An Automatic Lecture Recording System Using Pan-Tilt-Zoom Camera to Track Lecturer and Handwritten Data

Hsien-Chou Liao[*], Ming-Ho Pan, Min-Chih Chang, and Kun-Wei Lin

Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taiwan, R.O.C.

Abstract: For automatic recording of the traditional lecture with a wide blackboard by using a Pan-Tilt-Zoom (PTZ) camera, a system, called ACRS (After Class Reviewing System), was proposed. From the point of view of students reviewing the lecture, the handwritten data is more important than the video of the lecturer. So, the handwritten data and the lecturer were considered together to record a lecture most suitable for student review. In the experimental study, the video recorded by ACRS was compared with that recorded manually by a cameraman. About 83% of the video time was close to the video recorded manually and suitable for review by students.

Keywords: Lecture recording; automatic lecture tracking; video recording; object tracking.

1. Introduction

   Lecture recording is a very important function used in distance learning and e-learning. Cameras are the common recording devices used to capture multimedia content, such as the lecturer, projected slides, handwritten data, or even the participants/students joining the lecture. Generally, two types of cameras are used: fixed and active, i.e. PTZ (pan-tilt-zoom). Each is used depending upon the specific purpose of the lecture recording. Fixed cameras are usually used to record the lecturer, projected slide, and so on. If a lecturer is recorded using a fixed camera, he/she should sit or stand within a limited area; otherwise, he/she will be outside the camera's FOV (field of view). Sometimes two or more fixed cameras are simultaneously used to record all aspects of a lecture in order to make the recording more complete. Although fixed cameras are easily installed and operated for lecture recording, some problems may be encountered. For example, the combination or post-production of videos from different cameras requires additional effort. Students may feel somewhat bored while watching a video where the background does not change. It may be difficult to recognize the text or handwritten data in the video frames.
   PTZ cameras are increasingly used for lecture recording. The PTZ camera can be used either to cover a large area or to capture high resolution videos with its pan/tilt/zoom capabilities. The rotation of PTZ cameras makes the recorded video more interesting. A PTZ camera may also be incorporated with a fixed camera, thereby providing more sophisticated functions. In such a dual

---

camera system, the lecturer is usually the target to be tracked. He/she is detected in the image of the fixed camera. The PTZ camera is then controlled to keep capturing the high resolution video of the lecturer. The main problem of the PTZ camera is that it must be operated either manually by a cameraman or automatically by some control mechanism. Some work related to this problem is presented in the next section.

In this study, a single PTZ camera was used for automatically recording the traditional lecture with a wide blackboard. Although many digital teaching materials are commonly available today, the traditional blackboard lecture style is still popular. Two samples are shown in Figure 1. The left photo depicts a junior high school physics lecture and the right photo shows a digital system design lecture at our university. If one fixed camera is used to record such a lecture, as the figure shows, the handwritten data is difficult to recognize. One current solution is to install multiple fixed cameras with each camera used to capture a designated part of the blackboard. They are then switched according to the location of the lecturer which is detected by sensors, e.g. infrared sensors. The above solution is not good enough, especially when the lecturer or handwritten data is located near the boundary area of two cameras. Some studies utilized the HD (High Definition) fixed camera for lecture recording [1-2]. If such a camera is used to record a video of the whole blackboard, the resolution is still insufficient to clearly portray handwritten data in the video frames.
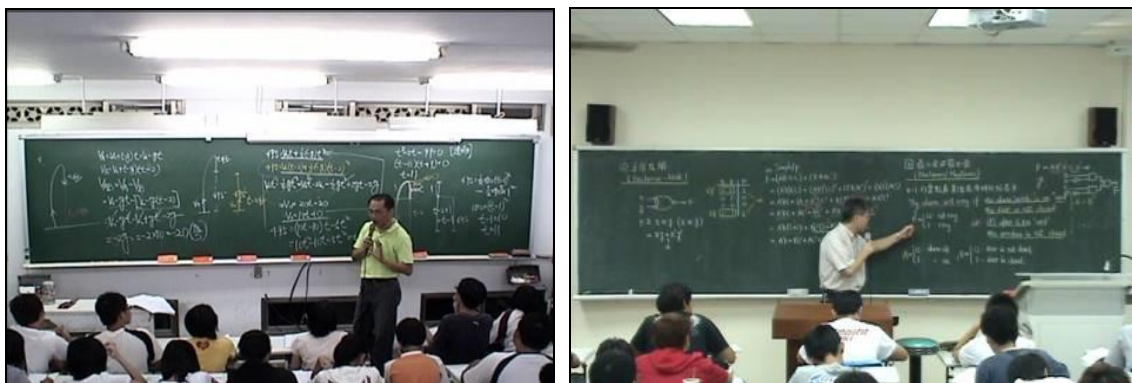

Figure 1. Two samples depicting traditional lectures with a wide blackboard

A system was therefore proposed in which a PTZ camera is controlled in such a way that the requirements for lecture recording are better satisfied. The lecturer and handwritten data are the two main components to be recorded. However, the handwritten data is more important than video of the lecturer to the students reviewing the lecture. The tracking of both handwritten data and the lecturer was considered in order to record a lecture most suitable for student review. An experiment was also designed to quantitatively evaluate the system's performance. The utilization of appropriate methods to achieve the above goal was also emphasized in this study. However, the unnecessary details of the methods used to do this will be omitted in the following sections.

The remainder of this paper has been organized as follows: Section 2 reviews related work; Section 3 presents the proposed system; Section 4 describes the experimental study; and Section 5 presents the conclusion and the need for future work as indicated by the results of this study.

2. Related work

There are several types of lecture material; lecturer/student audio/video, projected slides,

traditional or digital blackboard/whiteboard, and others. In this section, the related work of lecture recording is classified into three categories according to camera use: no camera, fixed camera, and the PTZ camera. In the first category, no camera is used for lecture recording. As an example, Zupancic and Horz proposed a system, called AOF (Authoring on the Fly) [3]. AOF recording consists of different media streams, including the lecturer's voice, projected slides, and the handwritten data presented on a digital whiteboard. It allows users to automatically produce a multi-media document with an HTML overview. Winslow et al. addressed issues to access lecture video content from mobile devices, including low bandwidth and small screen size [4]. They focused on two key elements of lecture video, slide projection and use of a laser point gesture. The lecture video was sliced on the client side to reduce the bandwidth needs. The ability to see the laser point was often lost due to compression. It was iconified on the client side by analyzing the videos to interpret the gestures made with a laser pointer. In addition, Dickson et al. developed an automated lecture recording system called PAOL (Presentation Automatically Organized from Lectures) [5]. One component of PAOL was designed to capture images of any significant content which was sent to a digital projector. The lecturer is not required to change his/her presentation style in any way other than wearing a microphone during a presentation.

