Techniques for Enhancing Pervasive Learning in Standard Natural Classroom

Chenping Lu, Jiaji Zhou, Liping Shen, and Ruimin Shen

Computer Science Department, Shanghai Jiao Tong University, Dong Chuan Rd. 800, 200240, Shanghai, China
fullfree@sjtu.edu.cn, zhoujiaji@live.cn

Abstract. Standard Natural Classroom (SNC) is a real-time classroom based on smart space and design principles of e-learning, aiming at creating face-to-face, interactive and pervasive learning scene for students who are far from live classroom. We use various techniques in developing different kinds of components in SNC. Two components among them are specially described in this paper: E-pen and Emotion Understanding. E-pen focuses on helping teachers mark on the projection screen and several recognition algorithms are mentioned, while Emotion Understanding focuses on affective learning and is used to estimate students' emotion.

Keywords: SNC, E-pen, Emotion Understanding, e-Learning

1 Introduction

Learning is a process of interaction, communication and thinking. In a world of information and knowledge today, people have the requirement to be educated whenever they want and wherever they are in order to keep up with the changing world. E-Learning is what we have created to fulfill such a requirement. In the classical e-learning environment, however, there exists a critical defect that teacher and students are lack of interaction. Standard Natural Classroom (SNC) is a concept proposed by our e-Learning Laboratory, which tries to create pervasive learning scene for students who are far from live classroom.

In this paper, we focus on two techniques, so called components, contained in SNC as instances of technical details: E-pen and Emotion Understanding. They help SNC to provide integrated and friendly teaching environment. E-pen system consists of a laser pen, a digital camera connected to the computer and software to process the video captured by the camera. E-pen enables teachers to "write" directly on the projection by the laser pen which solves the problem that electrical courseware is not as interactive as the blackboard. Emotion understanding system is made up of physiological sensors and algorithms to process data received from the sensors attached to students in SNC. It helps teachers to grip the emotional states of students so that teachers could deliver courses in the best way that students could accept.

Cooperation among all components in SNC, however, is of great importance and an information fusion system is being developed.

2 Background

Our e-Learning system is kind of pervasive learning platform. It extends the real classrooms by using pervasive computing technologies. The system architecture of the platform is composed of three main parts: a) distributed Standard Natural Classrooms (SNC), providing human-machine interaction and context-aware services for teachers and students; b} media streaming for multi-mode terminals delivering interactive lectures; c) personalized web-based learning with dynamic learning services, collaborative learning communities, and personalized recommendations.

The core of our e-Learning system is SNC, which is equipped with numerous smart devices or sensors and specially developed software. We try to digitalize interactive lectures and deliver them to different kinds of terminals live online. Meanwhile, offline students may also attach vivid classrooms by different kinds of techniques in our e-Learning system. So well-designed SNC could help us provide lively and natural learning scene, which students certainly prefer.

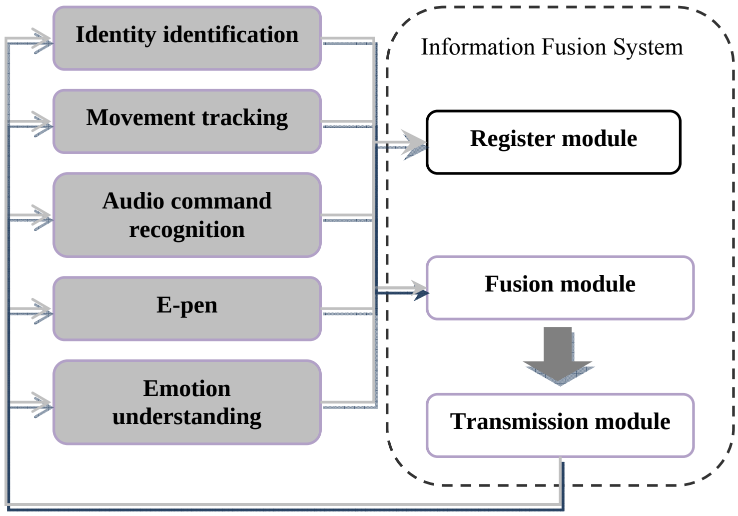

Fig. 1. Design of Standard Natural Classroom

As is shown in Fig.1, SNC has five components and an important information fusion system:

1. **Identity identification.** Identity is important information to context aware applications or other components in SNC, especially identities of teachers.
2. **Movement tracking.** In normal classroom, teacher often moves arbitrarily. It is helpful for students that one video flow always focuses on the teacher in order to accurately record what has happened on him.

