How are they Watching Me
Learning from Student Interactions with Multimedia Objects Captured from Classroom Presentations

Caio César Viel[1], Erick Lazaro Melo[1], Maria da Graça C. Pimentel[2] and Cesar A. C. Teixeira[1]
[1]*DC, Universidade Federal de São Carlos, São Carlos, SP, Brazil*
[2]*ICMC, Universidade de São Paulo, São Carlos, SP, Brazil*

Keywords: Interactive Multimedia, E-learning, Ubiquitous Capture, Capture and Access, NCL, Interactions.

Abstract: The performance of a teacher in the exposition of a subject is a rich experience that can be captured and transformed into a corresponding multimedia learning object, given the multimodal and multi-device nature of the presentation. Using as a starting point an interactive multimedia object which is an electronic version of a problem solving lecture recorded by the teacher, in this paper we report how a group of students interacts with one multimedia learning object composed of synchronized videos, audio, images and context information. The qualitative analysis of the data allows the teacher to infer useful information not only for refining the lecture content but also for improving its presentation. The case study presented illustrates how a similar analysis can be performed by other instructors with respect to their own lectures, and demonstrates both the power of capturing the multimodal and multi-device nature of the original presentation, and the utility of logging the student-multimedia learning object interaction.

1 INTRODUCTION

When lecturing to her students, the performance of an instructor in the classroom can be considered a multimodal and multi-device live presentation that can be captured and transformed into a corresponding multimedia learning object.

The classroom activity is the primary learning context in many courses (Abowd et al., 1999), so capturing such activities, lectures in special, may be interesting for several reasons. From the attendee's perspective, a student may use the recordings when solving assignments or to study for an exam, or a student who misses a class may still have access to what was presented by watching the recordings. From the instructor's perspective, a professor who will be absent from the campus may prepare a recorded lecture to deliver to the students. Moreover, a previously captured lecture may be improved and reused, or a portion of captured lecture may be used as a complementary learning object in different educational approaches. Last but not least, captured lectures can be a valuable resource for e-learning and distance education courses (Liu and Kender, 2004).

We are aware that there are strong divergences among educators as to the efficiency of the lecture format as a method of instruction in middle school and higher education. Ross, for example, states that "when I was younger, I used to say that it took 40 years for any change in significant higher education to take effect, because that was the time by when all the existing teachers would have retired. I now realize that I was not a cynic, but an optimist, since lectures are just the prevalent as they ever were" (Ross, 2011). However, as Ross himself acknowledges, lectures are still widely used in all levels of education. Moreover, Schwerdt and Wuppermann observe that "contrary to contemporary pedagogical thinking, we find students score higher on standardized tests in the subject in which their teachers spent more time on lecture-style presentations than in the subject in which the teacher devoted more time to problem-solving activities" (Schwerdt and Wuppermann, 2011).

Although recording lectures is common practice in several universities, producing quality video lectures demands a high operational cost. To reduce such costs, many tools for the (semi) automatic capture of lectures were developed in the past (Brotherton and Abowd, 2004), (Chou et al., 2010), (Dickson et al., 2010), (Halawa et al., 2011), (Nagai, 2009). However, such tools usually record only video streams and generate, as a result, a single video stream (e.g. a

Viel C., Melo ., Pimentel M. and A. C. Teixeira C..
How are they Watching Me - Learning from Student Interactions with Multimedia Objects Captured from Classroom Presentations.
DOI: 10.5220/0004454500050016
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (ICEIS-2013), pages 5-16
ISBN: 978-989-8565-61-7
Copyright © 2013 SCITEPRESS (Science and Technology Publications, Lda.)

podcast). In several scenarios, this may not be always enough to reproduce the classroom experience.