In the second category, a fixed camera(s) is used for lecture recording. The lecturer and the whiteboard are the two main targets to be recorded. For example, Dickson et al. extended the PAOL to enable the extraction and analysis of whiteboard content automatically [6-7]. A sequence of images from a fixed camera was analyzed to extract a series of key frames by removing the lecturer's image and enhancing the images for greater legibility. These key frames are important in enabling student review of the material and for use in distance education. Nagai used a high-definition (HD) camcorder in the recording system to support large-scale automatic lecture recording [1]. If a general fixed camera is used for lecture recording with a wide whiteboard, it must be operated manually by student staff. However, a HD camcorder enables the region of interest in the video frame to be traced and NTSC resolution video is able to be generated during post-processing. The transcoding from HD video to MPEG-2 video remains the most time-consuming problem to be solved. Vasenev et al. utilized an interactive whiteboard with full HD 1080p resolution and an HD camera to support live broadcasting [2]. The LernLabor system is very flexible to in handling various video input formats and supports Web-based learning or knowledge sharing options. In addition, Lin et al. proposed a method for structuring blackboard lecture videos recorded by a fixed camera [8]. The method was used to estimate the learning focus that learners need to give greater attention. A learning-focused attention model was developed and constructed by analyzing the lecturer's behavior and the content on the blackboard. However, current camera motions were utilized by professional videographers to emphasize certain lecture content on a wide blackboard. It is an issue to be considered further. Friedland and Rojas proposed a method to handle the split attention effect resulting from use of two separate windows showing the lecturer's video and blackboard content [9]. The lecturer image was extracted from the video stream recorded by a fixed camera and then pasted onto the blackboard image.

In the third category, single or multiple PTZ cameras are used for lecture recording. A PTZ camera is even incorporated with the fixed camera for some purposes. When a PTZ camera is used, there are two main considerations: determining what the target is and what kind of information is used to control the PTZ camera. The related studies of this category are summarized in Table 1.

Table 1. Summary of related studies using PTZ camera(s) for lecture recording

Related studies	Camera(s)	Target of PTZ camera	Information used
Wallick et al. [10]	fixed+PTZ	lecturer	camera image
Rowe and Casalaina [11]	fixed+PTZ	Lecturer, projected slide	(determined by cameraman)
Cavallaro et al. [12]	PTZ	face of lecturer	camera image
Debevc and Kosec [13]	fixed+PTZ	lecturer	camera image
Lampi et al. [14]	dual PTZ	face of lecturer and student	camera image
Zhang et al. [15]	dual PTZ	lecturer, students	camera image for lecturer audio (microphone) for student
Wei et al. [16]	fixed+PTZ	lecturer, students	audio (microphone)
Chou et al. [17]	PTZ	face of lecturer, projected slide	camera image
Hayden et al. [18]	PT+Z	lecturer, blackboard, projected slide	(determined by the user)
Ranjan et al. [19]	fixed+PTZ	face of participants	audio (microphone)

By observing the table, the following remarks can be made:
(1) Most of the studies utilized fixed and PTZ cameras. Some studies utilized single or dual PTZ cameras. One study (Hayden et al. [18]) is denoted as "PT+Z" that means a pan/tilt mechanism and a camera with zooming capability were integrated to provide a function similar to using a PTZ camera.
(2) The targets of the PTZ camera are mainly the lecturer and the teaching materials, i.e. the contents on the blackboard or the projector screen. The student or participant may also be the target thereby making the recorded lecture more complete and to improve its overall interest.
(3) The camera image is the main information used to control the PTZ camera. The microphone is also used in some studies. Two studies even provide functions to control the camera manually.
(4) The utilization of the PTZ camera in lecture recording has increased in recent years. Recent camera price decreases and a complete SDK (Software Development Kit) with supporting API (Application Programming Interface) may be considered as reasons for the increased use of these cameras. PTZ cameras are expected to be increasingly used for lecture recording.

As described, the PTZ camera is becoming popular for lecture recording. Its operation can be automated as much as possible unless there are special requirements. However, consideration of a system to record traditional blackboard lectures which may be easily reviewed by students has not yet been addressed in previous studies. The legibility of handwritten data on the blackboard is the main concern in terms of student access to the handwritten information. Therefore, a system described in this study has been designed which uses a single PTZ camera to address this obstacle.

3. The proposed system

For the traditional lecture with a wide blackboard, a system, called the ACRS (After Class Reviewing System) was designed to automatically record the lecture using a PTZ camera. A block diagram for the ACRS is shown in Figure 2. The system has been designed to support the simultaneous recording of lectures in four classrooms. The number of classrooms supported is dependent on the computing power and network bandwidth available. The recording process used in every classroom is the same. Before the lecture recording, an operator sets up the parameters for controlling the PTZ camera of a classroom, such as the IP address of the camera, the class schedule, the left and right boundary of the blackboard, and the zoom-in value. The main modules of the ACRS are the "Lecturer Detection" and "Handwritten Data Detection". The detection results are used by the "Camera Control" module to operate the PTZ camera appropriately. The "Video Recording" module records the live video from the PTZ camera.
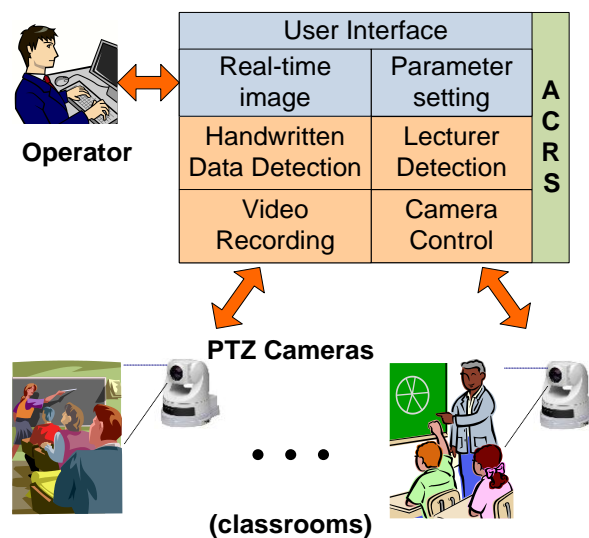

Figure 2. The system block diagram

For the lecture recording of a classroom, the process of the ACRS is divided into the wide-angle phase and the zoom-in phase as is shown in Figure 3. The upper-left part outlines the process of the wide-angle phase. In this phase, the system checks whether the time for the class is about to begin. If so, the system detects the presence of the lecturer using motion detection in a preset ROI (Region of Interest). When a moving object is found in the ROI, the system is switched to the zoom-in phase, as depicted on the right side of Figure 3. There are two sub-processes in this phase. The first consists of the Sobel edge detection, temporal differencing, and optimal thresholding, to locate the lecturer. This sub-process is complicated compared with that used in the wide-angle phase since the PTZ camera is controlled to pan left and right during the lecture. The second sub-process consists of the improved GMM (Gaussian Mixture Model) model, temporal differencing, and optimal thresholding, to locate the handwritten data on the blackboard. According to the detection results of the lecturer and the handwritten data, the camera is controlled according to the following rules:

(1) If the lecturer has not written any data on the blackboard, the camera is controlled to aim at the lecturer. The camera rotation is activated when the lecturer moves outside a preset

tolerance range (TR).

(2) When newly handwritten data is detected and if it is the first data or if the writing reaches the right boundary of the current FOV, the camera is controlled to rotate slowly to aim on the new data. If the new data does not reach the right boundary, it is appended into a tracking queue.

(3) Every piece of handwritten data is associated with an effective time (ET), e.g. 30 seconds. If the ET of the current handwritten data has expired, it is deleted from the tracking queue. The camera is controlled to then aim on the next handwritten data in the queue. If the ETs of all the handwritten data have expired, i.e. if the queue is empty, the camera is controlled to track the lecturer using Rule (1).

(4) If the camera aims on the handwritten data and the lecturer leaves the current FOV for a preset period, the system switches back to the wide-angle phase in order to detect and track the lecturer again.

