Towards Intelligent Interaction in Classroom

Pengfei Xu, Guanghui Han, Wen Li, Zhongke Wu, and Mingquan Zhou

College of Information Science and Technology
Beijing Normal Univiersity, China
pengfeixu@gmail.com, hangh@mail.bnu.edu.cn, liwenbnu@gmail.com,
zhongkewu@gmail.com, mqzhou@bnu.edu.cn

Abstract. In classroom environments, complex and valuable communication takes place. To augment and record these communications effectively, various computer-based systems were deigned in the past decade. In fact, with advancements of multimedia technology and interaction technologies, research in this field has already brought some of these systems into regular usage. The main contribution of this paper is to give an overview of the human-computer interaction technologies and approaches used in intelligent classroom systems. Current challenges in intelligent interaction in classroom are also discussed. Improving these interaction techniques has a significant effect on the overall system performance and user experience.

1 Introduction

In the past decade, various intelligent systems were developed for augmenting and recording lecture activity in the classroom environment. Classroom 2000 project [1] at Geogia Tech, for example, views the teaching and learning as a multimedia authoring activity, and can capture not only the audio stream, video stream in classroom, but also the annotation made by instructors and students. Smart Classroom project [16], on the other hand, bridges the gap between tele-education and traditional classroom activities through applying various interaction technologies and intelligent video capture system. These systems covers a rich set of interesting topics, including their system architecture, hardware setup, indexing and summarization of captured content, delivery of content to remote audience, synchronization between audio/video streams and visual aids. However, in this paper, we will focus on three interesting modules related to the interactions between the intelligent classroom systems and human (instructors and audience in this context). For other aspects about these systems, we refer readers to [10] and [4].

Unlike in traditional HCI context, students and teachers are normally deeply engaged in the learning and teaching activity in classroom environment. This makes HCI in classroom a rather special and interesting topic. As existing systems vary significantly in their system architecture, hardware setup and operation details, review of the interaction between human and these systems is not a straightforward task. To help our discussion, we roughly categorize human-computer interaction in existing smart classroom systems into three groups of modules: virtual assistant modules, virtual pen modules, and virtual film crew modules.

C. Stephanidis (Ed.): Universal Access in HCI, Part III, HCII 2009, LNCS 5616, pp. 150–156, 2009.

In rest of this paper, different approaches and techniques to implement these modules are reviewed and discussed. To be specific, various virtual assistant techniques are reviewed in Section 2, with discussion of the challenges faced. In Section 3, different virtual pen technologies and their characteristics are investigated. Section 4 focused on the virtual film crew modules which capture lecture video automatically. Section 5 concludes this paper.

2 Virtual Assistants

In this paper, virtual assistants refer to the hardware or software modules which transform activities in classroom, into commands to hardware or software components of the computing system appropriately. It can be further classified into passive ones, which wait for user's explicit commands passively, and active ones, which can act without the need for explicit user commands. Active virtual assistant, for example, first detects events such as user logging on to the classroom computer, and then automatically turns off the lights, lowers screen, turns on the projector, and switches the projector to computer input [3]. In this case, the virtual assistant performs actively, such that instructors need not give explicit command to the system. Instead, systems often act intelligently based on certain inference-based rules, pre-defined by system designers or operators. As this kind of intelligence could suffer from incompleteness of the inferences, in practical systems, proper manual override functionality must be provided to give user a channel to revoke or change the inferred actions quickly [3].

Intelligence in passive virtual assistant, on the other hand, is normally mapping-based, instead of inference-based. The mapping is often between a vocabulary of natural instructions and a vocabulary of computer commands. Computer vision or speech recognition based techniques are normally used to perform this mapping. For example, icon recognition is used as an input method to control video playback and run PowerPoint slide shows for instructors in [5]. Instructors may draw on a traditional whiteboard, and camera will capture the whiteboard image. Going through a pipeline of image processing and computer vision processes, such as edge detection and feature matching, the icon is transformed into certain command, like playback or stop. In [16], the teacher can send a mouse click message to the computer using a traditional laser pen, by fixing the laser spot on the point of interest for a second and then circle that point. This way of clicking, however, is troublesome, as reported that some students complained the relatively difficulty of performing a simple clicking. Thus, speech recognition technique is also adopted in [16]. Speech recognition modules can recognize the voice of teacher, such as “Jump to Next Page” or “Go ahead”, into a predefined command, such as PageUp or PageDown, to the slide presentation program.