The classroom itself can be viewed as a rich multimedia environment where audiovisual information is combined with annotating activities (Abowd et al., 1999). Furthermore, the context of the class (e.g. the slide being presented, what the lecturer says and her body language) and how the different audiovisual contents relate to each other are also important. For instance, sometimes it is necessary to relate the slide presentation with the whiteboard for the comprehension of an exercise or lesson (Dickson et al., 2012). In addition, the interaction between the lecturer and the students is also a valuable part of the learning process.

In this work, capturing a presentation means recording the audio and one or more video streams of the speaker, the images presented on the screen or projector, the writings and drawings made on whiteboards, and capturing relevant contextual information – the aim is to use the captured information to automatically generate an interactive multimedia object, as proposed by the Linking by Interacting paradigm (Pimentel et al., 2000). We refer to as an "interactive multi-video object" the composition of several videos, audio and some static media, properly synchronized and with facilities for flexible interaction and browsing.

From the multi-video object, the lecture may be reconstituted and explored in dimensions not achievable in the classroom. The student may be able, for example, to obtain multiple synchronized audiovisual content that includes the slide presentation, the whiteboard content, video streams with focus on the lecturer's face or the lecturer's full body, or the lecturer's web browsing, among others. The student may choose at any time what content is more appropriated to be exhibited in full screen. The student may also be able to perform semantic browsing using points of interest like slides transitions and the position of lecturer in the classroom. Moreover, facilities can be provided for users to annotate the captured lecture while watching it, as suggested by the Watch-and-Comment paradigm (Cattelan et al., 2008).

In this paper we report how a group of students interacts with a multimedia learning object composed of synchronized videos, audio, images and context information, and discuss how the analysis of the interaction data allows the instructor to infer useful information for improving the lecture. The case study illustrates how a similar analysis can be performed by other instructors with respect to their own presentations, and demonstrates both the power of capturing the multimodal and multi-device nature of the original presentations, and the utility of logging the student-multimedia learning object interaction.

This paper is organized as follows: in Section 2 we discuss related works; in Section 3 we describe our proposed model to capture live lectures; in Section 4 we present our current prototype implementation; in Section 5 we present one case study in which one instructor used the prototype to capture one problem solving session and generate an associated multimedia learning object; in Section 6 we detail lessons learned from the instructor after a qualitative analysis of the interaction a group of students had with the learning object; and in Section 7 we present our final remarks.

2 RELATED WORK

Several authors report results from building systems designed to capture lectures. The AutoAuditorium records classroom activities using a spotting and a tracking camera controlled by computers. The camera orchestration is carried out in real-time using some heuristics based on audiovisual production. The main idea is to create a "TV-like" production without the usual cameraman, video director, audio engineer and other professionals (Bianchi, 2004).

Lampi et al. consider the use of multiple cameras to record lectures. The authors use sensors and computational vision techniques to do the cameraman's job. They also use a finite state machine to define, at each moment, which camera stream should be included in the final stream (Lampi et al., 2008).

Nagai uses an environment with a high definition camera (*Advanced Video Coding High Definition* - AVCHD) placed at the back of the classroom. The camera can record the whole lecture scene (lecturer, whiteboard, slide presentation, students, etc.). By using tracking techniques, the camera performs digital zoom to what is considered the focus of attention at different moments (Nagai, 2009).

Chou et al. use tracking techniques to detect the lecturer's movements and screens (whiteboard, slide presentation) changes. A camera action table is then queried to get what must be done (zoom in, zoom out, pane, etc.) in order to highlight the image that must be the focus of attention (Chou et al., 2010).

All the aforementioned works differ from the work reported in this paper in that the resulting product of the lecture capturing process is a single video stream instead of a multi-video object.

In the work of Liu et al., lectures are captured in a similar process to the ones mentioned before, resulting a single video stream. The difference is that the set of slides used in the presentation is added to the video stream. However, the slides are not synchronized with the video (Liu and Kender, 2004). Given that the result is single-video-stream, students do not have autonomy to choose the camera that gives them the best view of the lecture for each situation, or to focus their point of interest, as allowed in our multi-video object.