In Rule (1), TR is used to adjust the frequency of the camera rotation. If TR is set to 0, the camera is controlled to focus on the lecturer all the time. If the lecturer is used to walking around on the platform, it can cause the recorded video to shake frequently and it becomes unsuitable for students to review. In Rule (3), the ET is designed to prevent the same piece of handwritten data being aimed at too long based on Rule (2). The above rules are depicted in the lower-right part of the process.
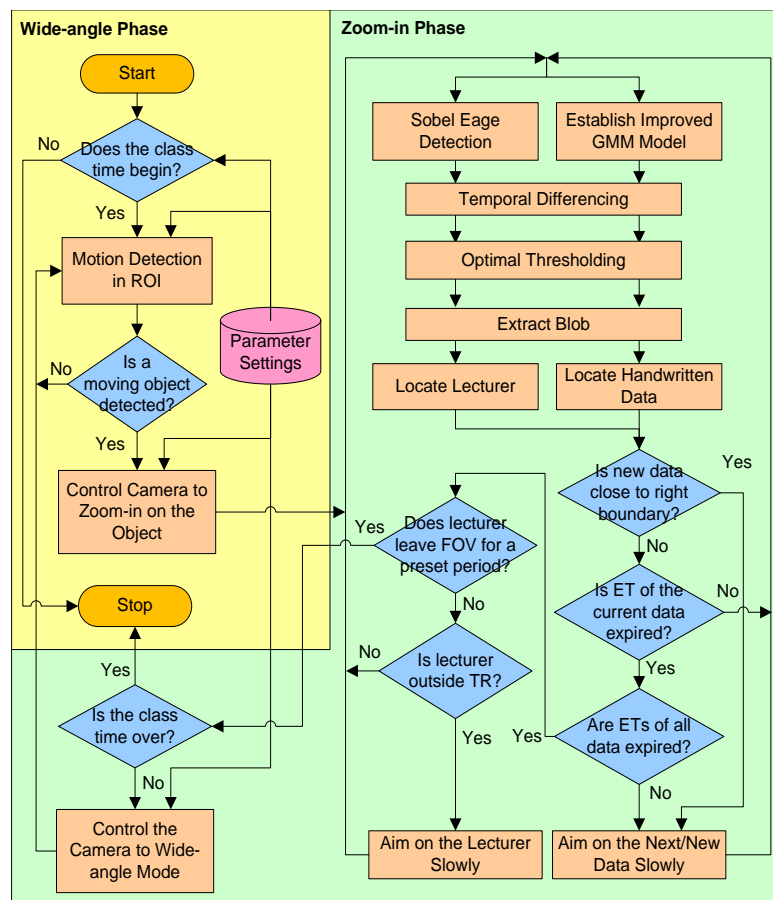

Figure 3. The process of the ACRS system

The above phases are presented in detail in the following subsections.

3.1. Wide-angle camera

In this phase, the detection of moving objects is performed in a preset ROI and at the preset class time in order to check whether the lecturer has entered the area. After the lecturer enters the ROI, the system enters the zoom-in phase. An example of the preset ROI is shown in Figure 2. marked by a red rectangle. The setting of the lower boundary of the ROI must prevent detection of the students as they are being seated.

There are two methods popularly used for detection of moving objects; background model subtraction and temporal differencing. The rotation and zoom capabilities make it difficult to establish the background model, so a temporal differencing method is used to detect the lecturer. The method computes the absolute difference of two successive frames for every pixel in the ROI. When a moving object is detected, the system controls the PTZ camera to zoom-in and aim on the lecturer. The system then enters the zoom-in phase. A phase switching example is shown in Figure 5. If a student is misdetected as he is being seated, the system will enter the zoom-in phase. However, the system will switch back to a wide-angle phase when no object has been detected within the zoom-in image during a preset period.
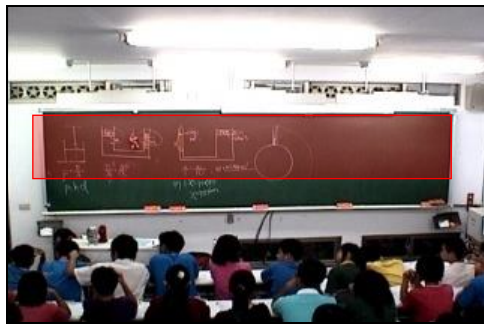

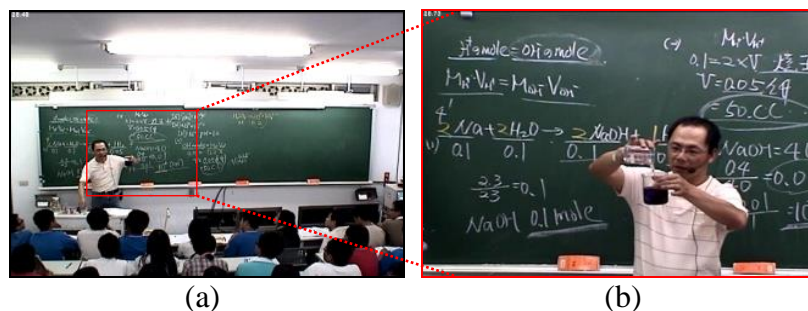**Figure 4.** The setting of ROI in the wide-angle phase

(a)   (b)

Figure 5. The scenario of phase switching: (a) wide-angle phase; (b) zoom-in phase

3.2. Zoom-in phase

The purpose of this phase is to record the lecture in high resolution for student review. Initially, it is intuitive to track the lecturer in this phase. However, the lecturer moves frequently and so do the recorded videos. Recognizing this, we quickly found that the recorded video is not suitable for subsequent review by students. Students experience difficulty in recognizing handwritten data on the blackboard, even though handwritten data is the key content needed by the students reviewing the lecture. Therefore, the process of this phase tracks the lecturer before he writes

data on the blackboard. After the handwritten data is detected, the PTZ camera is controlled to aim on the handwritten data. The camera is then controlled to pan right slowly when either the ET of the current data has expired, or when the new handwritten data reaches the right boundary of the current FOV. During the handwritten data tracking, the system keeps detecting the lecturer simultaneously. If the lecturer leaves the FOV completely for a preset period, e.g., ten seconds, it means he may continue his lecture at some location. In this case, the system will be switched back to the wide-angle phase to restart the process. The lecturer and handwritten data detection methods are presented as follows.

(1) Lecturer detection

During the lecturer tracking period, a tolerance range (TR) is defined to avoid frequent camera rotation. An example of the TR setting is shown in Figure 6. The range is indicated by two vertical dashed lines. If the TR is set to zero, the system will keep the lecturer in the center area of the FOV. It is similar to the lecture tracking method presented in some previous studies.
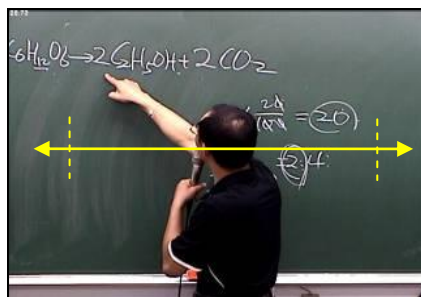

Figure 6. The setting of the tolerance range (TR)

Next, the motion of a lecturer is focused mainly on his or her hands, arms, or head. The temporal differencing method of two successive frames causes only the moving parts, i.e., hands, arms, or head to be detected. The whole lecturer is easily missing. For this reason, the Sobel edge detection method is performed before the temporal differencing. It is useful to preserve the contour of the lecturer. An optimal thresholding method proposed by Sonka et al. [20] is also performed to extract a clear contour. Then, the morphing operations, including erosion and dilation, are used to eliminate the small fragments. Finally, the largest "blob" (binary large object) is recognized as the lecturer. The above steps are demonstrated by the example shown in Figure 7.

