Multimed Tools Appl (2018) 77:28749–28775
https://doi.org/10.1007/s11042-018-6017-2

Student engagement study based on multi-cue detection and recognition in an intelligent learning environment

Yuanyuan Liu[1] · Jingying Chen[2] · Mulan Zhang[2] · Chuan Rao[2]

Received: 16 July 2017 / Revised: 24 March 2018 / Accepted: 16 April 2018 /
Published online: 5 May 2018

Abstract Student engagement has great impact on learning performance. It's necessary to investigate student engagement objectively from learning behavior. In this paper, we propose a student engagement study approach in an intelligent learning environment, which automatically detects and analyses multiple learning behavioral cues based on five modules, i.e., attendance management, teacher-student (T&S) communication, visual focus of attention (VFOA) recognition, smile detection and engagement analysis. Attendance management matches the student's identity and locates his/her profile using face recognition. T&S communication provides an additional channel of Question and Answer (Q&A) between a teacher and students for students' behavioral engagement analysis via their cell phones. VFOA recognition is used to recognize students' cognitive engagement through capturing students' attention based on the estimated head poses, visual environment cues and prior states in class. Smile detection achieves students' affective engagement through spontaneous smile expression classification. Finally, a tree-structural engagement model is proposed to decide student engagement based on multi-cues of one's behavioral, cognitive and affective engagement. We thoroughly evaluated each module for engagement study on some public available datasets and practical video sequences in class applications. The experimental results suggest that the proposed approach can automatically detect and analyze student class engagement objectively and effectively.

Keywords Multi-cue behavior detection · Class engagement study · VFOA recognition · Smile detection · Intelligent learning environment

✉ Jingying Chen
chenjy@mail.ccnu.edu.cn

Yuanyuan Liu
liuyy@cug.edu.cn

[1] Faculty of Information Engineering, China University of Geosciences, Wuhan, China

[2] National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China

1 Introduction

1.1 The overview of student class engagement study in the intelligent learning environment

Student engagement study is important for student satisfaction and learning effectiveness assessment in class. It is known to benefit learning and is a very important component of teaching and learning [4, 34]. Most researches consider that student class engagement is the closet relative to student attendance, class interactive behaviors, attention targets and affective states in learning process [36]. In a traditional class environment, student class engagement study mostly can be obtained by designed questionnaire surveys after class. Yet it lacks real-time learning surveillance and analysis in class. Moreover, many obstacles inhibit real-time learning behavioral data surveillance and analysis in class, such as limited class time, rigid seating arrangement and students' reservations to speak out, and so on [7, 34].

Assessing a student's engagement has never been easy. In recent years, technologies embedded in an class learning environment can provide opportunities for teachers and students to obtain efficient learning data for engagement study, e.g., WebAnn, Epost, intelligent E-learning system, affective aided system, learning attention tracking system, etc., [9, 14, 15, 17, 21, 29, 31, 35, 38]. In [35], the identified variables of student engagement were integrated into the INQPRO learning environment. Two variations of Bayesian Network model were handcrafted with the prior probabilities learned using the interaction logs of 54 students. Guerra et al. [15] presented the Mastery Grids system, an intelligent interface for on-line learning content that combined an open learner model and social comparison to support self-regulated learning and learning engagement. The results showed how Mastery Grids interacts with different factors, like gender and achievement-goal orientation, and ultimately, its impact on student engagement, performance, and motivation. Graesser et al. [14] investigated the relationship between emotions and learning by tracking the emotions that college students experienced while learning about computer literacy using an animated pedagogical agent, namely AutoTutor. Beverly et al. [38] recognized learner affective states while studying mathematics through the use of multi-sensory data. Podder et al. [29] proposed an engagement model using multiple descriptors from learners' feedback, eye tracking and saliency modeling in classroom education. Koji et al. [21] proposed a student engagement model based on digitized materials, including textbooks and collection event logs of tablets used by students. In [17], Guo et al. presented an empirical study system of how video production decisions affect student engagement in on-line educational videos. They measured engagement by how long students were watching each video, and whether they attempted to answer post-video assessment problems. These literatures have achieved improved results for student engagement study in on-line learning or classroom learning. However, complicated operation with PC or using invasive sensors makes these technologies hard to popularize in a natural classroom for multi-student engagement study. In a real and natural range classroom, non-invasive student engagement study in class-interactive environments is still a challenging task due to complex environment and various learning progress of students.

Due to the rapid growth in computer technology enhanced education, the intelligent learning system for student engagement study has received widespread attention from many research communities. There are some existing approaches for learning engagement data surveillance and analysis using computer vision, machine learning and special sensors. In recent years, student attendance using face recognition technologies becomes popular, which could replace traditional record table through artificial way [1]. Nowadays, most

students have smart cell phones and bring them to the classroom. Students and teachers can benefit from the additional channel of communication via their cellphones in classroom education [31]. Visual focus of attention (VFOA) and spontaneous smile expression are important for a student's cognitive and affective engagement understanding and analysis in class [10, 12]. The VFOA of a student can be defined as the student or the object that a student is focusing his/her visual attention on learning targets [3]. Researches on VFOA recognition can be classified into two types, i.e., invasive way based on special sensors and non-invasive way based on visual cues. Invasive systems are generally accurate and reliable depending on expensive sensors (e.g. SMI Eye Tracking Glasses or cameras mounted on a helmet) [28]. However, the discomfort and restriction disrupt a user's natural behaviors. In most of current work, non-invasive way has been shown that head orientation (head pose and location) can be reasonably utilized as an approximation of people's VFOA [3]. Facial expression recognition aims to classify a given facial image into one of the six commonly used emotion types, which include anger, disgust, fear, smile, sad and surprise proposed by Paul Ekman [40]. Among these six facial expressions, smile is distinct facial configuration. It's very informative in real-class applications. Spontaneous smile detection can be used to assess a learner' affective engagement in technology-enhanced learning (TEL) environment [3, 19]. In a crowded class, it is never easy to detect and recognize these learning data with each student, and analyze class engagement level [16, 27]. Hence, design and develop an intelligent system for automatically detect and recognize the learning behaviors in student engagement study is still an open and challenging problem in class education.

In this paper, we develop an intelligent learning environment for student engagement study based on multi-cue detection in class. The intelligent system automatically detects and analyses multiple learning cues based on five modules, i.e., attendance management, teacher-student (T&S) communication, VFOA recognition, smile detection and engagement analysis, as shown in Fig. 1. When a student enters the system environment, attendance management matches his/her identity and locates his/her profile using face recognition with the Camera1 at the entrance of environment (see Fig. 2). During the class, T&S communicate module assists the student and teacher interaction via the designed Question and Answer (Q&A) application runs on their own cell phones and keeps all the communication records for behavioral engagement. Meanwhile, in order to detect cognitive and affective engagement, students' VFOA and spontaneous smile expression cues are recognized from head poses estimation using the proposed hybrid multi-layered random forest method (HMRF) with a overhead Camera2 in the environment (see Fig. 2). Finally, student class engagement can be fused to decide based on multi-cues of one's behavioral engagement, cognitive engagement and affective engagement. The configuration of the intelligent learning environment is shown in Fig. 2.

1.2 Our contributions and paper organization

This paper is an extension version of our previous ICNC-FSKD2016 conference paper [24]. The main differences from the conference paper are as follows: First, the ICNC-FSKD2016 conference paper only proposed an intelligent system for interactive learning including three modules, i.e., attendance management, T&S communication, and VFOA recognition. The extended paper proposed the approach for student engagement study, which automatically detects and analyzes multiple learning cues based on five modules, i.e., attendance management, T&S communication, VFOA recognition, smile detection and engagement analysis. The detected multiple learning cues can be used for real-time learning behavioral data surveillance and analysis in class. Second, the hybrid multi-layered random forest method is proposed

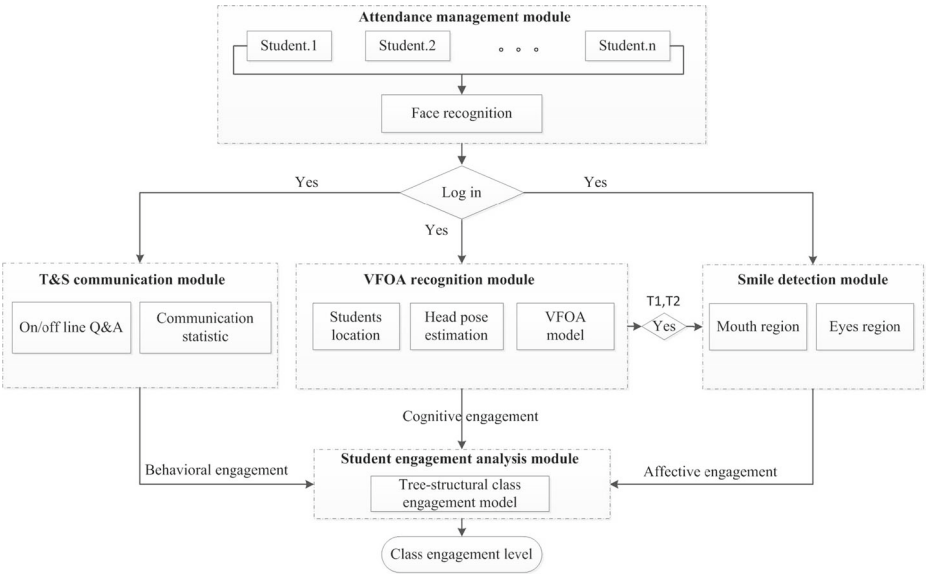


Fig. 1 Architecture of class engagement study in the intelligent system environment. It includes five modules, i.e., attendance management, T&S communication, VFOA recognition, affect recognition and engagement analysis

to detect students' smile expression in this paper, which is never proposed in the previous conference paper. It can recognized smile expression in the classroom environment robustly and efficiently. Third, in this paper, we present a novel tree-structural class engagement decision model to analyze a student engagement level based on one's behavioral engagement, cognitive engagement and affective engagement in class. It is also not proposed in the previous ICNC-FSKD paper. Finally, in our experimental section, we add many new experiments for each module. We evaluate our engagement study system on five modules.

