# CLASSROOM MULTIMEDIA INTEGRATION FOR ADVANCED E-PRESENTATIONS

*Weihong Li[1*], Hao Tang[2*], Chad McKittrick[2+] and Zhigang Zhu[1,2*]*
[1]Department of Computer Science, Graduate Center, The City University of New York, NY 10016
[2]Department of Computer Science, City College, The City University of New York, NY 10031
**{wli, tang, zhu}@cs.ccny.cuny.edu, +chad.mckittrick@csfb.com*

## ABSTRACT

*We have developed several basic components for our Virtualized Classroom project: automated data collection, intelligent media integration, and flexible user interfaces. As an example of media integration, we show that when the PPT slides are projected onto the digital whiteboard where the instructor annotates, modifies, or expands the PPT presentation, geometrically registering high-resolution images from the two sources yields high quality digital presentations (e-presentations). We use a low-cost digital camera as a bridge to align the two sources and present a hybrid registration approach to align the PPT and handwriting images. Experimental results are presented to validate our approach.*

## 1. INTRODUCTION

Multimedia materials from classrooms are rich sources of information. Today's first-generation e-learning systems primarily adopt a “record-and-playback” approach, which does not leverage the processing capabilities that we believe will underlie the next generation of more automated, flexible, and interactive e-learning systems. Using a portable presentation system with very cost-effective multimedia sensors, we have developed the following basic components for our Virtualized Classroom project-automated data collection, intelligent media processing and integration algorithms, and user-customized e-presentation interface designs. In this paper, we will mainly focus on alignment and integration of images with PPT printing notes and images with whiteboard handwriting contents. The result is an ability to dynamically allow the instructor to add handwritten material generated in a different medium (i.e., the whiteboard) onto the projected slides. This is done in high resolution using low-resolution video as the alignment automatically. This approach has further potential towards content-based media integration with handwriting recognition and content-based video representation.

We have two notes here to further motivate our work. First, although it might be true that a presenter seldom adds things on a well-prepared, academic-oriented PPT presentation, adding notes and drawings on a PPT lecture in a classroom scenario enriches interactions and discussions between the teacher and students. Second, an alternative approach for implementing handwriting notes and drawings is to enable annotation on computer screen directly, particularly when a Tablet PC is used for presentation. However, our experience is that some instructors, particularly with many writing needed, and/or for a discussion style class, prefer to give lectures in front of a large physical board (for easy reading/writing and natural interaction) rather than in front of a small screen (with small fonts and the need to change presentation nodes).

The organization of the paper is as follows. Some related work is discussed in Section 2. In Section 3, we give a brief description of our classroom sensors and the automated data collection and synchronization mechanism. Section 4 presents our hybrid media integration approach. Section 5 introduces our user-customized e-presentation interface. Finally, we conclude our work in Section 6.

## 2. RELATED WORK

Projects focused on record and playback technologies include: the Georgia Tech Classroom 2000 (eClass) [1], the CMU Just-In-Time system [4], and the UMass RIPPLES/MANIC [9], among others. Automated production is a major issue for these e-lecture systems, because to improve the quality of presentations many systems require significant manual effort in analog data collection, digitization, and synchronization. We note that some systems have introduced forms of automatic production. Cornell’s lecture browser [6] includes lecturer tracking, slide change detection and segmentation, and matching slide projections with digital slides. Auto-Auditorium [2] uses multiple cameras that automatically switch based on context. However we are not aware any work trying to automatically register PPT presentations with handwriting contents.

Another class of related work includes automated keystone correction [8,10] and projector array alignment [3] using video cameras and computer vision techniques. We share the similar ideas of using a video camera and the homography of the planar projection. However, our task is more challenging: we need to dynamically register three different types of images (PPT slide, handwriting, and video images) on-line.

## 3. MEDIA CAPTURE AND SYNCHRONIZATION

The commonly used classroom presentation tools are PowerPoint (PPT) slides, overhead projections, and blackboards /whiteboards. In order to digitize the classroom handwriting contents to create e-lectures, we use a low-cost Mimio Virtual Ink digital whiteboard system to substitute the use of the blackboard. Therefore, the "sensors" we are using in our portable presentation system for a classroom setting are a PPT slide capturer, a Mimio Virtual Ink handwriting system, and a digital video camera and a wireless microphone (to capture classroom video and audio). All the sensors are controlled by a laptop or a Tablet PC used for the presentation and can be easily managed by the instructor/lecturer using the system [13].