If half the number of white pixels of the detected lecturer is outside the TR and no handwritten data is detected or tracked, the system rotates the camera to keep the lecturer in the center of the camera's FOV. If the lecturer cannot be detected after camera rotation, the system will be switched back to a wide-angle phase to detect the lecturer again.

(2) Handwritten data detection

The lecturer is assumed to write data in a left to right and top to bottom order. A new set of handwritten data is expected to appear on the upper part of the blackboard. Therefore, only the upper part is processed in order to detect the handwritten data and to shorten the processing time. In order to detect the area of the upper part, the first zoom-in image is captured and no handwritten data is expected on it as shown in Figure 8(a). The image is processed using the same optimal thresholding method with the result as shown in Figure 8(b). The upper boundary

of the blackboard is detected when the length of the continuous white pixels is larger than half the image width as indicated by the red dashed line. Then, the upper part of the blackboard is one third of the image height based on the upper boundary as shown in Figure 8(c).
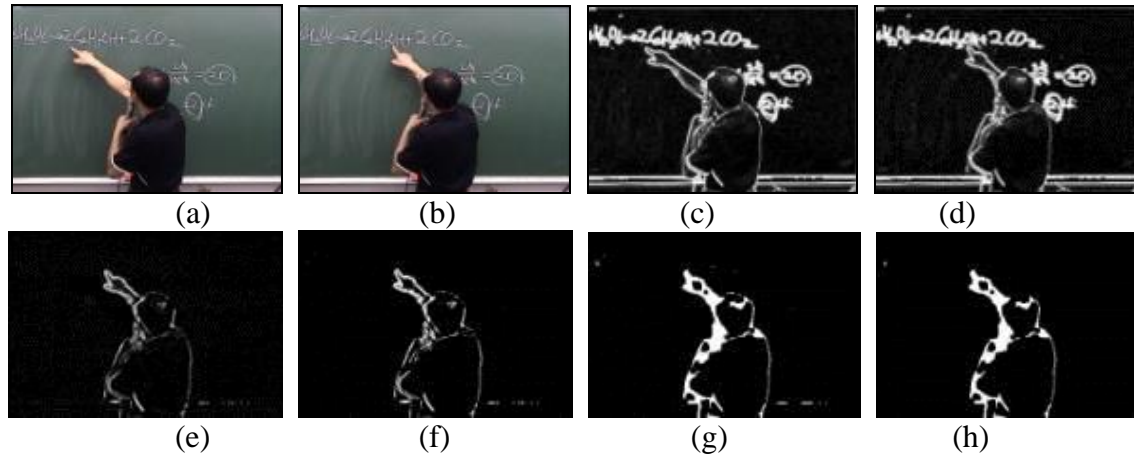

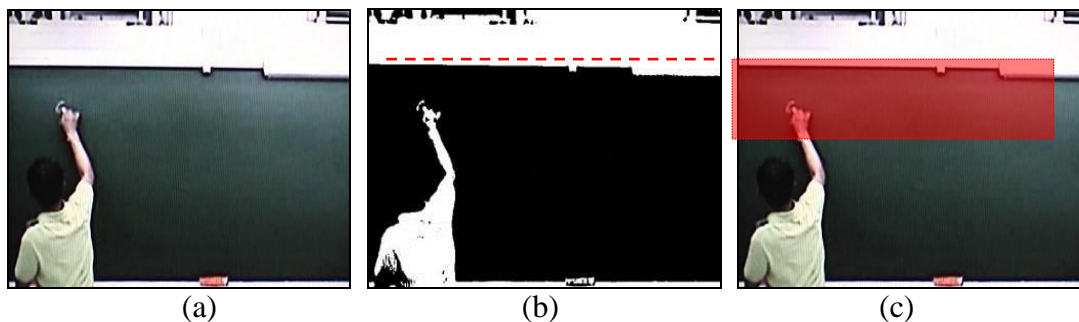Figure 7. An example of the lecturer detection in the zoom-in phase (a)~(b) two successive frames

Figure 8. An example of upper area detection: (a) zoom-in image; (b) optimal binarization and detected upper boundary; (c) the detected upper area

After the upper part of the blackboard is detected, a background model is constructed for the detection of handwritten data. It is constructed dynamically as the camera rotates. The popular Gaussian Mixture Model (GMM) was chosen and improved here for constructing the background model. The GMM model is constructed immediately after every camera rotation. However, the GMM is a computationally more time-consuming method since the model is constructed for every pixel. Although the model construction is restricted to the upper part of the blackboard, the construction time is shortened further for effective handwritten data detection. The steps of the improved GMM model are as follows:

(1) Construct the initial GMM model based on the frames of the first second following the camera rotation. Most of the pixels are included in the background model.
(2) For those pixels not included in the model, an interpolation method is used to generate their pixel values. Their value is the average of the left and right pixel values.

An example of the above steps is shown in Figure 9. It demonstrates that the interpolation method is useful since the blackboard image is not complicated.

(a)   (b)

Figure 9. GMM background model construction (a) initial model (b) the result of interpolation

After the background model is constructed, a simple background subtraction method and the same optimal binarization method mentioned previously are used to detect the new handwritten data. The positions of handwritten data are recorded in a tracking queue. When the range of newly handwritten data reaches the right boundary of the current FOV, the camera is controlled to let the newly handwritten data appear in the left side of camera's FOV. Therefore, any handwritten data will be recorded in the video for students to review. The above process of the handwritten data detection and tracking is shown in Figure 10. In Figure 10(a), the improved GMM model has been constructed and starts to detect handwritten data. When a handwritten data is detected, its range is marked by a pink line with arrows. The range is also expanded with the handwritten data as shown in Figure 10(b) and Figure 10(c). The camera aims on the current handwritten data even though the lecturer is outside the TR as shown in Figure 10(d). When a newly handwritten data is detected and reaches the right boundary of the current FOV as shown in Figure 10(e). The camera is controlled to rotate slowly to aim on the new data as shown in Figure 10(e).

In advance, every piece of handwritten data is associated with an effective time (ET). It is designed to prevent the system from focusing on the same handwritten data for too long. The system will ignore a piece of handwritten data after its ET has expired. After that, the system will focus on the next handwritten data in the tracking queue. If no data appears in the queue, the system will continue tracking the lecturer once again. If during the period of the tracking of the handwritten data the lecturer leaves the FOV for a preset period, the system will switch back to the wide-angle phase.

4. Experimental study

In order to evaluate the performance of the proposed ACRS system, a prototype system was implemented using Microsoft Visual C# 2008 and AForge.NET Framework. The camera was an AXIS-214 PTZ camera with an 18X optical zoom. In the beginning of the prototype development, the system was tested in the lecture room. It was found to be difficult to debug and to repeat the testing of the same error. In order to shorten the development time, several lecture videos were recorded in wide-angle mode as shown in Figure 11(a). The video was then played on a LCD monitor as shown in Figure 11(b). The PTZ camera was controlled to aim on the monitor to perform the process of ACRS enabling the video to be replayed to repeat testing of the same error. The same environment was also used in the experimental study.

A screen shot of the prototype system designed for recording four lectures simultaneously is shown in Figure 12. On the left is the real-time camera image of the selected lecture. The parameters listed on the right side include the IP address, the ROI of the wide-angle phase, the zoom value and boundary of the zoom-in phase, and so on.