ClassX is a tool designed for online lecture delivery (Halawa et al., 2011) (Pang et al., 2011). A live lecture is captured by means of an AVCHD stream split in several virtual standard resolution cameras. By using tracking techniques, the most appropriated virtual camera for a given moment is chosen and streamed to the remote students. The students have the opportunity to choose a different stream from another virtual camera or even watch the original AVCHD stream, and a synchronized slide presentation is offered — but no other navigation facilities are available the students.

REPLAY is a system for producing, manipulating and sharing lecture videos (Schulte et al., 2008). Besides offering similar features to the aforementioned systems, REPLAY uses computer vision to recognize written words, and deploys MPEG-7 to index the videos. Although REPLAY allows more navigation alternatives than the previous systems, it does not produce an independent multi-video object.

Other authors report the use of other features such as image processing and audio transcription (Dickson et al., 2012), (Dickson et al., 2010)), (Brotherton and Abowd, 2004), (Cattelan et al., 2003), the result being hypermedia documents that offer interfaces providing different ways of indexing the recorded information. The model for capturing and recovering lectures presented in this paper allows more flexibility. This flexibility results from the ability to specify the context information that must be captured, and to specify how this context information should be combined to generate a multi-video object, or to promote live interventions in the classroom during the capture process — for example in the case that there is a change in the illumination of the room because the light was off.

3 UBIQUITOUS CAPTURE AND AUTHORING

In order to produce quality lecture videos, the conventional lecture recording process usually requires the presence of audiovisual professionals. Our infrastructure offers a self-service approach, allowing the instructor to record a lecture herself. Some solutions usually rely on computational vision, tracking techniques and sensors to perform camera orchestrations in a attempt to produce a single video or audio stream output.

As detailed elsewhere, the model we have proposed goes a step further (Viel et al., 2013). As depicted in Figure 1, the model aims at capturing all the content presented in the classroom. The capture process is pervasive, does not rely on human mediation and generates automatically an interactive multi-video object which preserves as much as possible of the lecture content and context.

An environment, usually a classroom, is instrumented with physical devices (Figure 1(1)), such as video cameras, microphones, whiteboards, interactive whiteboards and slide projectors. The instrumented classroom may also contain sensors, such as temperature sensors and luminosity sensors, and secondary screens, such as notebooks, TVs, tablets, etc. The video cameras should be placed in points where they can frame important classroom's points (instructors, students, whiteboard, slide presentation, etc.).

Computer devices capture all the content produced by the physical devices used in the classroom (e.g. whiteboards and slides) and represent them as video, audio and data streams (Figure 1(2)). Cameras produce video and audio streams, microphones produce audio streams and sensors produce data streams. By capturing the screen output from the secondary screens or by intercepting the signal sent to the slide projector, we can also produce video streams. The electronic whiteboard can produce both data and video streams. By capturing its strokes we can generate a data stream; intercepting the signal sent to its projector, we can generate a video stream.

All such streams are stored (Figure 1(3)) for further use in the multi-video object generation. The streams are also sent to the *capture controller* (Figure 1(4)), a component responsible for managing the capture process. The *capture controller* uses signal analysis to analyse the captured streams and to send commands (Figure 1(5)) back to the physical devices and actuators (Figure 1(6)) present in the classroom.

The instructions in the *capture controller* are defined in a customizable action table. The action table can be used to define actions for certain events which may occur during the capture process. For instance, zooming into the image of a specific camera when the lecturer starts talking, or activating an actuator in order to reduce the light intensity when the lecturer starts a slide presentation.