The PowerPoint *slide capturer* (top-right in Fig. 1) was modified from the Berkeley PPT Recording Add-In, provided by [7]. We added the start date and time information of the presentation being recorded in order to synchronize the recorded PPT slides with the accompanying whiteboard pages. The slide capturer also controls the synchronous audio recording from the wireless microphone. In addition to the pages of slides in one of the image formats (e.g., JPEG), a PPT log file is automatically generated with the timing information and the titles of all the slides in the presentation.

The Mimio Virtual Ink ultrasonic position capture system (top-left in Fig. 1) is capable of recording handwriting and drawing contents on a normal whiteboard of 2.4 m x 1.2 m (8 ft x 4 ft), with a 100-dpi resolution. The capture bar is a two-foot ultrasonic tracking array positioned along the upper-left edge of the whiteboard. The electronic marker sleeves transmit an ultrasonic signal to the capture bar, which triangulates the pen's position on the board as a instructor writes. The whiteboard presentation is saved in a series of html files, one html file (with timing information) for each whiteboard page saved as a JPEG image.

For the best use of the above sensors, we assume that the instructor will use a computer projector to project PPT slides onto a whiteboard. The handwriting contents written on the whiteboard could be on the top of the slide projections. A video camera (top middle in Fig. 1) is used to automatically collect video of the classroom activities, mainly the projector, whiteboard, and the instructor. After the instructor sets up the sensors and starts the presentation, everything is automatically saved for him/her. The synchronization of the PPT slides, whiteboard pages, and audio and video streams is enabled by a simple stream synchronization algorithm which uses the timing information in the PPT log file, the whiteboard log files, and the video/audio streams [13].

## 4. HYBRID MEDIA INTEGRATION APPROACH

When the PPT slides are projected onto the whiteboard, we need to geometrically register the images from the two sources to create high-quality e-presentations. If all the devices could remain stationary during the lecture, we could require the instructor to mark at least the four corners of the PPT projection area in the whiteboard provided in the Mimio "calibration" step [13]. However, this simple approach has several drawbacks: the occlusion of the instructor to the projections usually causes difficulty to accurately mark the projector's four corners; sometimes, the corners are too high to be reached; in other occasions, the instructor may forget to mark the four corners. In addition, it is hard to keep all the devices stationary with a portable representation; if either the projector or the whiteboard capture bar moves, the instructor has to mark the four points again.

We propose a completely automatic and "mind-free" approach. Since the two sources (PPT slides and handwriting pages) do not share the same contents, we use a low-cost digital camera (which is also used to capture the classroom video) as a bridge to align the two source images. Here is the overview of our approach. First, the corresponding PPT slides, handwriting pages, and video frames are matched up via the synchronization mechanism in the following manner (Fig. 1). When the instructor changes slides, a new slide image S is saved by the PPT slide capturer and a slide-only video frame $V_1$ is extracted (after a delay of few frames to ensure a stable frame). When a new handwriting page H is saved by the Mimio Virtual Ink system, the timing information will locate this page with the right slide S, and the video frame $V_2$ with both slide projection and handwriting contents will be saved automatically. Second, the PPT slide S and the handwriting page H are registered by way of the video frames $V_1$ and $V_2$. The slide-video registration is implemented by finding the boundary of the projection area in the video frame $V_1$. The video-whiteboard registration is fulfilled by matching the handwriting contents saved in the Mimio page H and the video frames ($V_1$ and $V_2$).