In this experimental study, the lecture automatically recorded by ACRS was compared with that manually recorded by a cameraman. Only the pan value was adjusted in the zoom-in phase.

So, the series of pan values of the PTZ camera was recorded every three seconds during the lecture recording. The pan values of the camera controlled automatically and manually were recorded separately. The absolute difference of two pan values, determined at the same time and denoted as ΔP, was computed to realize how close the control of the ACRS was to that of the cameraman. Two lecture videos were also used for the testing. One was the lecture of a junior high school physics class and the other an engineering mathematics (EM) class at our university. The length of the testing videos was 30 minutes. The ET of the handwritten data was set to 15, 30, and 60 seconds separately. The ΔPs of the two testing videos are depicted in Figures 13(a) and 13(b), respectively. The ΔP must be as small as possible. By observing the figure, it can be seen that most of the ΔPs are less than 1.5 degrees. The ΔPs of the lecture recorded under the ET for 30 seconds is lower than that recorded under the ET for 15 or 60 seconds.

In order to evaluate the above results quantitatively, a threshold, called maximal ΔP, is used to compute the percentage of those ΔPs that are less or equal than this threshold. The maximal ΔP is set from 0.1 to 2.4 degrees. The percentages of two testing videos under different maximal ΔPs are listed in Table 2. The corresponding curves are depicted in Figure 14. By observing the results listed in Table 1, we found that the percentage of ΔPs with 30 seconds ET is better than that with 15 or 60 seconds. This is the same as the above observation shown in Figure 13.

In addition, the ΔPs with 1.2 and 1.8 degrees are close to 1/4 and 3/8 camera image, respectively. Some examples of such ΔPs are shown in Figure 15. By observing Figure 15(b) and (d), the images are still suitable for student review. Therefore, 1.8 degrees is deemed as a reasonable upper bound of the maximal ΔP in this experimental study. When the maximal ΔP equals 1.8 degrees and the ET equals 30 seconds, the percentages of the two testing videos are 83 %and 83.4% as listed in Table 2. It means that most of the time, the recorded video is suitable for student review.
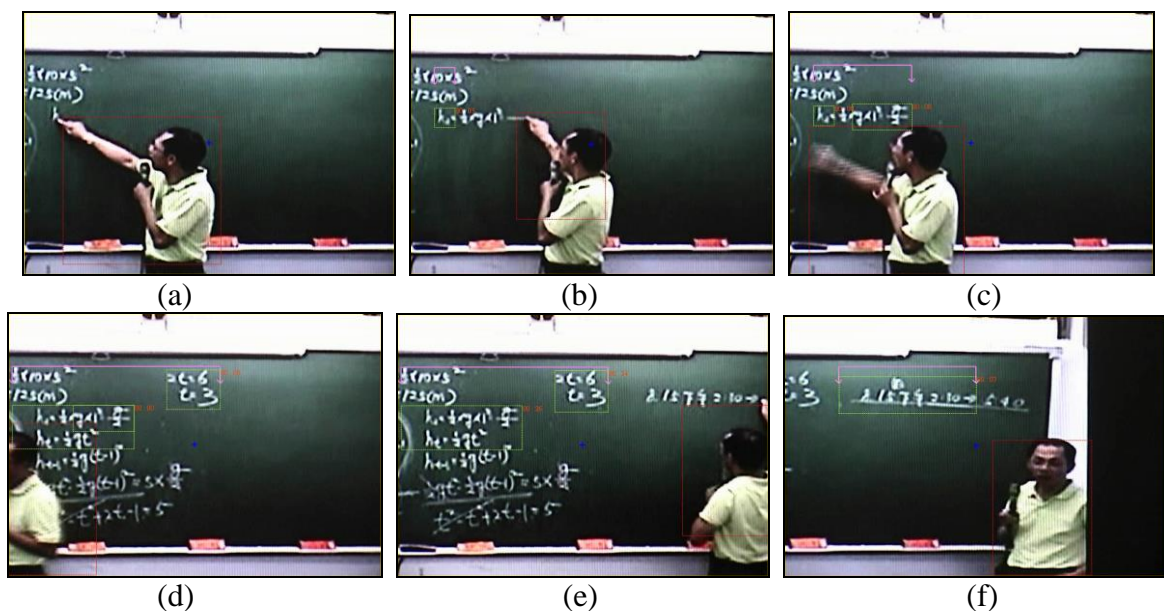

(a)	(b)	(c)
(d)	(e)	(f)

Figure 10. An example of the handwritten data tracking: (a) the lecturer starts to write data; (b) data is Detected; (c) the detected range is expanded; (d) data is tracking; (e) new data reaches the right boundary of FOV; (f) camera is controlled to aim on the new data

Table 2. The percentages of the maximal ΔPs under various ETs

Videos Max. ΔP (degree)	Video 1-Physics			Video 2-EM		
	ET			ET		
	15s	30s	60s	15s	30s	60s
0.1	16.2	6.8	6.8	20.9	8.2	0.0
0.2	19.7	12.7	31.0	21.1	32.2	6.4
0.3	45.6	38.8	63.9	27.1	53.6	7.2
0.4	52.2	64.3	64.3	27.5	57.1	12.5
0.5	55.8	64.7	64.3	31.4	58.1	18.3
0.6	65.7	68.6	64.5	45.4	71.4	33.7
0.7	65.9	68.8	64.7	46.0	71.5	34.1
0.8	65.9	68.8	64.7	46.8	75.6	35.3
0.9	72.5	69.6	65.5	62.6	76.6	55.8
1.0	79.0	69.6	73.3	62.8	77.8	60.6
1.1	79.0	69.6	73.5	65.5	77.8	63.4
1.2 (1/4 image)	79.3	78.8	73.5	67.6	81.1	73.3
1.3	79.5	79.1	73.7	68.0	81.1	73.9
1.4	79.5	79.1	73.7	80.5	81.3	73.9
1.5	85.0	79.5	73.9	81.1	81.5	76.4
1.6	85.0	81.7	73.9	81.1	82.9	76.4
1.7	85.0	81.7	73.9	82.1	83.0	76.8
1.8 (3/8 image)	**85.4**	**83.0**	**73.9**	**82.3**	**83.4**	**77.4**
1.9	85.4	88.1	73.9	82.3	83.4	85.2
2.0	85.4	88.1	76.2	84.8	83.8	85.2
2.1	85.8	88.1	92.2	85.4	86.4	85.4
2.2	85.8	97.1	92.2	85.4	86.4	85.4
2.3	86.7	97.1	92.6	85.6	86.6	85.6
2.4	88.1	97.5	92.8	86.2	86.9	85.6

(unit: percentage)
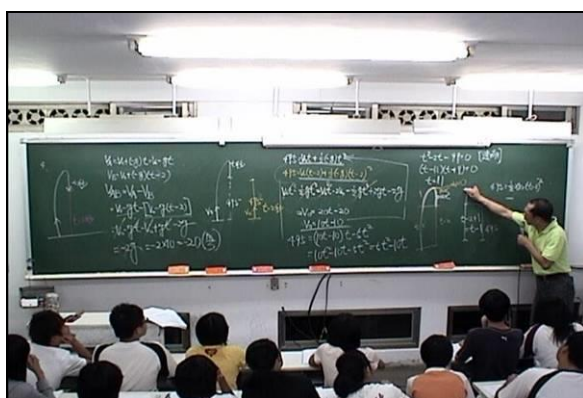
(a)   (b)