Using passive virtual assistants, teachers can input a certain command to the computer system without approaching the computer to control the standard input devices. Instead, certain commands are generated by using traditional, computer-irrelevant methods, such as drawing an icon, using laser pen, and hand gesture. However, comparing to standard input devices, computer vision based or speech recognition based input methods often suffer the issue of robustness. For example, an icon may not be well-recognized due to occlusion or lighting condition. Another class of errors occurs when non-command action, e.g. instructions to students, is taken as a command to the computer system and thus misinterpreted. The later problem can be alleviated by setting constraints of when certain input method shall be used and when shall not [5]. However, setting and remembering these constraints could be troublesome for system designers and users. Another practical but less-mentioned solution is to use composition of two input methods, for example, voice recognition and hand gesture. Assume both methods have correct detection rates of 100% and false detection rates of 10% (this number is not overstated, considering the instructor is interacting with students at the most time during the lecture), then using composition of two methods, the final false detection rate could be reduced to 1%.

3 Virtual Pens

In classroom, perhaps one of the most valuable contents is the annotations traditionally written by teachers and students. In this paper, virtual pen refers the hardware and software module to digitalize the handwriting by teachers or students in intelligent classroom system. The most direct solution to virtual pen is perhaps to use digital whiteboards or touch-sensitive screens. For example, a pen-sensitive 67-inch diagonal screen is used in Classroom 2000 project [1]. When teachers draw on the screen, the system will generate corresponding lines digitally. With these devices, recording of teacher's drawing and annotations become a trivial task. Similarly, a wall-size touch-sensitive SmartBoard device is used as the display screen in the Smart Classroom project [16]. However, these solutions not only demand costly hardware setup, but also involve maintenance issue as the device is costly and not portable.

A less costly and more portable alternative approach is to use a system comprising of a positioning subsystem and a projector subsystem. For example, taking advantage of the positioning capability of infrared-vision based Wii Remote, John Lee captures the drawing of an infrared pen digitally, and the drawing will be presented by a projector [17]. As projectors are ordinary devices in current classroom setting, thus the real cost of this system is the positioning subsystem, which is often relatively cheap and portable. However, before adopting this approach, one should pay attention to the limitations of the positioning subsystem. For example, in John Lee's solution, if instructor blocks the light from infrared pen to infrared camera while he is writing, the positioning subsystem fails and so does the whole system. Another positioning solution which uses the laser spot to position the focus point is presented in [16]. Though its advantage of low-cost, it suffers the same problem as the previous solution, and its precision and robustness may not be as good as the previous one. For example, a red point in the PowerPoint slides can be misunderstood as the red point generated by the laser pen by the vision-based Laser2Cursor algorithm.

Another notable solution of virtual pen is the Mimio Virtual Ink system [12]. In authors' opinion, the biggest advantage of Mimio Virtual Ink system over other solutions is that the annotation is recorded without changing the traditional way of madding it. In Mimio Virtual Ink system, instructors could write on a traditional whiteboard using ordinary whiteboard pens enclosed in Mimio-supplied shell. A

Mimio ultrasonic sensor attached on the side of the whiteboard could determine the position of this shell and thus record teachers' writing. This system has been used successfully for automatic data collection in the Virtual Classroom project, and has been extended to support recall annotation made on different places of the whiteboard by moving handheld devices over the whiteboard [7].

Nevertheless, the techniques mentioned above may not be suitable for recording of annotation made by students. In Classroom 2000 project [1], students' annotations are captured using pen-based computing devices. With recent progress in mobile-based interaction techniques [14], feasibility of mobile-based techniques for recording students' annotation might worth a new look.

It should also be noted that with virtual pens' positioning capability, they can be easily deployed to input commands to computer systems, just as mouse devices. In another word, virtual pens technologies can support virtual assistants' functionality. For example, the Mimio Virtual Ink system can also be used as a remote mouse and/or keyboard to control the presentation of slides.