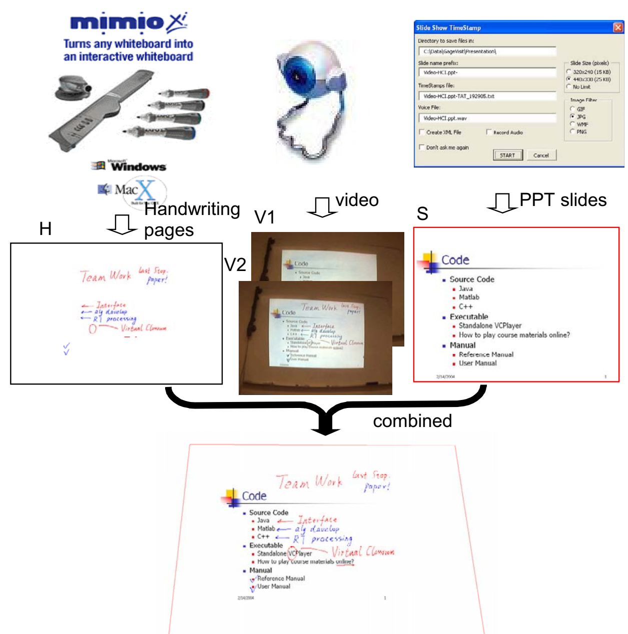



**Fig. 1. Multi-sensor data collection, synchronization, and integration in a classroom setting.**

In the following two sub-sections, we discuss the two main parts of our registration approach, namely slide-video registration and whiteboard-video registration, for the purpose of slide-whiteboard integration.

### 4.1. Slide and Video Registration

First we assume that the projection area is always in the camera's field of view (Fig. 1). After we remove the radial distortion of the digital camera using the method described in [12], we register the video frames with PPT slides using *a priori* knowledge of the projection area, e.g., the projection boundary. Since the projection area is usually significantly brighter than other areas, we first generate a binary image from the rectified video frame. Then, we use the Laplacian-of-Gaussian (LoG) operator to obtain the edges (zero-crossings). Finally, we use the Hough Transform to extract four boundary lines and calculate the coordinates of the four corners of the projection area (Fig. 2a). A 3x3 projective transformation matrix $\mathbf{A}_1$ is calculated between the quadrilateral formed by the four boundary lines and the rectangular frame boundary of the PPT slide image. The projective mapping from the video frame point $\mathbf{x}_v$ to the slide image point $\mathbf{x}_s$ is represented by $\mathbf{x}_s \cong \mathbf{A}_1 \mathbf{x}_v$ . Fig. 2b shows the video and PPT alignment result that uses the projective mapping matrix $\mathbf{A}_1$. The video frame is transformed to the PPT slide coordinates so that the orthogonal view of the PPT slide remains. The perspective distortion of the original video frame is obvious by looking at the shape of transformed video frame.

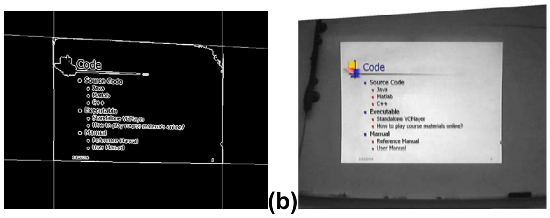
(a) (b)



**Fig. 2. (a) Boundary detection and fitting (b) Video-slide matching**

Sometimes, very cluttered notes in a slide will affect the boundary extraction results using the Hough Transform. Therefore, we scan the edge map of the video frame from its four borders in order to only keep the boundary edge pixels. Since we use the Hough Transform for boundary extraction, we do not need to have all the edge points of the complete projection boundary.

### 4.2. Whiteboard and Video Registration

The matching of the whiteboard handwriting page and the video frame is more challenging since the camera cannot "see" the invisible frame boundary of the whiteboard page. Therefore, we propose to match the handwriting contents from the Mimio Virtual Ink pages and the video frames. As a pre-processing step, we obtain the handwriting contents by subtracting the video frame with PPT projection only from video frame with both PPT projection and Mimio handwriting. Since the difference image shows obvious differences in places with handwriting contents, a simple thresholding process can reveal most of the handwriting contents.

We have developed a two-stage content matching approach [5] for registering handwriting contents from the two different sources. In the initial matching stage, we try to find some robust features to translate and scale the images in order to roughly align the two. In order to better match the two handwriting contents from the two sources, with different thickness of the strokes, we first run a thinning algorithm as proposed in [11] on both video and Mimio images to get skeletons of the handwriting. Then, the centroids of both thinned images are calculated. The thinning process is useful to obtain the centroid with the least influence by the unbalanced stroke thickness. The difference in the locations of the two centroids, C and c, in the video and the Mimio images, respectively, gives the translation between two images: $t = C - c$. To determine scale factor between these two images, we generate a centroid-centered polar profile for each image by measuring the distances of all the outmost stroke points in all directions from each centroid. After image translation and scaling, the two images are roughly aligned (Fig. 3), which gives a good starting point for the refinement step.