3. **Audio commands recognition.** It is interesting to recognize teachers' frequent used commands, which may be used further to free hands. For example, we try to use the recognized content from teacher's audio to tone up the content-based multimedia retrieval.
4. **E-pen.** When teacher is away from dais, he may control his courseware or simulate handwriting on the projection drawing by a simple laser-pen.
5. **Emotion understanding.** Emotion understanding component is to inform the teacher the emotional states of the students.
6. **Information fusion system.** This system is used to collect and fuse data, thus make all components communicate with each other easily. For the moment, we have a prototype system called *InfoIntegrator*. As is shown in Fig.1, a register module manages the elements; a fusion module receives and fuses information from elements; a transmission module transmits information to elements. Besides, a special self-defined communication protocol is used inside this system.

In this paper, we'll focus on techniques of SNC for enhancing pervasive learning. So the information fusion system will not be deep discussed. Due to the space limit, details of *E-pen* and *Emotion Understanding* will be discussed in the next sections as examples of components implementation.

3 Design and Implementation of E-pen

3.1 Requirements of E-pen

In a classroom, when the teacher is away from dais, he can control his courseware or simulate handwriting on the projection drawing by a simple laser-pen. Therefore, E-pen may help making lessons more freely and personally in SNC. To achieve this, E-pen should not only work properly and efficiently itself, but meet the needs of information fusion in SNC as well, so as to cooperate with other components.

3.2 Structure and Implementation Details of E-pen

The actual scene of E-pen is shown in Fig.2.

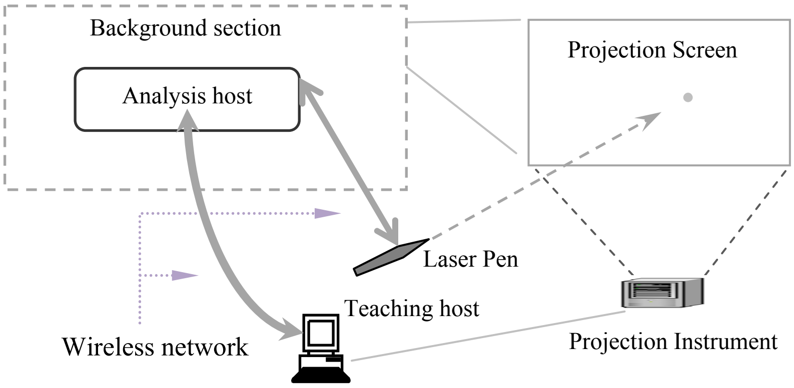

Fig. 2. Actual Scene of E-pen

3.2.1 Work Flow of E-pen

1. Background Section in Actual Scene of E-pen

As shown in Fig.2, an indication tool, such as a laser-pen, is used to produce light points on the projection screen. A camera is used to sample the video information of the projection screen and transmit the video information to *the analysis host*, who runs *the light point recognition algorithm* to get the light point position. Before introduction to this algorithm, some specific words are explained as follows:

Light Point Position (LPP): The target position to find in this algorithm.

Suspicious Position (SP): Potential light point position which needs further estimation.

The images, or called frames, caught from camera arrive 1-by-1, each of which will be dealt with through the steps in Table-1.