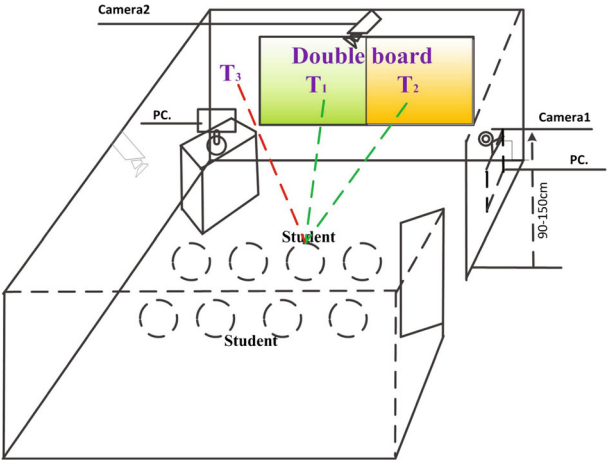


Fig. 2 The geometric configuration of the intelligent system environment

The main contributions of the paper are as follows:

1. Different from traditional student engagement study based on questionnaire surveys after class, we develop an intelligent learning environment that automatically detects and analyses student engagement based multi-cues in class objectively and effectively.
2. We investigate student engagement according to multi-cues of information, including student attendance, T&S communication, VFOA recognition, smile detection and engagement analysis.
3. We propose approaches to recognize students' cognitive and affective states during learning using head pose estimation, VFOA recognition and smile detection.
4. We present a novel tree-structural class engagement decision model to analyze a student engagement level based on one's behavioral engagement, cognitive engagement and affective engagement in class.

The remainder of the paper is organized as follows. In Section 2, we present the module of attendance management in the system. Section 3 describes behavior engagement based on T&S communication module. Section 4 gives cognitive engagement based on VFOA recognition. Section 5 presents affective engagement achievement using spontaneous smile detection. Section 6 analyzes the student class engagement level based on multi-cues of information. Section 7 discusses and analyses the experiments and results on each module for engagement study using the intelligent system, meanwhile, evaluates the performances on four practical class applications. Conclusion and future work are given in Section 8.

2 Attendance management in the intelligent system

Student and teacher should first register their information with the attendance management module, i.e., ID, name, subject, and photo. When a student enters the system environment, attendance management matches his/her identity and locates his/her profile using face recognition with the Camera1 at the entrance of the classroom (see Fig. 2). After identity matching, this module automatically sends each attended student's information into the teacher, and generates a unique QR code associating with each attended student to authorize him/her to access the T&S communication module. This module helps to track each attended student's learning interaction in class.

3 Behavioral engagement based on T&S communication

When students and a teacher log in to a course, the T&S communication module provides an additional interactive channel of Q&A between a teacher and students for students' behavioral engagement analysis via their cell phones. Different from a traditional class mode, our proposed T&S communicate module keeps all the communication records and analyzes statistical accuracies of students' answers from Q&A application. A student logins in the T&S communication module using the QR code provided from the attendance management module. Figure 3 shows the framework of the T&S behavior communication. The specific steps are in the following.

During the class, a teacher can release a class topic or exercises through the on-line question function. Meanwhile, students can receive the topic or exercises and submit their answers through the application on their own cellphone with on-line answer function. Then, with the communication statistic function, students and the teacher can get the statistical

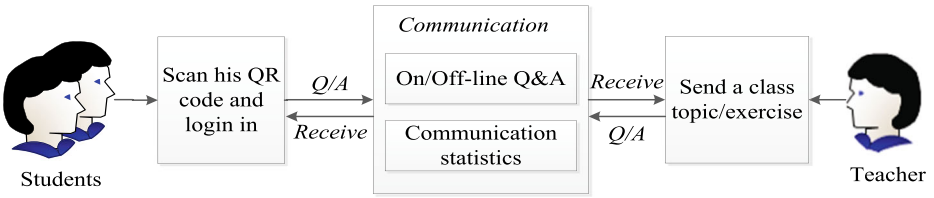


Fig. 3 The pipeline of T&S communication in class

analysis data of interactions, including each student's communication times (including answers and questions) and the average accuracy of one's answers, etc. This statistical data can help that the teacher objectively understand the student behavioral engagement and improve the teaching scheme. Inspired by [30], according to student's communication times and the accuracy of the answers, a student's behavior engagement model E_b based on T&S communication is defined as:

$$E_b = \begin{cases} Positive, & t \in \left(\frac{3}{4}N, +\infty\right) \text{ or } a \in (80\%, +\infty) \\ Negative, & t \in \left(-\infty, \frac{N}{4}\right] \text{ or } a \in (-\infty, 50\%] \text{ or} \\ & t \in \left(\frac{N}{4}, \frac{3}{4}N\right] \text{and } a \in (50\%, 80\%] \end{cases} \quad (1)$$

where N is the total number of communication times between a teacher and a student in whole class, t is the number of a student's communication times during every five minutes of a class recording, and a represents the statistic average accuracy of a student's answers. The threshold values in the Eq. (1) are the empirical values set by a teacher in class. The student's behavior engagement model is Positive when his/her communication times is higher than $3N/4$ or the accuracy of the answers is over 80%. Otherwise, one's behavior engagement model can be considered as Negative.

In addition, students can also get the registered information of other students and their teacher in the same class. It's convenient for students and a teacher to set up a discussion group after a class.

4 Cognitive engagement based on VFOA recognition

Different from traditional cognitive engagement based on questionnaire surveys after class, we propose a novel VFOA recognition method for cognitive engagement study in class. VFOA recognition module captures each student's cognitive engagement on different attention targets during the class. VFOA recognition method includes three stages, i.e., student positioning, head pose estimation and VFOA recognition. The flowchart of the method is shown in Fig. 4. First, student positioning associates each student's identity with his/her 2D position in a video sequence via face detection, face tracking and identity matching. Then, the hybrid multi-layered random forest algorithm is proposed to estimate continuous head poses and tip of the nose location in a hierarchical way, integrating classification and regression random forests. Finally, the VFOA is recognized and tracked based on head pose, prior state and environmental cues. Environmental cues include the physical placement of targets and the participant's 3D position in the classroom. Prior state comprises the prior recognized attention state and some prior 3D cues in the classroom.

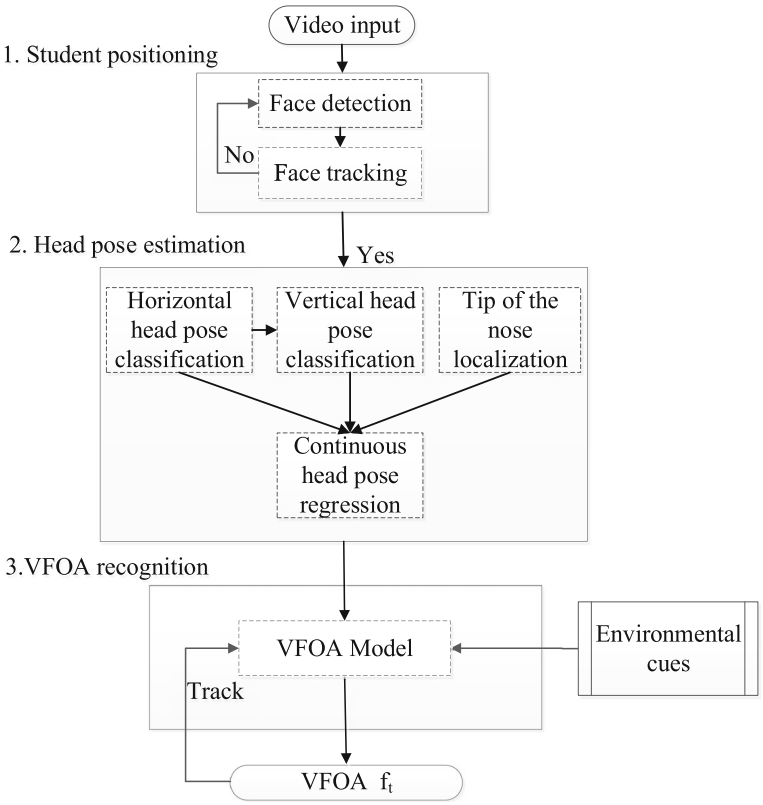


Fig. 4 The flowchart of VFOA recognition

4.1 Student positioning

Due to the fixed seats of students in class, student positioning associates each student's identity with his/her 2D position in a video sequence captured using the overhead camera (see Fig. 2) in the class environment. 2D position of multi-students can be obtained using face detection and tracking. In a natural and wide classroom, multi-student face detection and tracking are two challenging tasks due to wide angles and low resolution.