Figure 11. The experimental design: (a) the lecture video for testing; (b) the setup of the camera and the LCD monitor
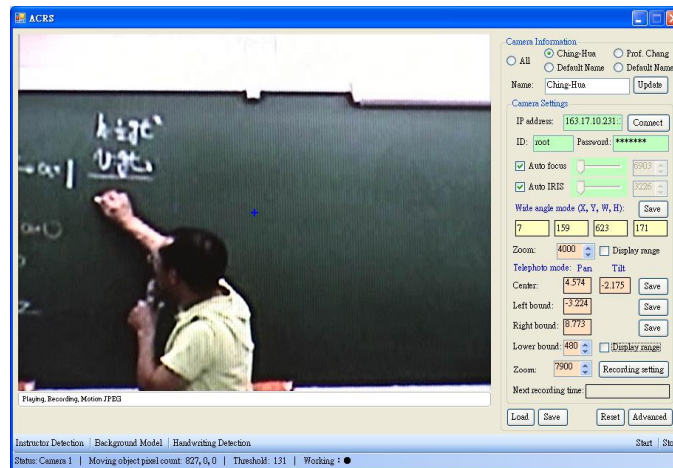
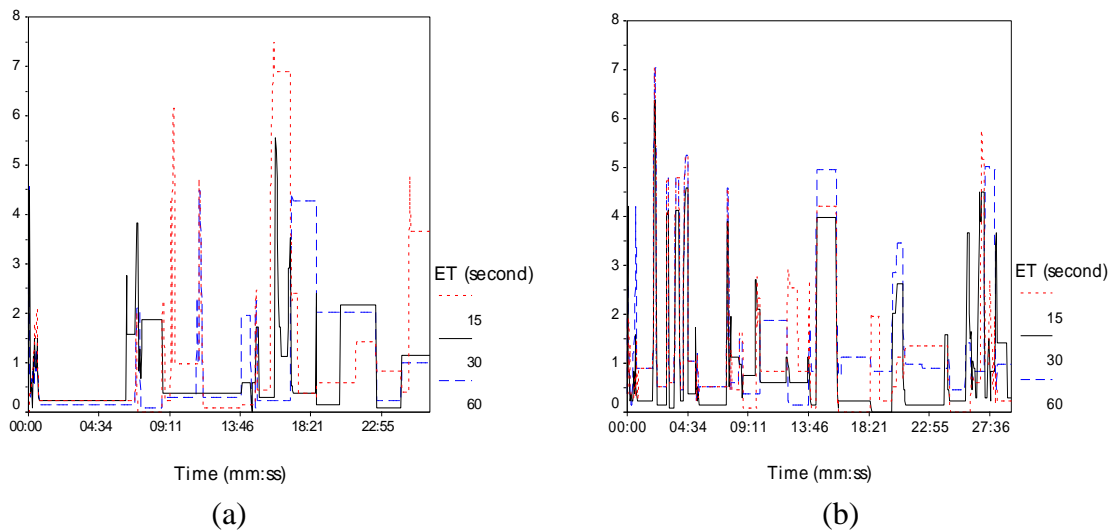**Figure 12.** The screen shot of the prototype

(a)   (b)

Figure 13. The ΔP of the videos 1 and 2 under various ETs: (a) video 1; (b) video 2

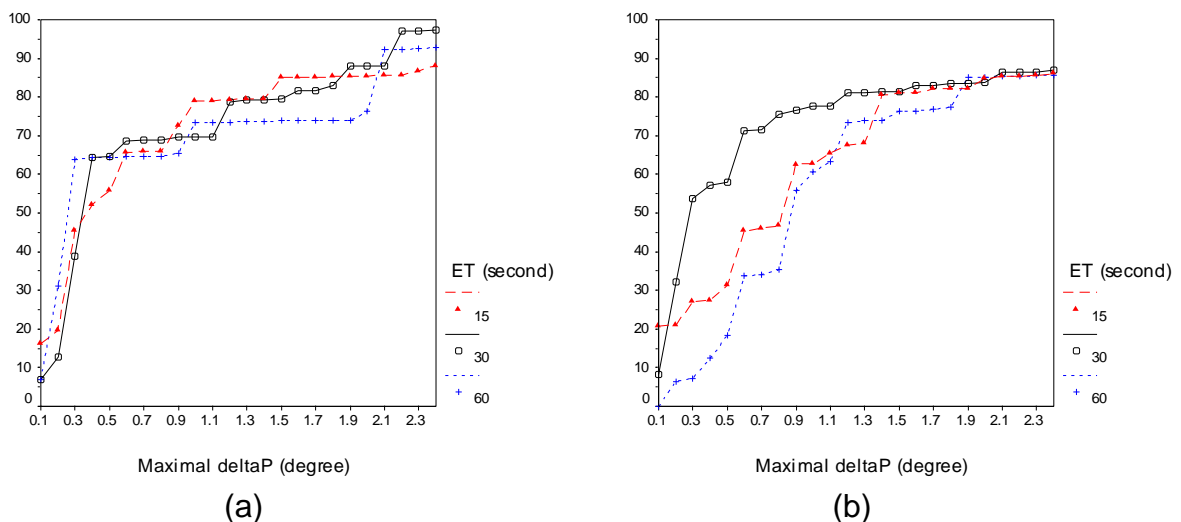(a)   (b)

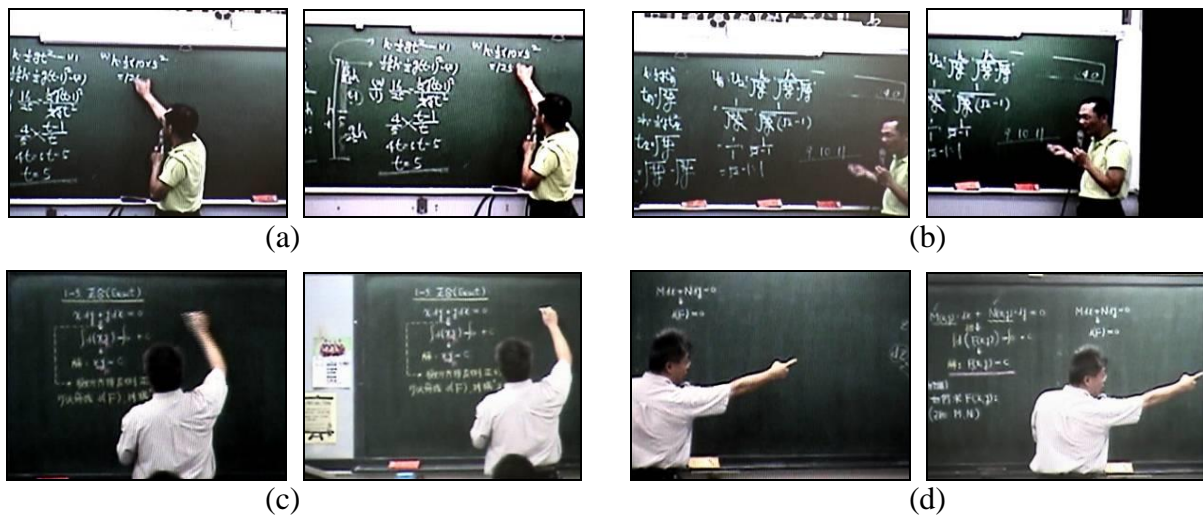**Figure 14.** The ΔP percentage of the videos 1 and 2 vs. maximal ΔP: (a) video 1; (b) video 2

Figure 15. Some examples of reasonable ΔPs (the results of manual control and ACRS are shown in the left and right photos, respectively): (a) 1.2 degrees of video 1; (b) 1.8 degree of video 1; (c) 1.2 degrees of video 2; (d) 1.8 degree of video 2