The entire algorithm can be concluded as: scan-by-line, merge-across-line. In the implementation, we do some optimizations mentioned below, together with the Time Complexity analysis:

Table 1. Process of Light Point Recognition Algorithm on One Frame

Step	Action	Additional aspect
i	Get the current frame as: img_1	
ii	Compare old img with img_1, get all the possible light point coordinates as *SPs*.	*SPs* are estimated based on gray level.
iii	Pick out SPs aggregated together and record the average coordinates of them to represent LPPs.	Details are omitted here to avoid verbosity and will be given in other papers before long.
iv	Reject those LPPs with unreasonable coordinates.	Reject those with too more or too less SPs and those cover too large area.
v	Get new img by old img and img_1	$img(i,j)=img(i,j)\times a+img\_1(i,j)\times(1-a)$

Step (ii), (iii) and (v) can be executed pixel-by-pixel. So each pixel in one frame would be considered only once. It is easy to have that the Time Complexity is O(MN), if we neglect Step (v), where M, N represent the width and height of frame respectively, determined by the camera. Step (v) could be done independently after all other steps and does not influence the Time Complexity.

2. Other Sections in Actual Scene of E-pen

Besides background part, other parts are working as follows:

Light point positions are transmitted to the teaching host through network, point-by-point. When a point arrives, the teaching host does a coordinate-conversion between camera video and projection screen. Then it begins to display the track information of those points. Two steps are needs in display:

First, point track is formed by connecting points one-by-one, using simple lines or curves. At the same time, we offer another function for users: light point position represents mouse position of teaching host, and if light point disappears, a left-click event occurs in teaching host. Users may choose from those two functions. The Second step below works only if the former function is chosen.

Second, after getting numbers of points, we try to find track information in those points. If succeeded, we use those tracks, for instance, simple graphs or letters, to replace the track in the first step. This step could be partitioned as below:

1. Points partitioning. There are two ways to choose: Time-based partition and User-behavior-based partition.
2. Do track recognition for each partition based on the track-type users set in advance. Types of track include simple shapes, digits and letters. In E-pen, handwriting recognition differs from [3], [6] because of the actual scene we've set up. We modify traditional BP networks and feature sets but it is still in developing as we'll mention in experimental results below. Details of our method would be given before long if we get the chance.
3. Process track optimization if necessary. Use improved light point track form to display instead of printed form.

3.2.2 Communications between E-pen and Other Components in SNC or Users

Accepted commands to receive from users or information transmission module: start E-pen; stop E-pen; change track-type.

Accepted information to be sent out from E-pen to information fusion module: E-pen started, E-pen stopped, track-type changed.

3.3 Experiments and Results

We've completely implemented the light point recognition algorithm and done some tests under different conditions. The results are shown in Table-2. Every test lasts for 2 minutes.

Table 2. Results of Light Point Recognition Algorithm Test

No.	Screen Size (Meters)	User Position (from Screen)	Result (fps)	Camera (fps)	Recall	Precision
1	2×2	2m	20	25	80%	100%
2	2×2	10m	20	25	80%	100%
3	1×0.8	2m	22	25	88%	100%
4	1×0.8	10m	22	25	88%	100%

We may conclude that the distance between the user and the screen doesn't influence the precision or the recall, but obviously, the further users stand from screen, the harder indication tools would be to use.

The smaller screen gets a higher recall in test. However, the precision is very good and we may summarize that the light point recognition algorithm is feasible. On the other hand, 20 or higher fps is enough for use, so noise-reduction has been added after light point recognition in order to reject the points that possibly results from the shake of hand and finally we get a frame-rate of 10 fps.

Track-recognition tests have been partly finished and we've got some results in Table-3, all of which are tested under Condition 1 of Table-2. It is inspired that shape recognition and digit recognition are practical to some extent.

Table 3. Results of Track-recognition Test

Test object	Classification object	Partition	Error rate
Simple shapes	ellipse, line, triangle, rectangle	User-behavior-based partition	6.5%
Digits	10 digits	User-behavior-based partition	5%
Letters	26 capitals or lowercases	Time-based partition	In developing

4 Preliminary Results of Emotion Understanding Component

People tend to focus on the cognitive ingredient of learning process, while most of them neglect the affective factor. "…expert human tutors... devote at least as much time and attention to the achievement of affective and emotional goals in tutoring, as they do to the achievement of the sorts of cognitive and informational goal…", concluded in the early work of intelligent tutoring systems by Lepper M. R. and Chabay R. W.[12].