Our model allows the instructor to split her presentation in different modules, an approach usually adopted in e-learning platforms.[1] A multi-video pre-

[1]Examples include http://www.coursera.org and http://www.edx.org

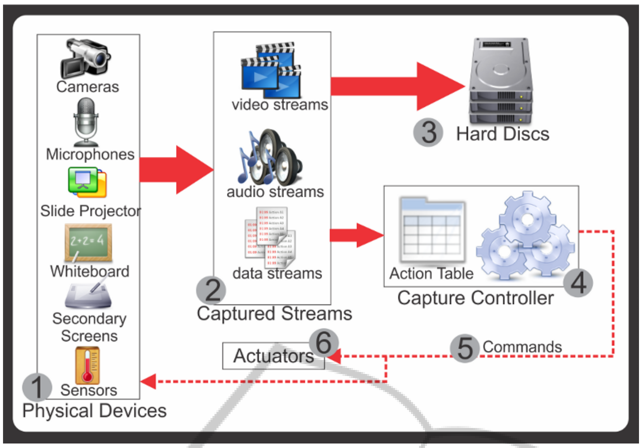


Figure 1: Capture Workflow.

sentation can be composed of one or more modules. This is useful to better organize the content of a lecture. The lecturer may, for instance prepare a problem solving presentation with one exercise per module. And, it also allows the lecturer to take breaks during the recording process and the students to navigate in the modules of the multi-video presentation.

Splitting the presentation into modules can also minimize the time need for repeating the recording in case of errors. For instance, if in one module the lecturer starts stuttering or becoming confused and wishes to make a retake, she only needs to record that module again. Reusing the modules to compose a new presentation is another advantage of splitting the recording process into modules — reuse is in fact one of the main ideas underlying learning objects.

Given that the processes of analysing and converting the captured streams can demand much computational power and time, once the capture process is finished the data is transferred to a server for further processing.

Considering points of Interest as moments in the lecture which may have particular importance for students, we designed recognizer components that use one or more captured streams to automatically detect potential points of interest. The points of interest can be used to provide a more semantic navigation over the multi-video object, allowing the students to seek for the next slide transition, for instance.

Some points of interest have been suggested in the literature ((Dickson et al., 2012), (Cattelan et al., 2003) and (Brotherton and Abowd, 2004)), while others were inspired on our own observation of real lectures. Examples of Points of interest are slide transition, whiteboard interaction and change the eye-gaze of the instructor.

The resulting multi-video learning object is composed of videos and other captured media. Although the multi-video object cannot reproduce several aspects of the live lecture experience (live interactions, odors, temperature, etc.), it offers other facilities to the students when they are interacting with the object.

4 PROTOTYPE

As a proof-of-concept of the model, we developed a prototype tool for capturing lectures and generating multi-video objects. This prototype was mainly developed in Python. Figure 2 depicts an overview of the prototype.

The prototype is composed of three main parts: the *Capturing tool* used to capture streams; the *Processing tool* in charge of stream analysis and the generation of the multi-video object; and the *Presentation tool*, which allows the user to playback the multi-video object.

Capturing Tool

The *Capturing tool*, named *Classrec*, (Figure 2(A)) performs the lecture capturing process. Each computer used in the capturing process runs an instance of *Classrec*, and one of these instances is selected to be the session manager (Figure 2(B)). It corresponds to the *Capture Controller* of the workflow (Figure 1). The session manager is responsible for handling the lecturer's stimulus and for controlling the other *Classrec* instances, keeping them synchronized.

The capturing process is based on video streams. *Classrec* captures content (video and audio streams) produced by AVCHD and outputs produced by computers (such as computer screens, slide presentations, etc.). It also records metadata about the lecture, such