In this module, a multi-view face detector based on a cascade of boosted classifiers with Haar-like features [37] has been trained to detect faces with LFW facial dataset [18] and CCNU classroom dataset [23] collected from an overhead camera under various poses, illumination and occlusion. In order to decrease false detection, detected face areas were averaged in the initial 30 frames of a sequence, firstly. Then, robust facial detection is performed within sub-windows, which are extensions of the average face areas.

After face detection in the initialization process, a face tracking method is used to track facial areas in the scene. The CAMShift [5, 39] method is a popular method for face tracking but rapid movement degrades s the performance. To improve the robustness of tracking, an advanced CAMShift method based on the position of skin color is used to track face areas and detect tracking failures in a sequence. The advanced CAMShift method makes use of

the skin color and motion position information to suppress tracking failures. The procedure of the advanced CAMShift method is as follows.

There are three steps in the advanced CAMShift tracking. In the first step, CAMShift is first used to track face areas. In the second step, when tracking in process after the initial 30 frames of sequence, tracking failure detection and updating are performed to suppress tracking loss. Skin-color detection is used to update face positions in the expanded tracking regions. If face areas are updated, CAMShift will continue to track face areas based on the updated positions. If no face area is detected by skin-color detection in the expanded tracking regions, it may indicate that there are abandoned faces due to students leaving their seats. The system will remove the regions without face areas. In the third step, it is required to re-initialize and update the students' positions at scheduled time intervals or leaving students coming back to the classroom.

Finally, the student identity from the attendance management module can be associated with one's face area for student positioning.

4.2 Head pose estimation

Random Forest (RF) is a popular ensemble method in computer vision because of its powerful regression and classification capability [6, 26]. In previous work, we proposed a Dirichilet-tree distribution enhanced random forests algorithm (D-RF) to estimate head pose and facial features in [22, 23]. In this paper, a more discriminative hybrid multi-layered random forest (HMRF) is proposed to estimate head pose in a hierarchical way integrating classification and regression random forests. Different from the previous work, the improved HMRF is a weighted combination of classification and regression D-RF, which estimates continues head poses based on combined texture and geometric features using the multi-task learning method. More detail on our previous work can be referred to [22, 23].

In order to train sub-forests in the HMRF, the training images have been divided into 4 hierarchical training sub-sets. First, face areas are located as described in the Student positioning stage and normalized to 125*125 pixels. Then we randomly extract 200 facial patches $\{P = \{F_i, H_i^m, D_i\}\}$ from each face area in sub-sets. The patch appearance F_i is defined as multiple texture feature channels $F = \{F_i^1, F_i^2, F_i^3\}$. F_i^1 contains the gray values of the raw facial patch with dimension as 31*31. F_i^2 represents the Gabor feature based principal component analysis (PCA) of facial patches with dimensions as 35*12. F_i^3 is the histogram distributions of the patches. The channel $H_i^m = \{h_i^1, (h_i^2|h_i^1), (h_i^3|h_i^2, h_i^1), (h_i^4|h_i^3, h_i^2, h_i^1), \theta_{yaw,pitch}\}$ contains the annotated discrete and continuous angles of training sub-sets in different sub-layers of HMRF, where h_i^1 are 3 yaw angles in the first sub-layer, $h_i^2|h_i^1$ are refined yaw angles refined from h_i^1 in the second sub-layer, $h_i^3|h_i^2, h_i^1$ are 3 pitch angles under condition of each yaw angle h_i^2 in the third layer, $h_i^4|h_i^3, h_i^2, h_i^1$ are refined angles based on the above annotated angles at leaves of the Dirichlet-tree in the fourth sub-layer. $\theta_{yaw,pitch}$ are continuous head pose angles in the fifth sub-layer. D_i is the offset vector from a patch centroid to the tip of the nose. The training procedure of each sub-forest in different sub-layers is similar to [23]. The head pose and tip of the nose position probabilities of patches $p(H_i{}^m, D_i|l) = N(H_i{}^m, D_i; \overline{H_i{}^m, D_i}, \Sigma_{H_i{}^m, D_i})$ have been stored in leaves l of the trained trees as the Gaussian probabilistic distribution, where $\overline{H_i{}^m, D_i}$ and $\Sigma_{H_i{}^m, D_i}$ are the mean and covariance matrix of head pose and tip of the nose probabilities in the i-th sub-layer of HMRF.

In order to obtain continuous head pose angles, a weighted composited measure is proposed to estimate continuous head pose based on multiple probabilities in the last fifth sub-layer, which is defined as:

$$\arg\max_{H}(w_m p(H^m) + \left(1.0 - \exp\left(-\frac{D_i}{\gamma}\right)\right) p(D_i)), \tag{2}$$

where γ is used to control the steepness of this function, the weight $w_m = P_S/P$ that is defined as the ratio of samples' number P_S of a subset to full samples' number P in each single tree of the HMRF.

Finally, the position of the nose tip D_i and head pose angles can be estimated using regression voting in the fifth sub-layer under the condition of the estimated coarse head poses from the first sub-layer to the fourth sub-layer.

4.3 VFOA recognition for cognitive engagement study

The objective of this module is to achieve student cognitive engagement based on VFOA recognition in class from a overhead camera (see Camera2 in Fig.2). A VFOA model is proposed to recognize and track attention based on head poses, prior state and visual environmental cues as shown in Fig. 5. $h_t^k = \{\theta_{yaw,pitch}\}$ represent the estimated head poses of the student k in horizontal and vertical directions at time t, c_t represents the environmental cues currently, f_t^k and f_{t-1}^k denote the VFOA states of the student k at time t and prior time $t-1$.

4.3.1 Environmental cues

VFOA recognition from a monocular camera is difficult due to unknown 3D cues. In order to solve this problem, we introduce an approximated method to obtain the environmental cues $c_t(T_i, B)$ based on some prior state. $T_{i=1,2}\{T_1, T_2\}$ are the physical placements of attention targets in the white board and could be measured previously (see Fig. 1). $B(x, y, z)$ is the 3D position of a person estimated from a monocular camera in the classroom. According to averaging 100 persons' sitting height (the height of tip of the nose) in the classroom, we fixed a 2D reference plane with its height B_y =120cm as the prior state of 3D cues. Then, 32 different points in this 2D reference plane were labeled according to their image coordinates. Homography Matrix H between the 2D reference plane and the image was computed

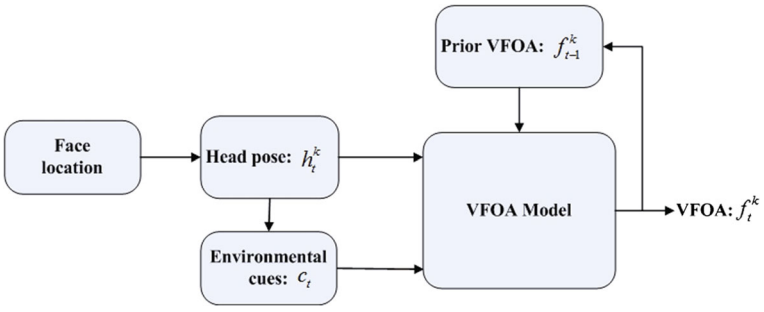


Fig. 5 The proposed model for VFOA recognition

based on these labeled points by Affine Transformation $h(\bullet)$. When recognizing, the reverse procedure could be performed, each person's position $B(x, y, z)$ can be obtained based on prior tip of the nose D_N and Homography Matrix H by Affine Transformation $h(\bullet)$,

$$B(x, y = 120, z) = h(D_N, H). \tag{3}$$

4.3.2 VFOA recognition

In the context, in order to recognize students' VFOA in the natural classroom, the attention target $T(x, y)$ of a student is computed under the estimated head pose $\theta_{yaw,pitch}$ and his position $B(x, y, z)$. The geometric relationship between $T(x, y)$, $\theta_{yaw,pitch}$ and $B(x, y, z)$ is shown in Fig. 6. The VFOA target $T(x, y)$ can be computed as:

$$T(x, y) = \left\{ \frac{B_z}{\cot(\theta_{yaw})} + B_x, B_y + \frac{\tan(\theta_{pitch}) \times B_z}{\cos(\theta_{yaw})} \right\}. \tag{4}$$

4.3.3 Cognitive engagement based on VFOA recognition

In order to achieve cognitive engagement from VFOA targets, we define cognitive engagement based on VFOA recognition in the following. If the VFOA target point $T(x, y)$ belongs to the double board, the cognitive engagement is T_1 (left board) or T_2 (right board), which means that VFOA target of a student is within the double broad. Otherwise, cognitive engagement T_3 is unfocused when the VFOA target point $T(x, y)$ is outside the double board. The relationship of cognitive engagement and VFOA f_t^k of a student k is as follows:

$$f_t^k = \sum_t \sum_{k=1, i \neq k}^{K} \delta(T_t^k - T_i), k = 1, ...K.i = 1, 2., \tag{5}$$

where $\delta(T_t^k - T_i) = 1$ represents a student's cognitive engagement on the VFOA targets T_1 or T_2. While $\delta(T_t^k - T_i) = 0$ represents that a student does not focus the double-slides, and one's cognitive engagement directly outputs T_3='un-focused'.