We have noticed the importance of affective factor of learning and launched a research project trying to develop a system to inspect the emotional states of learners during the learning process. The emotion understanding module will be integrated into SNC as the result of the project. In this paper we want to show some preliminary results of the project under progress.

4.1 Emotion Model

Several previous works in emotion detection and measuring are listed in Table-4.

Table 4. Related Works

Method	Related Works
EEG	Alicia Heraz et al, *Emomental Agent* [7]
Skin Conductivity	R. W. Picard et al., *Galvactivator* [15] ; Liping Shen et al., *XVast* [8]
Facial Image	Arman Savran et al. [9]
Pressure	C. J. Reynolds, *Pressure Mouse* [10]

These works show a strong implication that human emotion relates tightly with physiological signals. We believe that internal emotional state of a subject could be revealed by analyzing his or her physiological responses. Based on this hypothesis, we designed a system trying to find out the relationship by applying machine learning algorithms.

We model the emotional state of a person as a pair of sets:

$$\mathbf{S=\{P,E\}} \tag{1}$$

P represents a set of *Physiological Signals*. **E** is a set of *Emotional States*. They are described in detail in the next two sections.

Emotional State Set (E)

The emotional state set defines emotion directly related to our research.

Russell's circumplex model of affect [11] is our reference model since we consider it simple and suited our purpose. Russell's model divides human emotion into two base states: *Valance* and *Arousal*. All emotions are represented as a dot in a 2D coordinate with x-axis representing *Valance* and y-axis representing *Arousal*. (Fig. 3)

Fig. 3. Russell's Circumplex Model

Though simple, Russell's emotion set are too rich for us, we need to select some particular emotions out which relate most tightly to our concern. In the previous work by Liping Shen et al. [11], they confine the emotion set to: *interest, engagement, confusion, frustration, boredom and hopefulness*. We reduce it more, we only consider four of them since *engagement* is more often mixed with *interest* and also *boredom* is more often mixed with *frustration*. The emotional state set is shown in Table-5.

Table 5. Emotional State Set

Emotional State Set	Description
Interest	Curious about the new knowledge, attentive, eager to learn
Confusion	Faced with problems, trying to solve the problems
Frustration	Completely unable to understand the course material, reluctant to learn
Hopefulness	Difficulties solved, pleased with the new findings, willing to explore more.

Physiological Signal Set (P)

Human affection is subtle and complex, no exception in the process of learning. Multiple-aspects of physiological signals are needed. On account of this reason, we choose to sample multi-channel signals of learners including SC (skin conductivity), BVP (Blood Volume Pressure) and EEG (electroencephalography).

Some frequencies of brainwave show tight relation with the brain activity (Table-6) so that we separate them from the raw EEG data.

Table 6. Brainwaves

Wave Type	Frequency	When wave is dominant
Delta	0-4 Hz	Deep sleep
Theta	4-8 Hz	Creativity, dream sleep, drifting thoughts
Alpha	8-12 Hz	Relaxation, calmness, abstract thinking
Beta	+12 Hz	Relaxed focus. High alertness, mental activity.

By putting these signals together, the Physiological Signal Set is composed of:

$$\mathbf{P} = \{(\text{SC, BVP, EEG\_RAW, Alpha, Beta, High Beta, Theta, Delta})\} \quad (\mathbf{2})$$

Table-7 describes the definition of each physiological signal in detail.

4.2 Experimentation

The experimentation is conducted by simulating a virtual distant education environment. The learner watches video records from a real classroom and each learning session takes 40 minutes which is just the same length as a real class in order to get emulational result. We conducted four separated learning sessions altogether with video records from an algorithm course, a mathematic course, a Chinese history course and an economic course respectively.