In addition to the above examples with a reasonable ΔP, several examples of ΔPs reaching 2.1 or 4.8 degrees are also presented and discussed here. Two examples of ΔP reaching 2.1 degrees are shown in Figure 16. Figure 16(a) and 16(b) are the examples for videos 1 and 2, respectively. In this instance, the lecturer had left the camera's FOV and the ET of the handwritten data had also expired. ACRS switched back to the wide-angle phase and relocated the lecturer again. Sometimes the recorded video is still suitable for student review as shown in Figure 16(a).
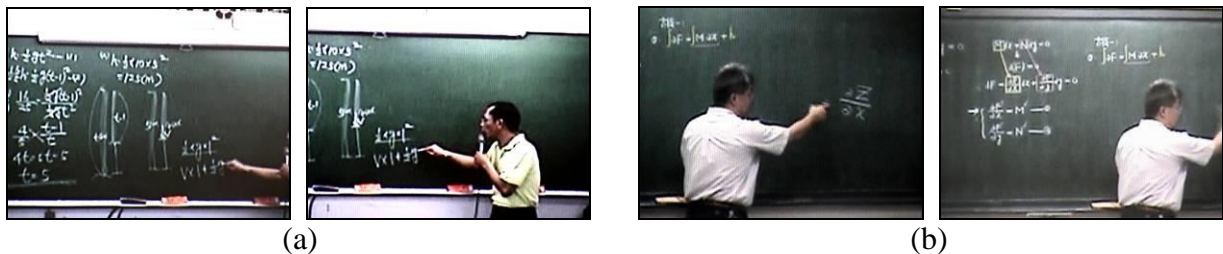

Figure 16. Two examples of the ΔP reaching 2.1 degrees (the results of manual control and ACRS are shown on the left and right sides, respectively): (a) video 1; (b) video 2

One example of ΔP reaching 4.8 degrees is shown in Figure 17. Figure 17(a) and 17(b) are the frames captured from the video recorded manually before and after a special case. Figs. 17(c) and 17(d) are frames captured from the video recorded by ACRS. The special case involved a lecturer who didn't follow the left-to-right order when writing data on the blackboard. Instead, the lecturer wrote data in the middle of the blackboard. Such a case easily occurs when the lecturer reaches the right boundary of the blackboard. It causes the recorded video to be unsuitable for student review.
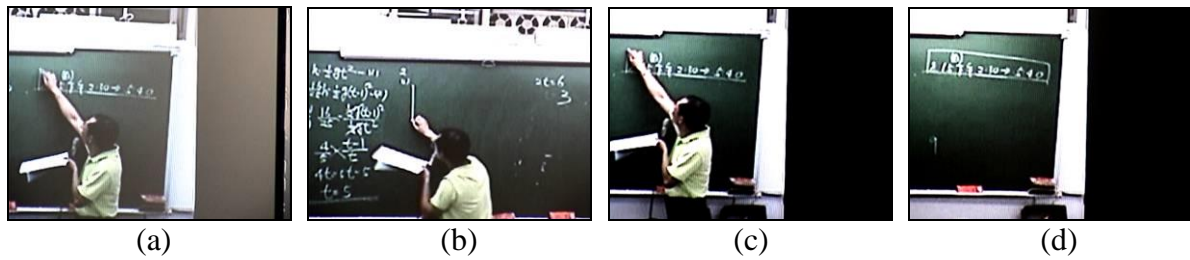
(a) (b) (c) (d)

Figure 17. An example of ΔP reaching 4.8 degrees: (a) and (b) manual control; (c) and (d) ACRS; (a) and (c) before; (b) and (d) after

Another special example of ΔP reaching 6.3 degrees is shown in Figure 18. Similar to the previous example, Figure 18.(a) and 18(b) are frames of manual control before and after the special case. Figs. 18(c) and 18(d) are frames of ACRS. The special case here involves a lecturer moving to the right and outside of the FOV to write data in the middle of the blackboard for a short time. Since the ET of the current handwritten data had not yet expired, the camera was still aimed at the same data as shown in Figure 18(d). However, it could be avoided if the lecturer followed the left-to-right order when writing data on the blackboard.
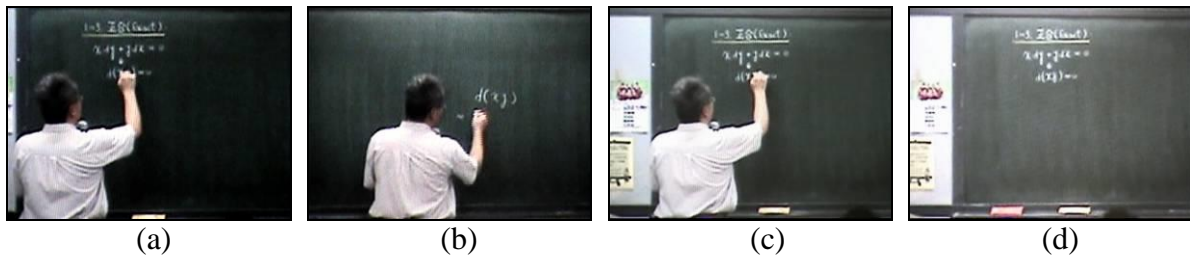

(a) (b) (c) (d)

Figure 18. An example of ΔP reaching 6.3 degrees: (a) and (b) manual control; (c) and (d) ACRS; (a) and (c) before; (b) and (d) after

According to the above experimental results, the prototype of ACRS can achieve about 83% of the video time being suitable for review by students compared with video recorded by a cameraman. Although a large ΔP exists in some special cases, the system can still control the camera back to the needed FOV after a short time.

4.1. Discussion

Although the prototype of ACRS seems to offer promising results, some issues remain:
(1) For some examples with large ΔP, as shown in the figures, these are caused mainly by the behavior of the lecturer. However, we believe that the lecturer might possibly cooperate more with the system in order to achieve a much better performance. If the lecturer can follow the left-to-right order and guide the PTZ camera to record the required video, the percentage will be improved to achieve performance similar to the video recorded manually by a cameraman.
(2) In general, the performance of such a system may be easily influenced by the ambient brightness when the camera detects the lecturer and the handwritten data. However, methods used in ACRS, such as the Sobel edge detection method, the GMM model, or the optimal thresholding method, can overcome the influence of the environmental brightness and provide performance similar to that achieved by an operator manually operating a

camera.

(3) In the experimental study, the tracking of the lecturer and handwritten data were considered together for the traditional blackboard lecture. In fact, ACRS can be adapted to track the lecturer by adjusting only two parameters, TR and ET, to zero. When TR equals zero, the PTZ camera will be controlled in such a way as to keep the lecturer in the center of the FOV. A zero ET means that the camera will not be controlled to focus on any detected handwritten data. The applicability of ACRS is wider than the current demonstration perhaps indicates.