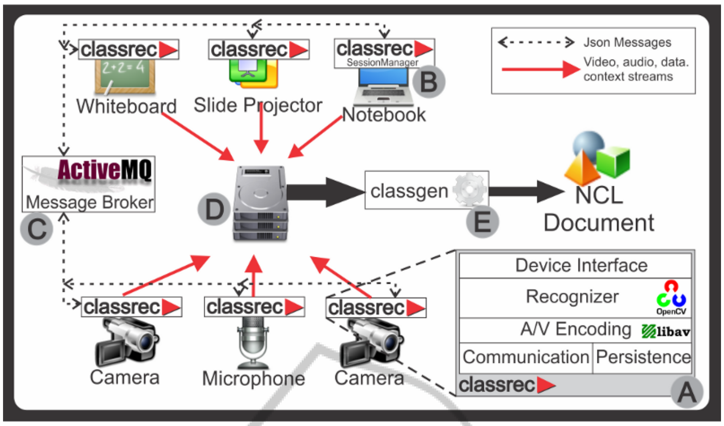


Figure 2: Prototype Overview.

as module structure, available streams and authoring information into an XML file.

We opted to capture the electronic whiteboard output as a video stream instead of its strokes. This was done because a video stream is more portable than strokes and, given the modern video encoding as h.264 Advanced Video Codec and the static nature of whiteboard outputs, the bit rate of the video stream is low. We could record a stroke stream, but it would require a specialized media player to play it back (as it is the case with other systems, e.g. (Müller and Ottmann, 2000)).

Some streams, such as slides, whiteboards and computer screens may contain segments with a lot of static content, but they are still captured as video streams. A possible improvement would be to replace the video for a combination of non-static content videos and a single image to represent a static segment (video with no changes during a period of time).

The communication among the different applications is carried out using the Apache ActiveMQ message broker (Figure 2(C)).

Processing Tool

The *Processing tool*, named *Classgen* (Figure 2(E)), performs the multi-video generation process. This tool uses as input the video streams and metadata recorded by *Capturing tool*. It also supports an XML configuration description language, which allows the specification of which recognizers (and its inputs) should be used, and the codecs that should be used to encode audio and video.

We have implemented recognizers capable of detecting (i) the presence of a lecturer in a video stream; (ii) if the lecturer is facing a camera; (iii) slides transitions; (iv) interactions with whiteboard or PC; and (v) a list of spoken keywords.

It is also possible to specify an orchestration of video streams in order to produce a new video stream. This is useful in environments with multiple cameras recording different angles of the lecturer. Through the XML configuration description language, it is possible to select which stream will be used in the orchestration and how to orchestrate then. For instance, it is possible to specify that when a recognizer detects the lecturer's face in video segments, the camera orchestration stream should include that segment.

Classgen uses the OpenCV library (Bradski, 2000) to perform pattern recognitions in order to identify points of interest for composing the context stream. The media manipulation during the orchestration process and the audio/video conversion is handled by the *libav* library.

Once the several processes associated with recognition of points of interest, orchestration and video conversion are concluded, the information they generate (the specification of the points of interest, the orchestration stream, and the converted streams) are stored in the *XML lecture*. The XML is then passed to a component of the *Processing tool* responsible for generating the final multi-video object (Figure 2(5)). Our prototype generates NCL[2] (ABNT, 2007) documents, but the *Classgen* can be extended to generate other types of multi-video objects, such as HTML5 pages or stand-alone desktop, tablet or smartphone applications.

The XML configuration description language can also describe the video streams (including the orchestration, if any) and points of interest will be used in the final multi-video object. It is also possible to generate different multi-video objects using the same recorded lecture (for instance, by using the orchestration stream or not).

[2]Nested Context Language - http://ncl.org.br/en

Presenting Tool

It is desirable to offer students a platform-independent way to access the captured lectures. We would like to avoid students having to install specific software to playback the lecturers. To fulfill this requirement we choose a web-based implementation.

The multi-video object generated from the capture imposed some challenges. In the scenario where we considered the generation of the object directly in HTML5 + JavaScript, a large development effort to implement the synchronization capabilities was estimated. We also noticed that most obstacles identified in the HTML5-based implementation would be easily overcome with the use of a declarative language specialized in media synchronization. However, there were no solutions to support it that did not demand external plug-ins.