Table 7. Physiological Signal Set

Signal	Description	Unit
SC	Skin Conductivity	μS
BVP	Blood Volume Pressure	N/A
EEG Raw	Raw brain wave	μV
Delta	Delta wave 2-4 Hz	μV
Theta	Theta wave 4-8 Hz	μV
Alpha	Alpha wave 8-13 Hz	μV
Beta	Beta wave 15-20 Hz	μV
High Beta	High Beta wave 20-40 Hz	μV

ProComp5 Infiniti™ encoder [16], a multi-modality device, is used to collect real-time physiological data of the learner during the experimentation.

To record the emotional state during the learning process, we arrange an assistant to watch the learner from a distance who is responsible for estimating and noting down the emotional state of the learner by his judgment. However we notice that even a well trained psychologist may not be able to catch every emotion transition by watching the learner, we ask the learner to do some self-evaluation during the learning process. In order not to attract the learner from learning to recording of emotional state, we design a program to help the learner. (Fig. 4) The learner needs only to click on a colored button once he feels the transition of his emotion. The program records the emotional states and the timestamp when the transition of states occurs. After a session is completed the program will generate a report of the emotional states of the learner during the learning session.

Fig. 4. Program Records Emotional States

We merge these two reports together to generate a summary of the emotional states and combine the emotional states to the data recorded by matching the timestamp when each piece of data is recorded. If the two reports conflict with each other, we trust the learner's self-evaluation.

4.3 Data Preprocessing

Raw data needs preprocess before they can be inputted into the algorithm.

Because EEG signal is so weak that prone to be interfered by tiny movements of the learner, we need first do some selection on the raw data by reviewing the recorded EEG waveform to delete obviously interfered data sections.

After removing noise data by naked eyes, we use algorithms to smooth the data.

4.4 Algorithms

We applied k-nearest neighbors (kNN)[13] and SVM to the preprocessed data, 80% of which are used as training data and the rest are used as testing data.

The maximum accuracy is reached when k is set between 2 and 4. Fig. 5(a) shows the results when k is set to 3.

We use libSVM [14] library and choose radial basis function as the kernel. Fig. 5(b) shows the results by applying SVM.

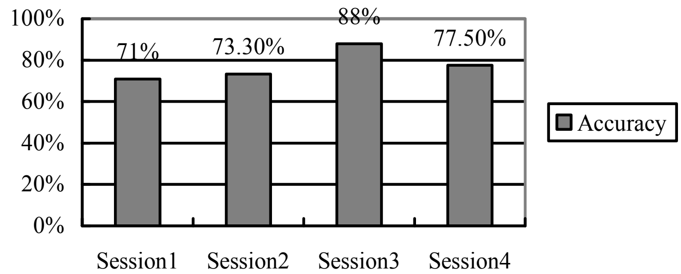

Fig. 5(a). kNN Results

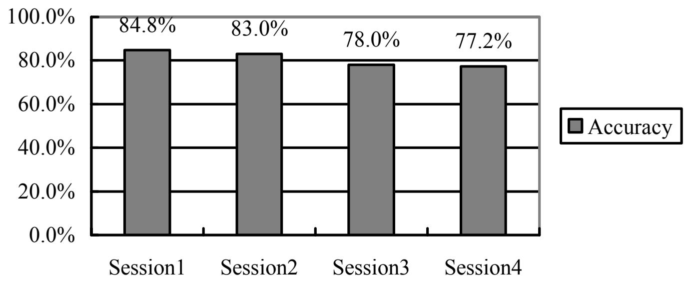

Fig. 5(b). SVM Results

4.5 Discussion of Results

Obviously the results strongly support our hypothesis that physiological responses of human body do reflect his or her emotional state.

This initial experimentation shows that SVM gives better result and is much more stable than kNN and is very prospective to be the algorithm we use for future research. The model we are currently using, however, is still rough and needs to be evolved. We expect higher accuracy so that we could apply the model to practice.