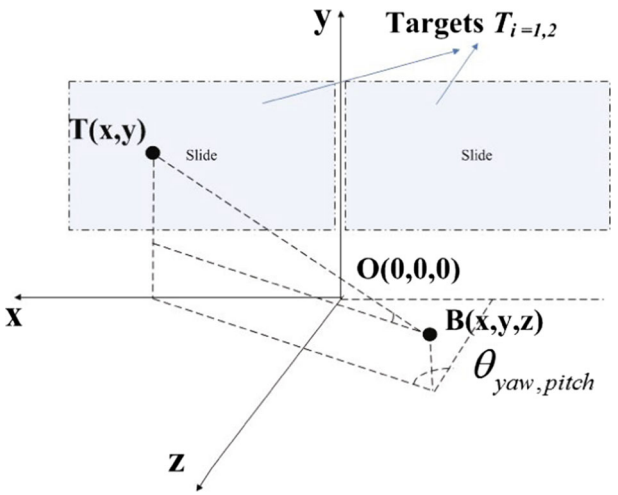


Fig. 6 Geometric relationship between the attention target, position and head pose of the student

To achieve cognitive engagement from the VFOA targets of multi-students in a video sequence, we rely on GMM to track the jointed model displayed in Fig. 5, and according to the jointed distribution of the estimated variables and prior state of VFOA as the VFOA model in the previous section is given by

$$p(f_t^k|f_{t-1}^k, h_t^k, c_t) \propto p(f_0^k)\prod_{t=1}^{T} p(h_t^k|c_t)p(f_t^k|f_{t-1}^k), \tag{6}$$

where c_t is the visual environmental cue. In a video sequence, the VFOA recognition is performed by estimating the optimal sequence of states which maximizes $p(f_t^k|f_{t-1}^k, h_t^k, c_t)$.

5 Affective engagement based on spontaneous smile detection

Among various facial expressions, smile is very informative for a student's affective engagement. Senechal et al. [32] and Chen et al. [8] have used the smile as important visual feature to detect one's affective state. In class, smile reflects the student's motion associated with learning interesting, hence, spontaneous smile detection is crucial to analyze the student's affective engagement. In this section, the proposed HMRF method is continuously used for spontaneous smile detection when the student's VFOA targe is within the double board in the class environment.

HMRF was carried out using 5700 smiling images and 14500 non-smiling images for smile detection training. These images were obtained from different sources, e.g., public smile expression databases, Internet and captured video sequences in a natural classroom scene. The procedure includes two cascaded sub-forest training, i.e., mouth sub-forest training and eyes sub-forest training. The mouth sub-forest has been grown with patches based features from the mouth sub-regions, and the eyes sub-forest has been grown from the eyes sub-regions.

For testing of smile detection, the local mouth and eyes areas have been detected by a cascade of Adaboost tree classifiers with Haar-like features [37] and are normalized based on the estimated head poses $\theta_{yaw,pitch}$, firstly. Then, the mouth and eye sub-forests are combined to determine that the face is smile/non-smile expression. The testing framework for smile detection is shown in Fig. 7. In order to eliminate the influence of head poses, we

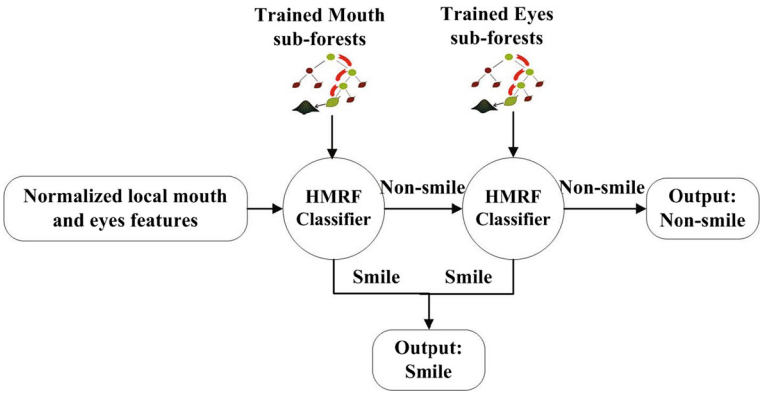


Fig. 7 The pipeline of spontaneous smile detection using HMRF in the module

normalize the local mouth and eyes areas, by estimated head poses. After pose normalization, the HMRF method is used to detect each student's spontaneous smile expression state in cascaded local mouth-eyes areas. Mouth sub-forest is firstly used to detect smile in the mouth area. When the detected result is "smile", the testing procedure is ending and outputs smiling state. When the detected result is "non-smile", eyes sub-forest compensates to detect the expression state in the eyes areas in a cascaded way. The final expression probabilities can be computed in the leaves of the sub-forests,

$$\begin{aligned} p(E_i = 1|\theta_{yaw,pitch}) =& p(E_i = 1|\theta_{yaw,pitch}, f_{mouth}) + \\ & p(E_i = 0|\theta_{yaw,pitch}, f_{mouth}) \cdot \\ & p(E_i = 1|\theta_{yaw,pitch}, f_{eyes}); \\ p(E_i = 0|\theta_{yaw,pitch}) =& p(E_i = 0|\theta_{yaw,pitch}, f_{mouth}) \cdot \\ & p(E_i = 0|\theta_{yaw,pitch}, f_{eyes}), \end{aligned} \tag{7}$$

where $p(E_i = 1|\theta_{yaw,pitch})$ represents the probability of the spontaneous smile expression stored in leaves of sub-forests, $p(E_i = 0|\theta_{yaw,pitch})$ represents the probability of no-smile expression, $\theta_{yaw,pitch}$ are the estimated head poses in horizontal and vertical directions, f_{mouth} and f_{eyes} represent the mouth and eyes sub-forests, respectively.

The weighted Gussian voting method is used to vote the leaves' probabilities of the relative sub-forests. We can obtain the final expression state as:

$$p(E|\theta_{yaw,pitch}) = \frac{1}{T} \sum_{T} \sum_{i} p_i(E|\theta_{yaw,pitch}, P), \tag{8}$$

where T is the number of trees within the sub-forests, P is the the image patches from the cascaded mouth and eyes sub-regions. Figure 8 shows some examples of smile detection on public facial expression datasets, e.g., GENKI [20], CK+ [25], and our CCNU dataset [23].

6 Class engagement analysis based on multi-cues

It' s known that assessing a student' s continuous engagement degree has never been easy in class. Many literatures and experts use discrete levels of the engagement to evaluate the degree of student engagement based on several learning data and feedbacks [35, 36]. Inspired by the papers, we propose the student class engagement model as shown in Fig. 9. In the class engagement analysis module, a novel tree-structural class engagement model is proposed to analyze student class engagement level based on multi-cues of one's behavioral engagement, cognitive engagement and affective engagement, which models the student class engagement as three discrete levels, i.e., "High level", "Moderate level" and "Low level".

When a student begins to attend the class, the tree-structural class engagement model assesses his/her behavioral engagement based on T&S communication module, firstly. If the behavioral engagement is positive, it could mean that the student's class engagement level is the "High level". Otherwise, the tree-structural decision model continuously needs to decide the cognitive engagement based on VFOA recognition module. If the cognitive engagement is recognized as T_3, it could obtain that the student's class engagement level is the "Low level", otherwise further discusses on affective engagement should be analyzed. Using the spontaneous smile classification, affective engagement as smile or non-smile of a

Fig. 8 The examples of smile detection on GENKI, CK+, and CCNU datasets. The red rectangles are the eyes and mouth areas extracted from the images

student can be detected on smile detection module. The class engagement level of a student will be decided as the "High level" when the affective engagement is detected as smile, otherwise, as the "Moderate level" when the affective engagement is detected as non-smile.

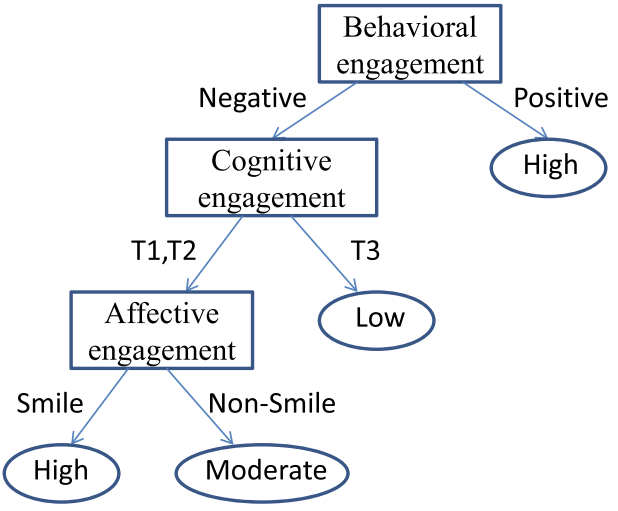


Fig. 9 The proposed tree-structural class engagement model based on multi-cues of one's behavioral engagement, cognitive engagement and affective engagement

Decisions from each module are fused to help teacher understand the students' learning states and adjust teaching scheme. For example, according to the positive results from behavior communication, the teacher could get known that the students understand the subject, and the teacher may not need to clarify it further. Moreover, if most of students' cognitive engagement is T_3 with negative behavioral engagement, it could mean that the student class engagement is low level. The teaching slide may be not attractive. The teacher may need to promote his slide contents and adjust his teaching process. On the other hand, if most of students' cognitive engagement is focused on VFOA targets T_1, T_2 with spontaneous smile expression, it may mean that the teaching efficiency is better. The student class engagement is high level. The class engagement study based on multiple cues of behavior, cognitive and affective engagement informations can get a better grasp on the students' learning status, which helps to improve the learning efficiency in class.