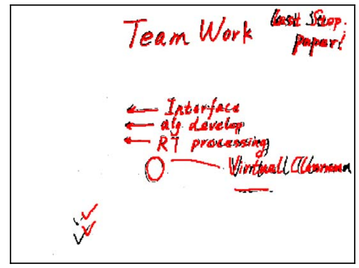



**Fig. 3. Initial matching result image (Video + Mimio). Translation t = (63, 19); scale s =1.938**

In the refinement stage, we use the connected components in the Mimio page as content-based matching primitives to find the best matches in the video handwriting image. Each connected component could be a single letter, part of a letter, a continuous figure, or even include several letters. We select those connected components of the Mimio page, whose sizes are sufficiently large for robust matches. Then, we use the pixel pattern under the bounding box of each selected primitive as the matching template to search for the best matched rectangular region in the video handwriting image using the normalized cross-correlation measures. Then, we choose those matches whose maximum normalized correlation values are above a threshold (e.g., 0.3) to calculate the final projective transformation matrix $\mathbf{A}_2$, which is used to align the Mimio page with the rectified video frame that has been aligned with the PPT digital slide image.

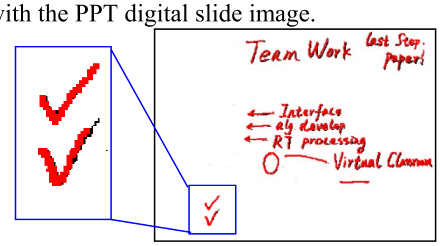



**Fig. 4. Fine registration result (Mimio contents in red and video contents in black)**

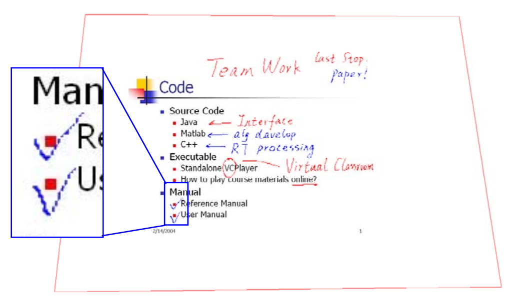



**Fig. 5. Mimio and PPT registered image**

Fig. 4 shows the fine alignment of the handwriting contents from video and Mimio, respectively, using the projective transformation matrix. Fig. 5 shows the result of registering the Mimio handwriting page to the corresponding PPT slide image. It can be seen that the two images are accurately aligned. More results can be found in [5] where handwriting contents are outside the projection area, and/or there are obvious keystone distortions in the slide projection.

## 5. USER-CUSTOMIZED PRESENTATION

Our current implementation of the Virtualized Classroom project includes a Virtualized Classroom Presentation System (VCPS) [13] which is designed as both an authoring tool and a presentation interface for different kinds of lectures, and has a user-selectable interface. The VCPS (developed in Java) includes two parts: the VCPS Creator and Player.

The user (an instructor or a student) who uses the system can customize the presentation by using the VCPS Creator to include different media (video, audio, slide, handwriting pages and/or integrated pages) in windows with user-selected sizes and positions. After the user opens a new Creator page, there will be a floating frame that has checkboxes on it. He/She will use these checkboxes to add or remove presentation components. Each time the user clicks on one of the checkboxes, a popup window will appear directing him/her to click on a certain file so that the program can load the proper information. In the VCPS Player, all the media contents are synchronized by using the timing information (from the PPT log file, the whiteboard html files, and the video/audio timing information). Using a pop-up table of contents (ToC), we can make full use of the space in the player. Fig. 6 shows a snapshot of the VCPS Player interface. We will show live demos for both authoring and playing VCPS e-lectures when presenting the paper.