As a result of these needs, we were motivated to propose and develop a multimedia presentation engine based on standard Web technologies. We conducted an implementation based on HTML5 + JavaScript that enables the presentation of multi-video NCL documents, named WebNCL[3] (Melo et al., 2012). Thanks to WebNCL, any device which has an HTML5-compatible browser (PC, Smart TV, Tablet, Smart Phone, etc.) can present NCL documents natively.

The choice for implementing support to the NCL language was taken because it is a powerful language for media synchronization, under active development and adopted as iDTV (ABNT, 2007) and IPTV standards (H.761, 2009). A good side effect of this choice was the possibility to reuse the content generated in different platforms.

Figure 3 shows running NCL learning objects generated by the prototype. The NCL document offers some facilities for students. One of these facilities is the synchronization of the captured audio/video. The multi-video object synchronizes the multiple audio/video streams, so students can see what was written in the whiteboard when the lecturer points to the slide presentation. This synchronization is essential to recover the whole audiovisual context of the captured lecture at a given moment. It is also possible to insert non-synchronized complementary media to the multi-video object like, for instance, an image from a textbook.

The multi-video object offers a more semantic and easy way to navigate in the captured lecture than timeline navigation, common in video (however, timeline navigation is still present). For instance, the student

[3]WebNCL is an open-source software, available at http://webncl.org

(a) Timeline

(b) Multiple Videos

(c) Full-screen

Figure 3: Multi-video learning objects.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

can move forward to the next slide transition or backwards to the previous one. When the lecturer begins to write something in the whiteboard, the student can skip all the writing process and see the final result. In a future implementation, students will also search for a keyword and move forward in the multi-video object to the point where the lecturer said “for instance”.

Similar to in-classroom lecture, wherein the student can pay attention to different spots (the lecturer, whiteboard, slide presentation, the textbook, or another screen), the multi-video object, which contains several navigation controls besides the timeline (Figure 3(a)), allows the student to choose whether he wants to see more than one video at the same time (Figure 3(b)), or which video stream he wishes to see in full screen (Figure 3(c)).

Finally, the student has the facility to make annotations in the multimedia object by means of the watch-and-comment paradigm. For instance, he can mark some part of the lecture as important or irrelevant, or he can delimit a snippet of the lecture which he did not understand for further research or to ask the professor or tutor. He can also make comments on the lecture via audio or text, in similar in-classroom students do with paper and pencil.

Instrumented Classroom

The capture-tool prototype was deployed in a multipurpose room (Figure 4). At the front of the room (Figure 4(a)) there is a conventional whiteboard, an electronic whiteboard and a notebook in which the presenter can browse the Web or use other software. The interactive whiteboard can be used to present slides (there is a Bluetooth presenter to control the presentation) and it allows drawing and writing over the screen. At the back of the room (Figure 4(b)) we placed two AVCHD, one with focus on the interactive whiteboard and the other with focus on the conventional whiteboard. We placed a webcam as a wide-shot cam, framing the whole front of the room. The cameras are locked cabinets when not in use.

(a) Front Side (b) Back Side

Figure 4: Instrumented Classroom.

We invited six instructors to use the prototype and record presentations. Four instructors recorded a lecture simulation (without students), one professor recorded a conventional lecture (with students), and one instructor recorded a problem solving class. We also used the prototype to record the presentation of term paper.

In the next sections, we report on results from analysing the interactions students had with the multimedia learning object resulting from the capture of the problem solving class.

5 CASE STUDY: CAPTURE LECTURE

Using the capture-tool prototype, one instructor captured one lecture: the capture was made in several modules, without students in the classroom. The students had access to the multimedia learning object to prepare to their final exam.

The lecture captured was a problem solving session for a Computer Organization course in which an instructor solved a total of 15 exercises. These exercises were related to each other and usually a subsequent exercise used some results from the previous one. The exercises also become more difficult as the presentation progressed.

The presentation was organized into 12 modules, performing a total of 1 hour and 18 minutes of content. The first 3 exercises were grouped in the module 1, module 5 contained 2 exercises, and all the other modules presented one exercise each.