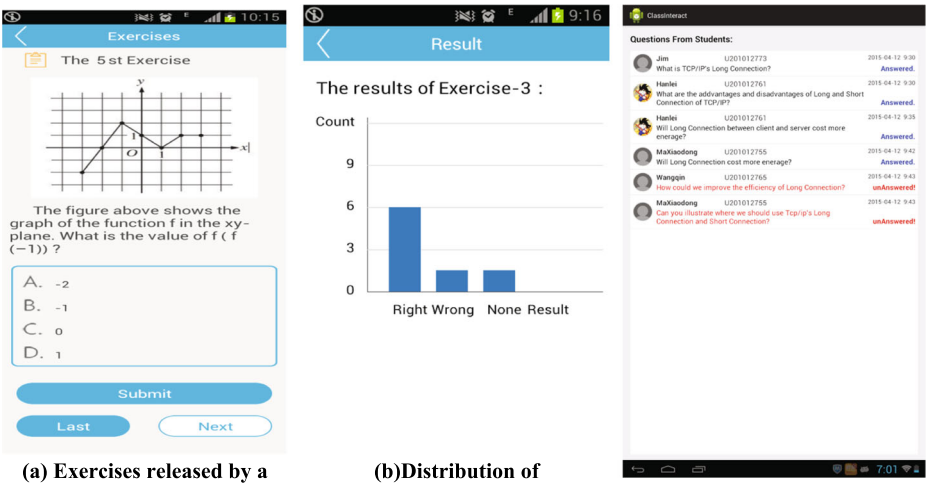
7 Experiments and summary

In this section, we thoroughly evaluated and discussed five function modules in the intelligent system, i.e., attendance management, T&S communication, VFOA recognition, spontaneous smile detection and engagement analysis module on some public available datasets (i.e., Pointing'04 dataset [13], LFW dataset [18] and CCNU classroom dataset [23].) and real-class videos. These datasets and videos were chosen since they contained unconstrained face images with poses ranging from $-90°$ to $+90°$ and spontaneous smile expression. The images were collected in the wild, and varied in poses, lighting conditions, resolutions, races, occlusions, and make-ups, etc.

7.1 Attendance management

Students and a teacher should first register their information with the attendance management module, i.e., ID, name, subject, and photo using face recognition. Figure 10 gives an example of the attendance management using face recognition in our intelligent learning environment. When a student enters the classroom, attendance management matches his/her identity using face recognition and generate his/her identity features for next modules.

Fig. 10 An example of attendance management using face recognition in the intelligent system



(a) Exercises released by a teacher **(b)Distribution of students' answers** **(c) Questions from students**

Fig. 11 Examples of the T&S communication for behavioral engagement analysis in a class

7.2 Behavioral engagement analysis from T&S communication

Figure 11 shows examples of the T&S communication for behavioral engagement analysis in a class via smart cellphones. The left image is the Q&A application on a student's cellphone, the middle image is the statistic distribution of answers submitted by students, and the right image shows questions from students received on the teacher's cellphone. Table 1 shows the statistic results of each student's behavioral engagement in a class, which have been achieved based on the number of communication times and accuracies of a student's answers in class. There are 8 students' behavioral engagement analysis in a real-class video.

During the class, each student's behavioral engagement can be also tracked and recorded in every five minutes, as shown in Fig. 12. The horizontal axis represents every time periods, the vertical axis represents two states of behavioral engagement, and different color spots represent different students in the class. It can help teachers to assess each student's learning efficiency objectively during the class. The overall distributions of the number of students with positive and negative behavioral engagement in a class can be shown in Fig. 13. The orange histogram shows the number of students with positive behavioral engagement in the class, while the navy blue histogram shows the number of students with negative behavioral

Table 1 The statistic distributions of each student's behavioral engagement based on communication times and answer accuracies in a class

Students	Stu.1	Stu.2	Stu.3	Stu.4	Stu.5	Stu.6	Stu.7	Stu.8
Communication times	4	9	8	6	4	9	4	5
Answer accuracies	60%	73%	88%	65%	40%	73%	53%	40%
Behavioral engagement	Negative	Positive	Positive	Negative	Negative	Positive	Negative	Negative

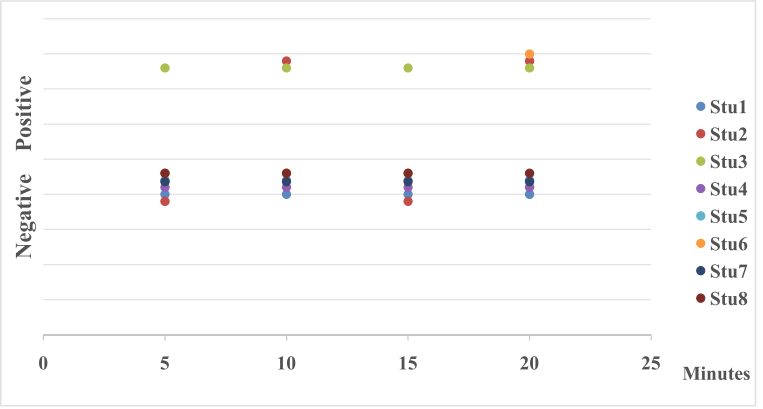


Fig. 12 The each student's behavioral engagement tracked during the class

engagement in the class. A teacher could improve his teaching scheme based on the analysis of student behavioral engagement in the class processing.

7.3 Cognitive engagement analysis from VFOA recognition

In the student positioning step, the average detected precision achieves 95.6% in the public LFW, collected CCNU classroom dataset and real-class videos. Figure 14 shows the ROC curve of face detection on the public LFW facial dataset [18] and our collected CCNU classroom dataset [23]. The average tracking time is 0.007s per frame on a PC with Intel(R) Core(TM) i5-2400 CPU@ 8GHz. Table 2 lists the true positive rate (TPN), true negative rate (TNR), and false negative rate (FNR) of our trained multi-view face detector and OpenCV

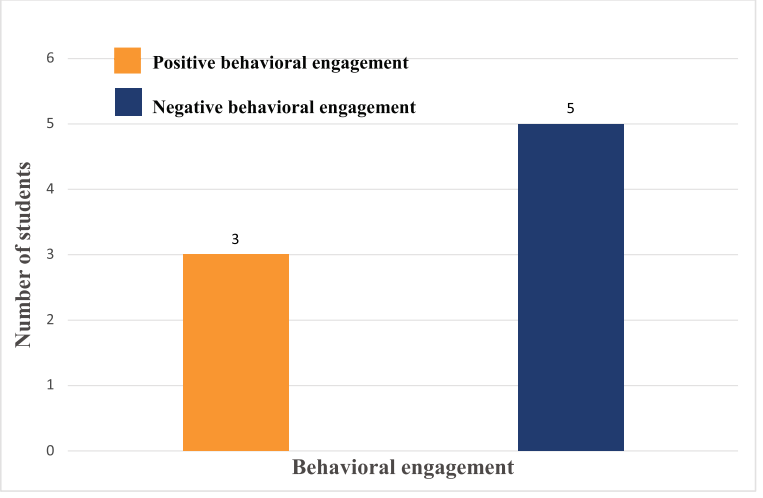


Fig. 13 The overall behavioral engagement distribution in a class

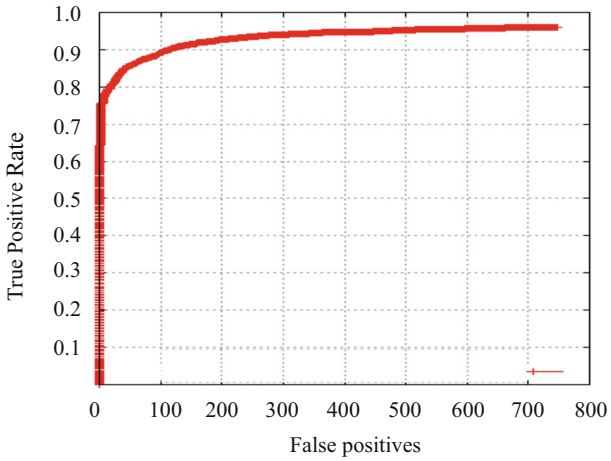


Fig. 14 The ROC curve on LFW and CCNU classroom datasets

face detector on LFW and CCNU classroom datasets. It is demonstrated that the multi-view face detector exhibited the better performance with pose and condition variation.

In this module, the advanced CAMShift method based on the position of skin color is used to track face areas and detect tracking failures in a sequence. Once the tracking failure has been detected from the overhead camera, skin-color detection will be also needed to confirm and update face positions in the expanded tracking regions. If face areas are updated, continuously tracking will be performed based on the updated positions. If no face area is detected by skin-color detection in the expanded tracking regions, the system will remove the regions due to abandoned faces. The successful detection-tracking examples of student positioning can be shown in Fig. 15.