4 Virtual Film Crews

Automatic lecture video capture is an important part of almost every intelligent classroom system. In classroom, the focus of interest changes constantly between different objects, such as the instructor's face, hand and body, the projected slides, whiteboard drawing and a model. It might not be difficult for professional cameramen and directors to produce a video tracks the objects of interest correctly and later present the lecture content to remote students in a quality comparable to the live experience. To automatic capture lecture video which is as good as the videos captured by the professionals is still a challenging task. Fortunately, with advancements in computer vision techniques, such as tracking and high-level scene understanding, the gap between computer-generated videos and human-generated videos is becoming smaller and smaller. In this section, common techniques used in virtual film crew modules are reviewed.

Instructor tracking can help the virtual film crew to target the camera to the instructor even when he/she is moving. It is perhaps the most widely adopted module in smart classroom systems, although detailed implementations of the tracking modules are different [2], [3], [5], [15], [16], [19]. For example, background subtraction based instructor tracking technique is used to help providing information to camera system in [5]. Device activities in classroom environment, for example, the instructor's use of a pen on the electronic whiteboard have also been used to provide additional information to the tracking algorithm [3].

Perhaps the most widely used hardware setting for instructor tracking is a combination of a static panoramic camera and a PTZ camera. For example, AutoAuditorium [2] used a static wide-angle camera to watch the entire area of interest, the image is then analyzed by software and the position of instructor is detected. The position information is used to point the PTZ camera to the object of interest to give a better shot of the object of interest. A similar approach is used in iCam system [15]: a static panoramic camera, which has a wide horizontal field view (74 degrees), is used to track the movement of the instructor. Tracking result is then used to guide the second camera, which is a PTZ camera, to keep the instructor at center of the view. To reduce the cost and burden of manual calibration of implanting two cameras, digital cropping and mechanical tracking could be combined together to reduce the hardware requirement from two cameras to one PTZ camera, while keeping the instructor at center of the view [18], [19]. The resolution of generated video will be smaller than the original one due to cropping.

During some periods of a lecture, for example Q&A or group discussion sessions, the virtual film crew needs to target the camera to capture the video of a specific student or a specific group of students. This task is still a challenging research problem due to variance of audience and activity. One solution is to use audio cue to help locating the students. In [15], [19], a microphone array and sound source localization algorithm is used to find the position of the student asking questions. Another possible solution is presented in [8]: wireless-based indoor positioning system is used to identify the absolute positions of all persons carrying the wireless device. This solution could provide the virtual film crew valuable information which cannot be easily detected by computer vision based technologies, such as the questioner's position in the audience or the questioner raising his/her hand, thus supports more sophisticated video productions.

Besides person tracking, event detection is another class of widely used modules in a virtual film crew. The definition of event may vary in different systems, and the corresponding event detection algorithm and behaviors of virtual film crew may also differ from each other significantly. In [6], three events are defined based on regions: "birth" of a region, "mature" of a region and "death" of a region. An attention model is then built upon these events to model where the instructor is directing the audience's attention. The position of attention is then used to guide the virtual film crew. In [13], however, events are defined as teacher's actions. The actions are recognized based on teacher's movement and extracted blackboard features. For example, the teacher's action of "writing on blackboard" is detected if 1) teacher's position is unchanged; 2) teacher's face is directing to the blackboard; 3) teacher's hand is moving; and 4) amount of characters on the blackboard is increasing. The detected event is then used to guide the selection of one appropriate camera from the three cameras in the classroom. In [16], events are defined as activities happened in the classroom, including teacher writing on board, teacher showing a model, remote student speaking and others. Device activities and hand gesture recognition module are used to detect these events, and then the detected event will determine in which mode the camera will operate. In [19], zoom level of the camera is controlled by the amount of motion detected throughout the tracking. If a speak stands still for a while, the camera will zoom in for clearer shot. If a speaker moves around, the camera will maintain a low zoom level. This can be also viewed as an event based approach.

Though automation capture systems could avoid the labor cost of cameramen and directors, the automation algorithms may also suffer from robustness problem considering the variance in classroom environments and user preferences. To cope with this problem, some systems support scripting of the system behaviors which means operators can change system behaviors according to their environment or preferences [19]. Furthermore, in FlySPEC [9], audience supervision is introduced to give a camera system with hybrid human and automatic control. It allows users to make the tradeoff between the operation effort of the camera system and the final quality of the captured video. However, in classroom environment, one should consider whether control of the camera system could become burden or distraction with respect to students' learning activity.