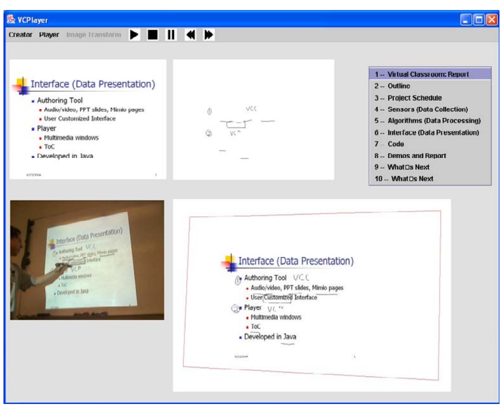



**Fig. 6. VCPS Player interface: integrated presentation of video, PPT and whiteboard pages.**

## 6. CONCLUSIONS

A multimedia system with portable classroom sensors is presented in this paper with three main components: automatic multimedia data collection and synchronization, intelligent multimedia integration, and user-selectable e-presentation interface design. As the core of this work, a hybrid approach for registering images from completely different sources has been proposed for the purpose of e-lecture production. The two different sources, PPT slide images and the handwriting contents, are connected by a video camera that can see both of them. We have implemented several techniques to achieve a fast and robust integration of different media. Experimental results are presented to validate our approach. This work could also be useful for other applications where images from different sources need to be geometrically registered.

We realize it is important to have a usability study of our Virtualized Classroom system in real classroom use. This is still being actively pursued by both us at CCNY and our collaborators at UMass-Amherst.


## 7. ACKNOWLEDGEMENTS

This work is partially supported by NSF under Grant No. CNS-0551598 and Grant No. CNS-0424539, CISDD and CUNY GRTI. We would like to thank Profs. Rick Adrion, Jim Kurose and Edward Riseman at UMass-Amherst and Dr. Parviz Kermani at IBM for the fruitful discussions and close collaborations in this project. We are grateful to anonymous reviewers for their insightful comments and suggestions.


## 8. REFERENCES


[1]. Gregory D. Abowd, Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment, *IBM Systems Journal*, Special issue on Pervasive Computing, 38(4), 508-530, Oct. 1999

[2]. M. Bianchi, "AutoAuditorium: a Fully Automatic Multi-Camera System to Televise Auditorium Presentations, *Proceedings of the Joint DARPA/NIST Smart Spaces Workshop*, Gaithersburg, MD, July 1998.

[3]. H. Chen, R. Sukthankar, G. Wallace, K. Li. Scalable Alignment of Large-Format Multi-Projector Displays Using Camera Homography Trees. *Proc. Visualization*, 2002

[4]. R. B. Dannenberg, "Are Just-In-Time Lectures Effective At Teaching?" June 1997 http://www.jitl.cs.cmu.edu/effectiv.htm

[5]. W. Li, H. Tang and Z. Zhu, Automated registration of high-resolution images from slide presentations and whiteboard handwritings via a low-cost digital video camera. *CVPR'04 Workshop on Image and Video Registration,* July, 2004

[6]. S. Mukhopadhyay and B. Smith, Passive Capture and Structuring of Lectures, *Proc. ACM Multimedia*, Oct 1999

[7]. PINY ELN: e-Learning Lecture Navigator, http://westlabp4-7.poly.edu/webpage/index.htm

[8]. R. Raskar and P. Beardsley, A Self Correcting Projector**,** In *CVPR'01*, Hawaii, Dec 2001

[9]. M. Stern, J. Steinberg, H.I. Lee, J. Padhye, J. Kurose, MANIC: Multimedia Asynchronous Networked Individualized Courseware, *Proc. Educational Multimedia and Hypermedia*, 1997.

[10]. R. Sukthankar, R. Stockton, M. Mullin. Smarter Presentations: Exploiting Homography in Camera-Projector Systems. *CVPR'01*

[11]. W.Xu, C.Wang, CGT: a Fast Thinning Algorithm Implemented on a Sequential Computer. *IEEE trans. Sys, Man, Cyb.*, Oct 1987

[12]. Z. Zhang, A flexible new technique for camera calibration, *IEEE-PAMI*, 22(11), pp. 1330-1334, 2000.

[13]. Z. Zhu, C. McKittrick and W. Li, Virtualized Classroom – automated production, media integration and user-customized presentation, *MDDE'04*.