect this confusing action, not only the event duration and frequency, but also the neighboring students' actions should be checked. As shown in Table I, one student (with ID U201705) was detected to whisper to others at 2017-3-15 10:53:03, then we found the student next to him (whose ID is U201706) having the same action at the same time, and this action lasted 37 seconds, so we inferred these two students were whispering from 2017-3-15 10:53:03 to 2017-3-15 10:53:40.

After filtering out all the meaningful learning behaviors, we counted the typical behavior of the students in the classroom. Firstly, we divided the above four behaviors(except sitting) into two categories: engaged (raising hand, standing) and disengaged (sleeping, whispering). Then we counted each student's frequency of these two types of behavior. Finally, we gave out the list of students having top five numbers of engaged and disengaged behavior, as shown in Fig.8 and Fig.9.

In Fig.8 and Fig.9, X-axis represents frequency of this behavior. The two bars represent the frequency that raising hand and standing have occurred in Fig.8, while representing whispering and sleeping in Fig.9. The first one on the list is the student with the top one numbers of engage behaviors. The rank is based on the total frequency of all kind of engaged behaviors or disengaged behaviors.

IV. Discussion

The testing results verify that our behavior measurement system can effectively detect and analyze the behavior in the school classroom environment. However, the system should be improved in the following aspects:

Firstly, our experiment was built in a small classroom, in which sixteen students were monitored by four Kinect devices. If we extend the system to a larger scenario, more Kinect should be equipped and will cause higher deployment cost. It is not scalable to deploy more Kinect devices under the ceiling in one classroom.

Secondly, we suspend the Kinect devices under the ceiling of the classroom, which brings a coordinate correction problem. The data collected by Kinect device is recorded in a coordinate system taking itself as the coordinate origin. It is required to convert the gesture data in oblique direction to the normal classroom coordinate system. What's more, it is difficult to distinguish some similar actions (such as bowing, writing and etc) because of the oblique direction.

TABLE I

SAMPLE DETECTION RESULTS STORED IN DATABASE

No.	Behavior	StudenID	e_{start}	e_{end}	e_{dur}(second)
1	hands up	U201702	2017-3-15 10:44:05	2017-3-15 10:44:12	7
2	standing	U201702	2017-3-15 10:44:25	2017-3-15 10:45:30	65
3	whispering	U201705	2017-3-15 10:53:03	2017-3-15 10:53:40	37
4	whispering	U201706	2017-3-15 10:53:03	2017-3-15 10:53:40	37
5	sleeping	U201711	2017-3-15 11:03:20	2017-3-15 11:10:00	400

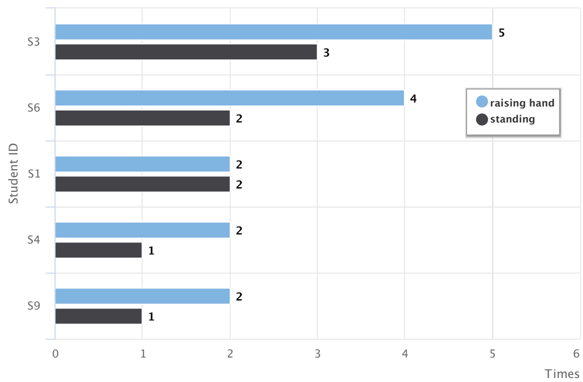

Fig. 8. Ranking of engaged behavior

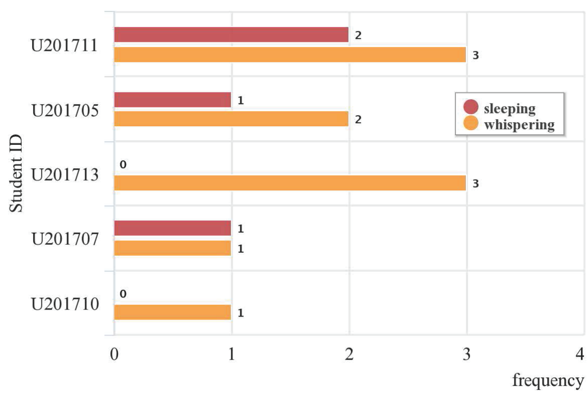

Fig. 9. Ranking of disengaged behavior

Finally, the measurement system is designed for detect the meaningful learning behavior, which is generally affected by the physical capabilities and habits of students. Thus, the criteria of behavior detection and analysis should vary with each individual.

V. CONCLUSION

The learning behavior is an important reference for teaching practice in education. In this paper, we implement a kind of automatic behavior measurement system to detect and analyze the learning behavior of students in the school classroom environment. Some representative gesture of students (such as sitting, raising hand, standing, sleeping and whispering) are detected and recorded. Specially, a queue-based analysis engine is proposed to distinguish the meaningful learning behaviors from those pointless actions. Experiments built in small classroom with sixteen students were conducted and the results reveal that our system can be utilized to measure the students' active behaviors in typical learning processes.

ACKNOWLEDGMENT

This work has been in part supported by the National Key Technology R&D Program of China under Grant 2015BAH33F04-05.

REFERENCES

[1] K. E. Godwin, M. V. Almeda, M. Petroccia, R. S. J. D. Baker, *et al.*, "Classroom Activities and Off-task Behavior in Elementary School Children," in *Proc. of CogSci'13*, 2013, pp. 2428–2433.

[2] E. S. Lane and S. E. Harris, "Research and Teaching: A New Tool for Measuring Student Behavioral Engagement in Large University Classes," *Journal of College Science Teaching*, vol. 44, no. 6, pp. 83–91, 2005.

[3] J. E. Caldwell, "Clickers in the Large Classroom: Current Research and Best-Practice Tips," *CBE Life Science Education*, vol. 1, no. 6, pp. 9–20, 2007.

[4] R. S. J. D. Baker, "Modeling and Understanding Students' Off-task Behavior in Intelligent Tutoring Systems," in *Proc. of ACM CHI'07*, 2007, pp. 1059–1068.

[5] A. A. Tamimi, O. N. A. Alallaf, and M. A. Alia, "Real-Time Group Face-Detection for an Intelligent Class-Attendance System," *International Journal of Information Technology and Computer Science*, vol. 7, no. 6, p. 66, 2015.

[6] S. Sempena, N. U. Maulidevi, and P. R. Aryan, "Human Action Recognition Using Dynamic Time Warping," in *Proc. of IEEE ICEEI'11*, 2011, pp. 1–5.

[7] M. Raptis, D. Kirovski, and H. Hoppe, "Real-time Classification of Dance Gestures from Skeleton Animation," in *Proc. of ACM SCA'11*, 2011, pp. 147–156.

[8] L. Xia, C. C. Chen, and J. K. Aggarwal, "Human Detection Using Depth Information by Kinect," *Applied Physics Letters*, vol. 85, no. 22, pp. 5418–5420, 2011.

[9] Z. Fan, J. Xu, W. Liu, F. Liu, *et al.*, "Kinect-based Dynamic Head Pose Recognition in Online Courses," in *Proc. of IEEE IMCEC'16*, 2016, pp. 448–453.

[10] W. L. Z. Fan, J. Xu and W. Cheng, "Gesture based misbehavior detection in online examination," in *Proc. of IEEE ICCSE'16*, 2016, pp. 234–238.