5 Conclusion and Future Work

We've introduced Standard Natural Classroom, a core member in our E-learning lab, which is constructed based on smart space and E-learning concept. Two technical components in SNC have been put forward: E-pen and Emotion Understanding. Our works on those two components results in some practical sub-system or integrated idea. Definite goals of them are also brought forward, which means further efforts on framework or experiment are in process.

The current E-pen component has already become practical except for part of the functions in track-recognition. Communications between E-pen software system and other components or users have also been completed. Our future work includes:

- Create a self-developed laser-pen instead of current ones, in order to facilitate the communication between users and teaching host and provide more functions.
- Continue implementing track-recognition and track-optimization algorithm. We are trying to specialize the methods mentioned in [4] and [5] so as to make them suitable for digit recognition and letter recognition in E-pen, respectively.

Our future work on emotion understanding component includes improving the model, as well as applying the emotion sensing to the actual classroom education or asynchronous self learning. We are planning to simply feed back the students emotions back to the lecturer in real-time, that the lecturer would adapt the lecture style, speed and content based on the students' emotional statistics. We also plans to provide personalized service based on the learner's emotions. This service will incorporate the learner's emotional states together with the learner's cognitive abilities, and his/her learning goals, to generate appropriate responses to the learner.

References

1. Picard R W, Papert S, Bender W, etc., Affective Learning - manifesto, BT Technology Journal, Vol. 22 No. 4, October 2004
2. Ruimin Shen, Fan Yang, Peng Han, An open learning model for web-based learning architecture, Web-Based Learning Theory, Research, and Practice, Shanghai, China, 2006
3. Y. Le Cun, B. Boser, J. S. Denker, et al. Handwritten Digit Recognition with a Back-Propagation Network, Advances in neural information processing systems 2, 1990
4. L. Bottou, C. Cortes, J. S. Denker, H. Drucker, et al. Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition. the International Conference on Pattern Recognition, pp. 77-82, Jerusalem, Israel, October 1994

5. J.J.O. Júnior, J.M. de C. Carvalho, C.O.A. Freitas, R. Sabourin, Evaluating NN and HMM Classifiers for Handwritten Word Recognition, the XV Brazilian Symposium on Computer Graphics and Image Processing, IEEE, 2002, pp. 210–217.
6. de Oliveira, J.J., Jr. de Carvalho, J.M. de A Freitas, C.O. Sabourin, R. Feature Sets Evaluation for Handwritten Word Recognition, Frontiers in Handwriting Recognition, 2002.
7. Alicia Heraz, Ryad Razaki and Claude Frasson, "Using machine learning to predict learner emotional state from brainwaves", In Proceedings of ICALT, Japan, 2007
8. Liping Shen, Enrique Leon, Victor Callaghan, Ruimin Shen, "Exploratory Research on an Affective E-Learning Model", In Proceedings of Workshop on Blended Learning, 2007
9. Arman Savran et.al. , "Emotion Detection in the Loop from Brain Signals and Facial Images", In Proceedings of eNTERFACE'06, Dubrovnik, 2006
10. C. J. Reynolds, The Sensing and Measurement of Frustration with Computers. Master thesis, Massachusetts Institute of Technology
11. Russell, J.A. (1980) A Circumplex Model of Affect, Journal of Personality and Social Psychology, 39:6, pp. 1161-1178, American Psychological Association.
12. Lepper M R and Chabay R W: 'Socialising the intelligent tutor: bringing empathy to computer tutors', in Mandl H and Lesgold A (Eds): 'Learning issues for intelligent tutoring systems', pp 242—257 (1988).
13. Dasarathy, B.V., Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, ISBN 0-8186-8930-7, 1991.
14. Chih-Chung Chang and Chih-Jen Lin, http://www.csie.ntu.edu.tw/~cjlin/libsvm/
15. Picard R W and Scheirer J: 'The Galvactivator: a glove that senses and communicates skin conductivity', 9th ICHCI, New Orleans (August 2001).
16. Thought Technology Ltd., http://www.thoughttechnology.com, Canada