In order to evaluate students' VFOA recognition, we firstly evaluated the accuracies of students' head poses and tip of nose location on Pointing'04 dataset [13], LFW dataset [18]and CCNU classroom dataset [23]. The parameters of training and testing in HMRF are similar to [11]. A 4-fold cross-validation was conducted. Among the three datasets, our method achieved the greatest performance with Pointing'04. The accuracy in yaw and pitch angles were in the range of 80% to 90%, respectively. Note that both LFW and CCNU datasets consist of great variation of poses, lighting, occlusions, etc. For two more challenging datasets, the accuracy was also above 80% for yaw rotation and above 70% for yaw and pitch rotation. The average error reached 8.2° for the Pointing'04 and those of the other two

Table 2 The TPN, TNR, and FNR (%) using two face detectors on LFW and real CCNU classroom datasets

Datasets	LWF dataset			CCNU classroom dataset		
	TPN	TNR	FNR	TPN	TNR	FNR
Multi-view face detector	96.3	2.4	1.6	94.5	3.3	4.1
OpenCV face detector	92.6	3.2	2.1	83.6	7.7	9.4

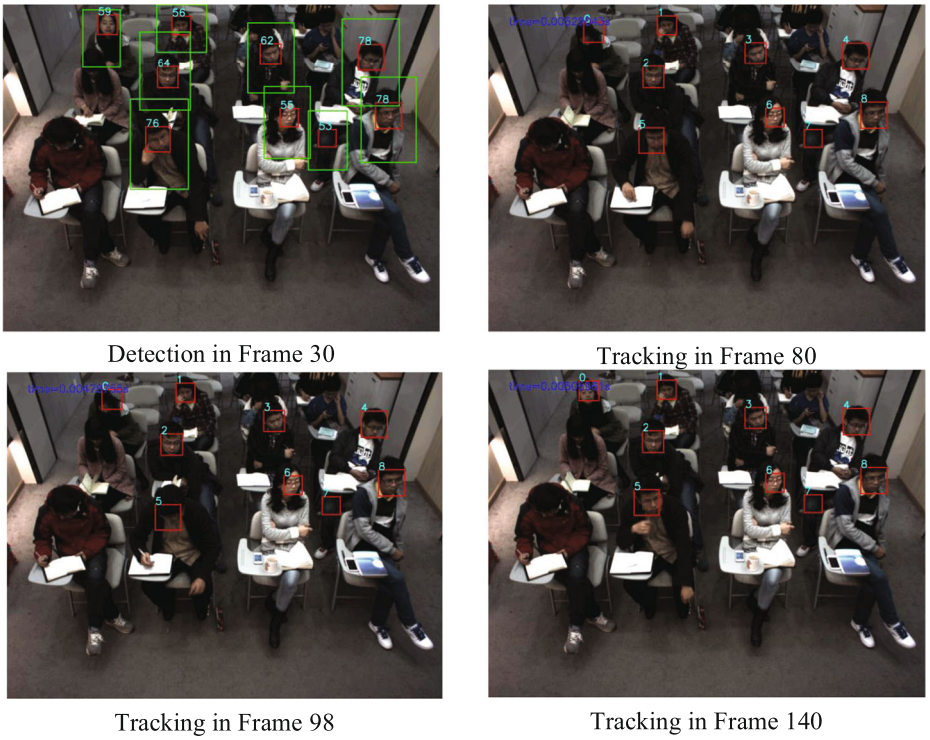


Fig. 15 The detection-tracking examples of student positioning in a real-class video

were close to 10°.in the class environment. The average accuracy using the HMRF algorithm is 74% on the CCNU dataset collected from an overhead camera in the wide range classroom. The average location error on tip of the nose reached 0.4 pixels on these three datasets. The all results in the Table 3 show the robustness of our approach for head pose estimation.

Table 4 provides the VFOA recognition results obtained using different head pose estimation algorithms, including the HMRF proposed in this paper, D-RF in [23], C-RF proposed in [11], RF in [6]. One can see that the best VFOA recognition rate can be obtained using our proposed HMRF.

Table 3 Accuracies and average errors (degrees) of HMRF method for head pose estimation and tip of nose location on different datasets

Datasets	θ_{yaw}	θ_{pitch}	$\theta_{yaw,pitch}$	Ave. Error	STD.	Tip of the nose
Pointing'04	90.8%	90.2%	81.5%	8.2°	4.0°	0.15 pixels
LFW	85.1%	89.4%	73.8%	13.2°	6.2°	0.16 pixels
CCNU	84.0%	89.5%	73.5%	13.5°	6.1°	0.7 pixels

Table 4 VFOA recognition rates with different head pose estimation algorithms

Algorithms	Recognition Rate (%)	Mean error (%)
HMRF	67.8	32.5
D-RF	63.5	38.4
C-RF	58.6	40.3
RF	50.2	46.5

The accuracies of cognitive engagement on different VFOA targets are provided in Fig. 16. The experiments have been performed on 6 videos in real classes. Each recording includes 8 persons in ten minutes long. The average accuracy reaches 67.8% using the proposed approach. The accuracies on T_1 , T_2 and T_3 are 62.28%, 60.81%, and 80.2% correspondingly. T_1 , T_2 and T_3 represent the results of cognitive engagement on different VFOA targets are right slide of the double board, left slide of the double board and outside of the double board, respectively. The accuracies on T_1 and T_2 are inferior to T_3 due to their smaller scales in the classroom.

Figure 17 shows an example of cognitive engagement based on VFOA recognition in our intelligent system in real class, where the estimated head poses are shown above the face rectangles and cognitive engagement on VFOA targets are below the rectangles. The pie chart in the lower right corner of Fig. 17 gives the cognitive engagement distributions of all students in class. The red represents cognitive engagement on the T_1 VFOA target, the blue represents cognitive engagement on the T_2 VFOA target, and the yellow represents cognitive engagement on the T_3 VFOA target. These results can help teacher understand student's learning status and provide appropriate teaching scheme objectively and effectively.

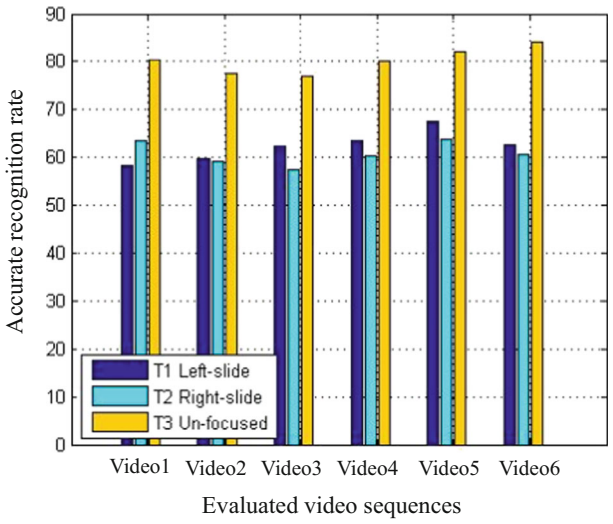


Fig. 16 The accuracies of cognitive engagement on different VFOA targets T_1, T_2 and T_3 in 6 real-class videos

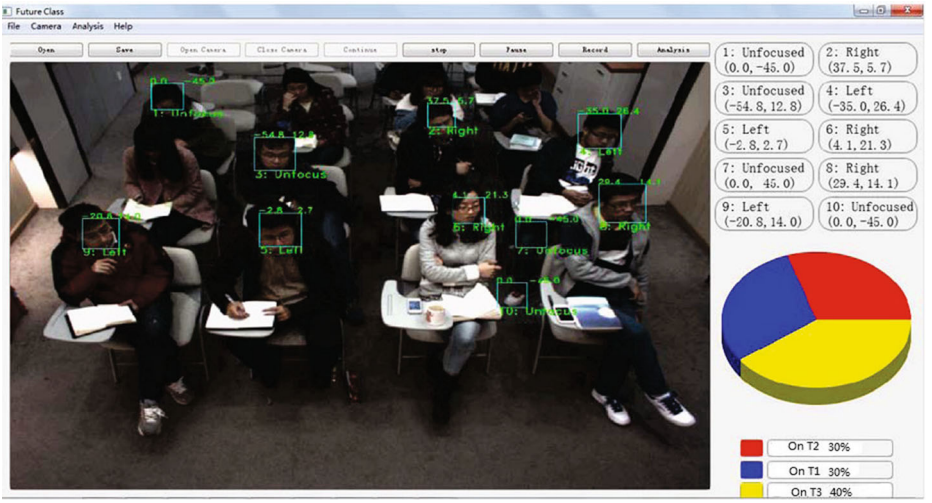


Fig. 17 An example of cognitive engagement based on VFOA recognition in the environment. "Left, Right, and Unfocused" represent the cognitive engagement on the T_1, T_2 and T_3 VFOA targets, respectively

Table 5 shows the statistic distributions of cognitive engagement based on VFOA recognition in 6 different real-class videos. The VFOA on T_1 or T_2 represents that students' cognitive engagement is positive, while VFOA on T_3 represents that students' cognitive engagement is negative. The distribution of cognitive engagement can be used for analyzing the final class engagement level objectively and effectively.

7.4 Affective engagement analysis from spontaneous smile detection

In the intelligent learning environment, we obtain each student's affective engagement based on spontaneous smile detection instead of traditional questionnaire surveys in class. Spontaneous smile detection has been tested from 500 images from CK+ expression dataset [25], 500 images from GENKI dataset [20], and 500 images from CCNU classroom dataset [23]. A 4-fold cross-validation was conducted. Table 6 lists the accuracies with respect to smile and non-smile expression of the HMRF method. One can see that the best performance is on the CK+ expression dataset, where the average accuracy is 96.6%. Owing to more noise and spontaneous expression on the CCNU dataset, however, where the average accuracy still achieves to 92.2% detected by the HMRF method. The each STD. is about 2.0%, which indicates the proposed method is robustness on different datasets.