5 Conclusion

Intelligent classroom systems focus on improvement of interaction in classroom and recording of valuable content in classroom. During the past decade, various commercial and experimental systems were designed and some of them are now in regular usage. However, some challenges, such as the robustness issue and variance between environments, activities and preferences, are still open for further research. Furthermore, with the introduction of new instruction method into classroom environments, such as mixed reality [11], the demand and difficulty of designing capable intelligent interaction and recording systems might increases.

References

1. Abowd, G.: Classroom 2000: an experiment with the instrumentation of a living educational environment. IBM Systems Journal 38(4), 508–530 (1999)
2. Bianchi, M.: Automatic video production of lectures using an intelligent and aware environment. In: Proceedings of the 3rd international Conference on Mobile and Ubiquitous Multimedia, pp. 117–123 (2004)
3. Cooperstock, J.: Classroom of the Future: Enhancing Education through Augmented Reality. In: Proc. Conf. Human-Computer Interaction (HCI Int'l 2001), pp. 688–692. Lawrence Erlbaum Assoc, Mahwah (2001)
4. Erol, B., Li, Y.: An overview of technologies for e-meeting and e-lecture. In: Proceedings of IEEE International Conference on Multimedia and Expo (2005)
5. Flachsbart, J., Franklin, D., Hammond, K.: Improving human computer interaction in a classroom environment using comptuer vision. In: ACM Multimedia (2000)
6. Heck, R., Wallick, M., Gleicher, M.: Virtual videography. ACM Trans. on Multimedia Computing Communications and Applications 3(1) (2007)
7. Holmquist, L., Sanneblad, J., Gaye, L.: Total Recall: In-place Viewing of Captured Whiteboard Annotations. In: Proc. Conf. Human-Computer Interaction (HCI Int'l 2003), Ft. Lauderdale, Florida, USA, April 5-10 (2003)
8. Lampi, F., Kopf, S., Benz, M., Effelsberg, W.: An automatic cameraman in a lecture recording system. In: Proceedings of EMME 2007, pp. 11–18 (2007)
9. Liu, Q., Kimber, D., Foote, J., Wilcox, L., Boreczky, J.: FLYSPEC: A Multi-User Video Camera System with Hybrid Human and Automatic Control. In: Proceedings of the 10th international conference on Multimedia, France (2002)
10. Liu, T., Kender, J.R.: Lecture videos for e-learning: Current research and challenges. In: Proceedings of IEEE International Workshop on Multimedia Content-based Analysis and Retrieval (2004)
11. Liu, W., Cheok, A., Lim, C., Theng, Y.: Mixed Reality Classroom: Learning from Entertainment. In: Proceedings of DIMEA 2007 Perth, Western Australia (2007)
12. Mimio Virtual Ink, `http://www.mimio.com/index.shtml`

13. Onishi, M., Fukunaga, K.: Shooting the lecture scene using computer controlled cameras based on situation understanding and evaluation of video images. In: International Conference on Pattern Recognition, ICPR (2004)
14. Ballagas, R., Borches, J., Rohs, M., Sheridan, J.: The Smart Phone: A ubiquitous input device. IEEE Pervasive Computing Magazine (2008) (January-March 2006)
15. Rui, Y., Gupta, A., Grudin, J., He, L.: Automating lecture capture and broadcast: technology and videography. ACM Multimedia Systems Journal 10(1), 3–15 (2004)
16. Shi, Y., Xie, E.A.: The smart classroom: Merging technologies for seamless tele-education. IEEE Pervasive Computing Magazine (April-June 2003)
17. Low-cost multi-point interactive whiteboards using Wiimote: `http://www.cs.cmu.edu/~johnny/projects/wii/`
18. Yokoi, T., Fujiyoshi, H.: Virtual camerawork for generating lecture video from high resolution images. In: Proceedings of IEEE International Conference on Multimedia and Expo (2004)
19. Zhang, C., Rui, Y., Crawford, J., He, L.: An Automated End-to-End Lecture Capture and Broadcasting System. Transactions on Multimedia Computing, Communications, and Applications (January 2008)