(4) Note that the handwritten data shown in the above examples seems blurry. This results from the fact that the video was recorded from the LCD monitor in the experimental environment. A snapshot of the video recorded from the real classroom with a resolution is 640×480 pixels is shown in Figure 19. The handwritten data is obviously clear and suitable for students to review.
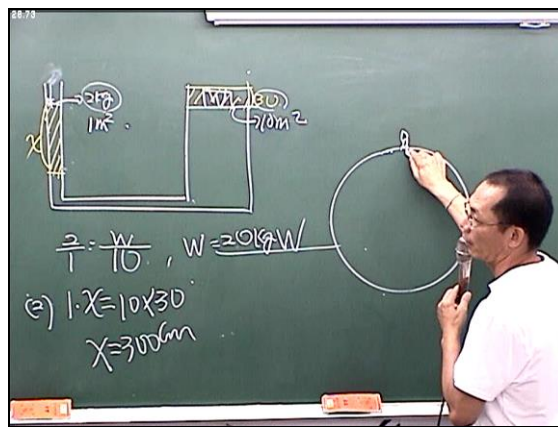

Figure 19. A snapshot of the recorded lecture video in the real classroom

5. Conclusion

Traditional lecture with a wide blackboard is still popular in various school settings from elementary school to university. The main purpose of this study was to design a system for lecture recording using a single PTZ camera. The recorded video must be suitable for students to review. It is an important issue, but one seldom addressed by previous studies. The design of the ACRS system focuses on such issues and enables the system to operate automatically. The detection of both the lecturer and the handwritten data is performed simultaneously. The tracking of the handwritten data is considered to be more important than a focus on the lecturer. The lecturer is tracked by the camera when there is no handwritten data in the tracking queue or when the lecturer leaves the FOV for a preset period. This provides the main advantage of ACRS as compared to the results of previous studies but of course, the system can perform lecturer tracking only by adjusting the TR and ET parameters. The experimental results show that the video recorded by ACRS is close to the quality of that recorded by a cameraman.

When a cameraman operates the PTZ camera for lecture recording, his eyes are similar to a fixed camera. Therefore, he or she knows whether the PTZ camera is being controlled to aim on the lecturer when the lecturer leaves the FOV of the camera. Many cases with a large ΔP occur during such situations. Therefore, a fixed camera will be integrated into the ACRS in the future. The role of the fixed camera is similar to that of the human eyes of the cameraman. It can be

used to monitor the whole scene; i.e., the blackboard as a whole. The detection of the lecturer and the handwritten data will be included in the image of the fixed camera. The PTZ camera will then be able to be controlled more precisely to record lecture video superior to that recorded with the current implementation of ACRS.

References

[1] Nagai, T. 2009. Automated Lecture Recording System with AVCHD Camcorder and Microserver. Presented in *Special Interest Group on University and College Computing Services (SIGUCCS) 2009 Fall Conference (SIGUCCS'09)*, ACM, St. Louis, Missouri, USA, 2009, 47-54.

[2] Vasenev, A., Zylka, J., Ecker, M., and Müller, W. 2011. The LernLabor System. Presented in *2011 IEEE Global Engineering Education Conference (EDUCON'11)*, IEEE, Amman, Jordan, 2011, 953-959.

[3] Zupancic, B. and Horz, H. 2002. Lecture Recording and its Use in a Traditional University Course. Presented in *the Seventh Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE'02)*, ACM, Aarhas, Denmark, 2002, 24-28.

[4] Winslow, A., Tung, Q., Fan Q., Torkkola, J., Swaminathan, R., Barnard, K., Amir, A., Efrat, A., and Gniady, C. 2009. Studying on the Move − Enriched Presentation Video for Mobile Devices. Presented in *the 2nd IEEE Workshop on Mobile Video Delivery (MoViD'09)*, IEEE, Rio de Janeiro, Brazil, 2009, 6 pages.

[5] Dickson, P. E., Arbour, D. T., and Gentzel, A. 2010. Evaluation of Automatic Classroom Capture for Computer Science Education. Presented in *the 15th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE'10)*, ACM, Ankara, Turkey, 2010, 88-92.

[6] Dickson, P. E., Adrion, W. R., and Hanson, A. 2006. Automatic Capture of Significant Points in a Computer Based Presentation. Presented in *the Eighth IEEE International Symposium on Multimedia (ISM'06)*, IEEE, San Diego, CA, USA, 2006, 6 pages.

[7] Dickson, P. E., Adrion, W. R., and Hanson, A. R. Whiteboard Content Extraction and Analysis for the Classroom Environment. Presented in *the Tenth IEEE International Symposium on Multimedia (ISM'08)*, IEEE, Berkeley, CA, USA, 2008, 702-707.

[8] Lin, Y. T., Tsai, H. Y., Chang, C. H., and Lee, G. C. 2010. Learning-focused Structuring for Blackboard Lecture Videos. Presented in *the 2010 IEEE Fourth International Conference on Semantic Computing (ICSC'10)*, IEEE, Berkeley, CA, USA, 2010, 149-155.

[9] Friedland, G. and Rojas, R. 2006. Human-Centered Webcasting of Interactive-Whiteboard Lectures. Presented in *the Eighth IEEE International Symposium on Multimedia (ISM'06)*, IEEE, San Diego, CA, USA, 2006, 6 pages.

[10] Wallick, M. N., Rui, Y., and He, L. 2004. A Portable Solution for Automatic Lecture Room Camera Management. Presented in *the 2004 IEEE International Conference on Multimedia and Expo (ICME'04)*, IEEE, Taipei, Taiwan, 2004, 4 pages.

[11] Rowe, L. A. and Casalaina, V. 2006. Capturing Conference Presentations. *IEEE Multimedia*, 13, 4: 76-84.

[12] Cavallaro, A., Chandrasekera, R., and Taj, M. 2007. Hands-on Experience in Image Processing: the Automated Lecture Cameraman. Presented in *the 2007 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07)*, IEEE, Honolulu, Hawaii, USA, III-721-724.

[13] Debevc, M. and Kosec, P. 2007. Equal Study: Lecture based Automated Video Recording System for All. Presented in *the Interactive Computer Aided Learning Conference (ICL'07)*, Villach, Austria, 2007, 8 pages.

[14] Lampi, F., Kopf, S., Benz, M., and Effelsberg, W. 2006. A Virtual Camera Team for Lecture Recording. *IEEE Multimedia*, 13, 4: 58-62.

[15] Zhang, C., Rui, Y., Crawford, J., and He, L. W. 2008. An Automated End-to-End Lecture Capture and Broadcasting System. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 4, 1: 23 pages.

[16] Wei, J. H., Wang, L. P., Li, M. Y., and Wang, Y. W. 2009. Advantages and Deficiencies of the Automated Multimedia Lecture Recording System in Lecture Video Production. Presented in *the 2009 International Forum on Computer Science-Technology and Applications (IFCSTA'09)*, IEEE, Chongqing, China, 2009, 271-273.

[17] Chou, H. P., Wang, J. M., Fuh, C. S., Lin, S. C., and Chen, S. W. 2010. Automated Lecture Recording System. Presented in *2010 International Conference on System Science and Engineering (ICSSE'10)*, IEEE, Taipei, Taiwan, 2010, 167-172.

[18] Hayden, D. S., Zhou, L., Astrauskas, M. J., and Black, J. A. 2010. Note-taker 2.0: the Next Step Toward Enabling Students Who Are Legally Blind to Take Notes in Class. Presented in *the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'10)*, ACM, Orlando, Florida, USA, 2010, 8 pages.

[19] Ranjan, A., Birnholtz, J., Henrikson, R., Balakrishnan, R., and Lee, D. 2010. Automatic Camera Control Using Unobtrusive Vision and Audio Tracking. Presented in *Graphics Interface (GI'10)*, ACM, Ottawa, Ontario, Canada, May 31th to June 2nd, 2010, 8 pages.

[20] Sonka, M., Hlavac, V., and Boyle R. 2007. "*Image Processing, Analysis, and Machine Vision*". third Ed., Thomson Learning, 180-181.