Table 5 The statistic distribution of cognitive engagement based on VFOA recognition in 6 real-class videos

Distribution	video1	video2	video3	video4	video5	video6
VFOA on T_1	36.0	45.2	40.5	26.1	33.2	48.5
VFOA on T_2	40.2	26.7	22.0	48.6	28.6	9.0
VFOA on T_3	24.8	29.1	32.5	25.3	38.2	42.5

Table 6 Average accuracies (%), average errors (%) and STD. (%) with respect to smile and non-smile expression of the proposed method on three challenging datasets

Datasets	Smile	Non-Smile	Ave. Acc.	Ave. Error	STD.
CK+ dataset	97.1	96.3	96.6	3.2	1.2
GENKI dataset	95.0	93.8	94.2	5.7	2.5
CCNU dataset	92.4	92.0	92.2	7.2	2.5

Different methods (SVM, Boosting difference [33], IMRF[41], ELM [2]) have been compared to our method on the GENKI dataset. We employed a 4-fold cross-validation and used the same training and testing sets. Table 7 shows the comparison of different methods on the spontaneous smile expression dataset. In ELM [2], Linear Discriminant Analysis (LDA) and SVM are used as classifiers with Local Binary Pattern (LBP) features. The average accuracy reached to 88.5%. In Boosting difference [33], pixel intensity difference(PID) is extracted as feature, while AdaBoost is used as classifier, whose average accuracy is 89.7%. IMRF [41] proposed iterative multi-output random forests, which can obtain 85.5% of the average accuracy. In this module, the HMRF based on mouth and eyes sub-forests achieved the average accuracy of 94.2% and obtained the best performance. Additionally, the smallest STD. (2.5%) indicates that our HMRF improved the spontaneous smile detection with great robustness.

The examples of multi-students' affective engagement based on spontaneous smile detection in class are shown in Fig. 18. One can see that our proposed method can automatically achieve multi-students' affective engagement from spontaneous smile expression in the intelligent system environment.

7.5 Student class engagement analysis in practical class applications

Examples of class engagement analysis with a student at different moments in class are shown in Table 8. Due to the space limitation of a table, we only select 4 moments during the class to show the student's class engagement levels, such as, 9:10am, 9:20am, 9:30am and 9:40am. The class engagement levels have been decided based on multi-cues by our proposed tree-structural class engagement model. From the table, the student's multi-cues including one's behavioral, cognitive and affective engagement can be detected and analyzed during the class process, and be fused to decide his/her class engagement level in the class engagement analysis module. When one's behavior engagement is 'Positive', the student class engagement level can be analyzed as 'High'. Otherwise, if one's behavior engagement is negative, his/her VFOA and expression state should be recognized by

Table 7 Comparison of different methods on spontaneous GENKI dataset

Different Methods	Smile (%)	Non-smile (%)	Ave. Acc.(%)	STD.(%)
SVM	84.4	83.6	84.1	3.0
Boosting difference [33]	88.3	89.9	89.7	2.6
IMRF[41]	85.7	85.2	85.5	2.7
ELM [2]	90.2	87.6	88.5	2.5
Our proposed method	95.0	93.8	94.2	2.5

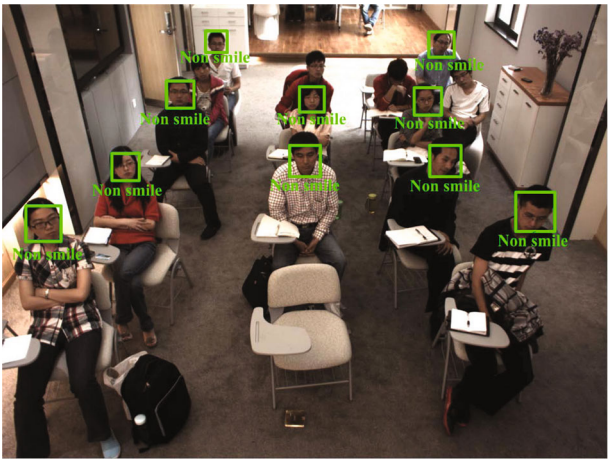


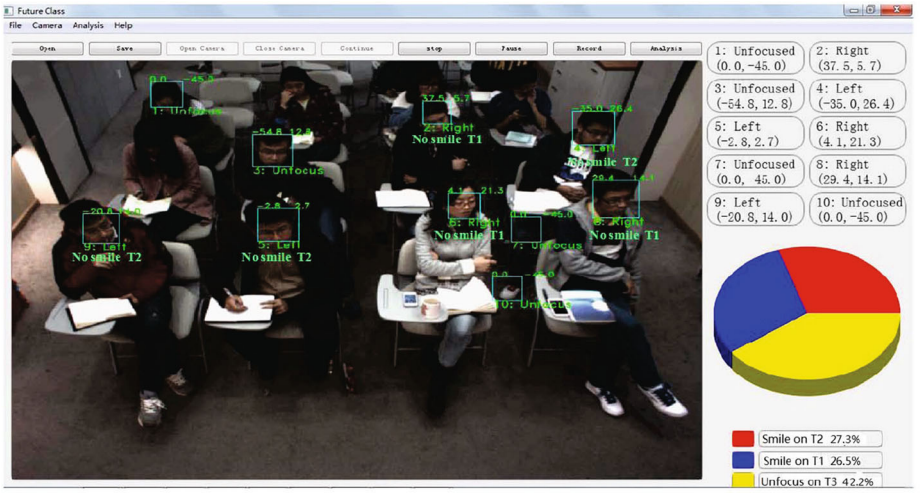
Fig. 18 The successful examples of affective engagement achievement using spontaneous smile detection in class.The proposed method can recognize spontaneous smile expression in the challenging class environment including occlusions, expressions, poses, low resolution, and make-ups, etc

computer vision analysis modules. Examples of multi-students' cognitive and affective engagement recognition is shown in Fig. 19. Table 9 provides the statistic distributions of student class engagement levels with all students in four real-class videos captured by the intelligent system. A teacher could achieve and study student class engagement based on multi-cues objectively and effectively, and improve his teaching scheme based on the analysis results of student class engagement.

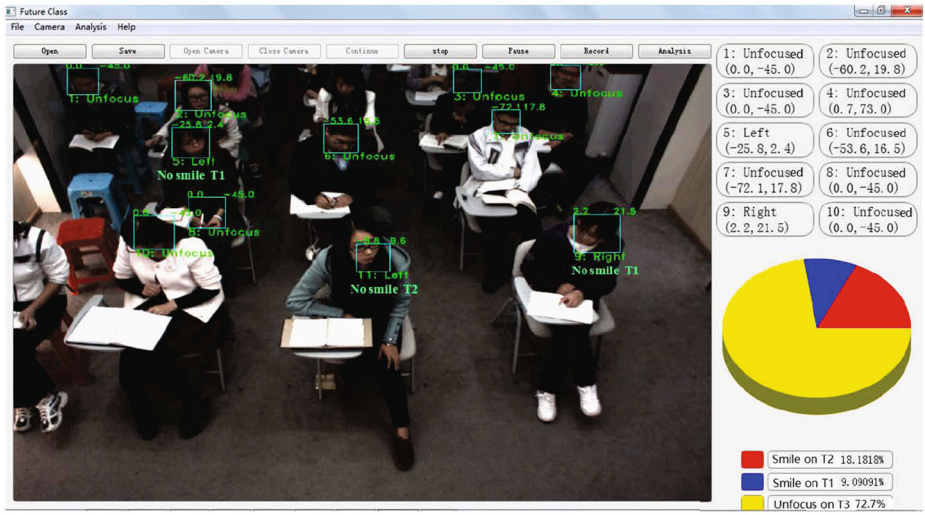
To evaluate actually the integrated class engagement, every student's learning records were stored on the system, then an expert in the area of education and human behavior helped in interpreting a class engagement for the classification in "High", "Moderate", or "Low" the degrees based on the obtained multi-cues of students. Additionally, in order to get the actual feed from the student, we send 100 questionnaires to students and teachers to access the true class affect and engagement distribution. According to the survey results and expert analysis, the average accuracy of student's engagement distribution is over 70% on the 8 real class videos. It objectively reflects the validity of our system assessment. In future work, we will do detailed and large-scale case assessment and validation for improving continuous class engagement study based on big data achievement using the intelligent system in the future version.

Table 8 An example of class engagement analysis with a student at different time (Am) of the class

Engagement analysis	Time 9:10	Time 9:20	Time 9:30	Time 9:40
Behavioral engagement	Negative	Negative	Positive	Negative
Cognitive engagement	T_1	T_3	–	T_2
Affective engagement	Non-smile	–	–	Non-smile
Engagement level	Moderate	Low	High	Moderate



(a)



(b)

Fig. 19 Examples of multi-students' cognitive and affective engagement recognition in the intelligent learning environment. When one's behavioral engagement is negative, his/her cognitive and affective engagement can be recognized by computer vision analysis

Table 9 The statistic distributions of student class engagement levels in 4 real-class videos

Class engagement level	video1	video2	video3	video4
High	23.2	28.9	37.0	39.5
Moderate	46.8	30.5	42.4	20.3
Low	30.0	40.6	20.6	40.2

8 Conclusion

In this paper, we propose an intelligent system for student engagement study, which automatically detects and analyses multiple learning cues based on five modules, i.e., attendance management, teacher-student (T&S) communication, VFOA recognition, smile detection and engagement analysis. We thoroughly evaluated each module for engagement study on some public available datasets and real video sequences in class. The proposed system achieved much improved performance and great robustness on each module, with an average accuracy of 95.6% on the student position, 67.8% on VFOA recognition, 94.2% on spontaneous smile detection and 70% on the class engagement analysis. The experimental results suggest that the proposed intelligent system can automatically analyze student class engagement objectively and effectively.

Our contributions are as follows. First, we propose an intelligent system that can automatically detect and analyze student class engagement objectively and effectively. Second, we investigate student engagement from multi-cues of information, including student attendance, T&S communication, VFOA recognition, smile detection and engagement analysis. Third, we propose approaches to recognize students' cognitive and affective states using head pose estimation, VFOA recognition and smile detection in the cclassroom scene. Finally, we present a novel tree-structural class engagement decision model to analyze a student engagement level based on one's behavioral engagement, cognitive engagement and affective engagement in class.

A preliminary evaluation has been carried out in the local intelligent class environment. The promising results suggest that the proposed system could detect and analyze multiple learning cues for student class engagement study objectively and effectively. It can help that a teacher understand students' learning performance and analyze students' engagement levels in real-time class. In future, we will work towards a large-scale class study where our proposed engagement study system will be carried out with more data in different class environments. The impact of the interactive learning and learning efficiency will be assessed through a range of measures including pre-tests and post-tests of various class sorts, along with analysis of the recorded engagement level.

Acknowledgements This work was supported by the National Social Science Foundation of China (Grant no. 16BSH107).

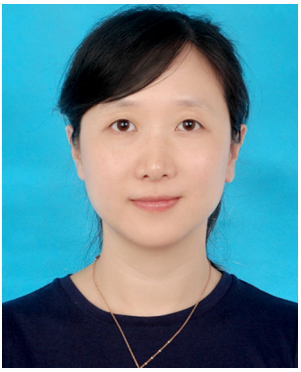
References

1. Agarwal M, Agrawal H, Jain N, Kumar M (2010) Face recognition using principle component analysis, eigenface and neural network. In: International Conference on Signal Acquisition and Processing, pp 310–314
2. An L, Yang S, Bir B (2015) Efficient smile detection by extreme learning machine. Neurocomputing 149:354–363
3. Ba SO, Odobez JM (2011) Multiperson visual focus of attention from head pose and meeting contextual cues. IEEE Trans Pattern Anal Mach Intell 33(1):101–116
4. Boyle JT, Nicol DJ (2003) Using classroom communication systems to support interaction and discussion in large class settings. Research in Learning Technology 11(3):43–57
5. Bradsky GR (1998) Computer vision face tracking for use in a perceptual user interface. Intel Technol J 2:1–15
6. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
7. Chen D, Hu Y, Wang L, Zomaya AY, Li X (2017) H-PARAFAC: hierarchical parallel factor analysis of multidimensional big data. IEEE Trans Parallel Distrib Syst 28(4):1091–1104

8. Chen J, Chen D, Li X (2014) Toward improving social communication skills upon multimodal sensory information. IEEE Trans Ind Inf 10(1):323–330
9. Chen J, Luo N, Liu Y, Liu L, Zhang K, Kolodziej J (2016) A hybrid intelligence-aided approach to affect-sensitive e-learning. Computing 98(1-2):215–233
10. Cohen M, Shimshoniand I, Rivlin E (2012) Detecting mutual awareness events. IEEE Trans Pattern Anal Mach Intell 34(12):2327–2340
11. Dantone M, Gall J, Fanelli G, Van Gool L (2012) Real time facial feature detection using conditional regression forests. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2578–2585
12. Gall J, Lempitsky V (2013) Class specific hough forests for object detection. In: Decision Forests for Computer Vision and Medical Image Analysis, pp 143–157
13. Gourier N, Hall D, Crowley J (2004) Estimating face orientation from robust detection of salient facial features in pointing. In: International conference on pattern recognition Workshop on Visual Observation of Deictic Gestures, pp 1379–1382
14. Graesser A, Chipman P, King B (2007) Emotions and learning with auto tutor. Frontiers in Artificaial Intelligence and Applications 158:569
15. Guerra J, Hosseini R, Somyurek S, Brusilovsk P (2016) An intelligent interface for learning content combining an open learner model and social comparison to support self-regulated learning and engagement. In: Proceedings of the 21st International Conference on Intelligent User Interfaces. ACM, pp 152–163
16. Gui J, Liu T, Tao D, Sun Z, Tan T (2016) Representative vector a unified framework for classical classifiers. IEEE Trans Cybernet 46(8):1877–1888
17. Guo P, Kim J, Rubin R (2014) How video production affects student an empirical study of mooc videos. In: Proceedings of the first ACM conference on Learning@ scale conference. ACM, pp 41–50
18. Huang GB, Ramesh M, Berg T, Learned-Mill E (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst
19. Ito A, Wang X, Suzuki M, Makino S (2005) Smile and laughter recognition using speech processing and face recognition from conversation video. In: IEEE International Conference on Cyberworlds, pp 437–444
20. Kahou SE, Froumenty P, Pal C (2014) Facial expression analysis based on high dimensional binary features. In: European Conference on Computer Vision, pp 135–147
21. Koji Y, Atsushi I, Shinji T, Hiroharu K (2016) Analysis of computer event logs to assess student engagement in classroom: a case study in the united states. In: 12th International Conference on Natural Computation Fuzzy Systems and Knowledge Discovery, pp 2098–2101
22. Liu Y, Chen J, Shan C (2014) Dirichlet-tree distribution enhanced random forests for facial feature detection. In: IEEE Conference on Image Processing, pp 235–238
23. Liu Y, Chen J, Gong Y, Zhang K, Liu L, Luo N (2015) Robust head pose estimation using dirichlet-tree dis-tribution enhanced random forests. Neurocomputing 127:42–53
24. Liu Y, Xie Z, Chen J (2016) An intelligent learning system for supporting interactive learning through student engagement study. IEEE, pp 618–623
25. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews (2010) The extended cohn-kanade dataset (ck+) a complete dataset for action unit and emotion specified expression. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp 94–101
26. Luo C, Wang Z, Wang S (2015) Locating facial landmarks using probabilistic random forest. IEEE Sign Process Lett 22(12):2324–2328
27. Mi J, Lei D, Gui J (2013) A novel method for recognizing face with partial occlusion via sparse representation. Optik-International Journal for Light and Electron Optics 124(24):6786–6789
28. Odobez JM, Ba SO (2007) A cognition and unsupervised map adaptation approach to the recognition of focus of attention from head pose. In: IEEE International Conference on Multimedia and Expo, pp 183–191
29. Podder P, Paul M, Debnath T, Murshed M (2015) An analysis of human engagement behaviour using descriptors from human feedback, eye tracking,and saliency modelling. In: Digital Image Computing: Techniques and Applications, pp 1–8
30. Rodgers T (2008) Student engagement in the e-learning process and the impact on their grades. International Journal of Cyber Society and Education 1(2):143–156
31. Scornavacca E, Huff S, Marshall S (2007) Developing a sms-based classroom interaction system. In: Proceedings of Mobile Learning Technologies and Applications, pp 47–54
32. Senechal T, Turcot J, Kaliouby R (2013) Smile or smirk? automatic detection of spontaneous asymmetric smiles to understand viewer experience. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp 1–8
33. Shan C (2012) Smile detection by boosting pixel differences. IEEE Trans Image Process 21(1):431–436
34. Siau K, Sheng H, Nah FF (2006) Use of a classroom response system to enhance classroom interactivity. IEEE Transactions on Education 49(3):398–403

35. Ting C, Cheah W, Ho C (2013) Student engagement modeling using bayesian networks. In: IEEE International Conference on Systems, Man, and Cybernetics, pp 2939–2944
36. Vazquez Rodriguez CA, Mejia Lavalle M, Pinto Elias R (2015) Modeling student engagement by means of nonverbal of nonverbal behavior and decision trees. In: International Conference on Mechatronics, Electronics and Automotive Engineering, pp 81–85
37. Viola M, Jones MJ (2003) Fast multi-view face detection. Mitsubishi Electric Research Lab TR-20003-96 3:14
38. Woolf B, Burleson W, Arroyo I (2009) Affect-aware tutors: recognising and responding to student affect. Int J Learn Technol 4(3-4):129–164
39. Yang C, Duraiswami R, Davis L (2005) Efficient mean-shift tracking via a new similarity measure. In: IEEE conference on Computer Vision and Pattern Recognition, pp 176–183
40. Zhang T, Zheng W, Cui Z, Zhong Y, Yan J, Yan K (2016) A deep neural network driven feature learning method for multi-view facial expression recognition. IEEE Trans Multimedia 18(12):2528–2536
41. Zhao X, Kim T, Luo W (2014) Unified face analysis by iterative multi-output random forests. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1–8

Yuanyuan Liu received B.E. degree from NanChang University, NanChang, China, in 2005, M.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 2007, and Ph.D degree from Central China Normal University. She is currently a lecturer in China University of Geosciences. Her research interests include intelligent system, computer vision and pattern recognition.



Jingying Chen received the bachelor's and master's degrees from the Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore, in 2001. She was a Post-doctor in INRIA, France, and a Research Fellow with University of St Andrews and University of Edinburgh, U.K. She is currently a Professor with the National Engineering Center for E-Learning, Central China Normal University, China. Her research interests include image processing, computer vision, pattern recognition, educational technology and human-machine interface.

Mulan Zhang received B.E. degree in communication engineering from Huazhong University of Science and Technology, Wuhan, China, in 2015. She is currently working for M.S. degree in the National Engineering Research Center for E-Learning, Central China Normal University, China. Her research interests include computer vision and machine learning.

Chuan Rao received B.E. degree from HuBei University, WuHan, china, in 2015. He is currently a graduate in Central China Normal University for M.E. degree. His research interests include information technology in education and image processing.