Figure 5 depicts the multimedia object generated from the presentation. There are four streams: (1) the capture of the projected slide, which contained the description of the exercise; (2) the camera focused on the conventional whiteboard; (3) the camera focused on the slide; and (4) the wide-shot camera. Although the generation process has a feature that allows the automatic orchestration of the cameras (e.g., the automatic selection of which video stream would be presented in the main (bigger) window), in this study case we did not use it. The aim was to exploit the students' interaction, forcing them to choose, for a better learning experience, which would be the video to be presented in the main window at each instant.

Figure 5: Problem Solving Presentation.

The multimedia object was made available for the students in the Web and, using the WebNCL's log API, we logged all the interactions carried out by the students, such as when and where the users clicked and to which point they seek in the presentation timeline. The loggged data were stored in a NoSQL database. We developed python scripts to extract information relative to how the students interacted with the multimedia object.

Figure 6 presents information about the time spent by the students, as well as the number of interactions they performed with the multimedia object. Each point in the horizontal axis represents a student (identified in the chart as letters from A to R). The blue bars show the amount of time each student spent watching the multimedia object (left vertical axis) and red bars show the number of interactions each student performed (right vertical axis).

The total duration of the 12 modules was 1 hour and 18 minutes. Eighteen students watched the presentation for at least 4 minutes. The average playback time of these 18 students is 3542.67 seconds (about 59 minutes) with a standard deviation of 2382.23 seconds (about 39 minutes). The average number of interactions of the students is 118.55 with a standard

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

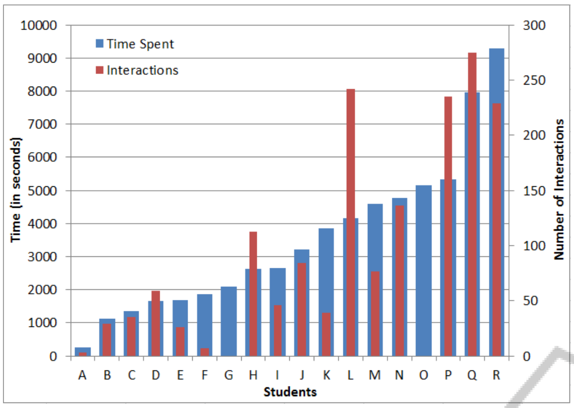


Figure 6: Students Interactions.

deviation of 99.58.

Figure 7 summarizes the number of interactions of each category performed by the students. The interactions were organized in the following categories:

- Main Video Selections: interactions carried out by the students in order to change the main video stream;
- Play/Pause: interactions causing the pause and the resume of the playblack;
- Timeline navigation: interactions that cause a move forward or backward through the timeline;
- Module Navigation: interactions that cause the change of the module currently watched;
- Points of Interest: interactions resulting from navigation by points of interest (e.g. slide transitions).

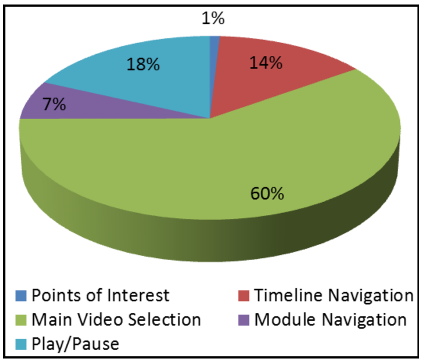


Figure 7: Interactions per Category.

Figure 8 summarizes how much time each module was watched. In order to get a better visualization, the values in the left vertical axis were normalized by the module time length. The blue bars represent the time in which the presentation was running (not paused) and the red bars are the time in which the presentation was paused. The green line represents the number of students that watched each module for at least 10% of their time length. The figure suggests that the modules in which the students spent more time were the module 2 and module 4. It also suggests that the number of different students that watched the modules decreases as the presentation progress

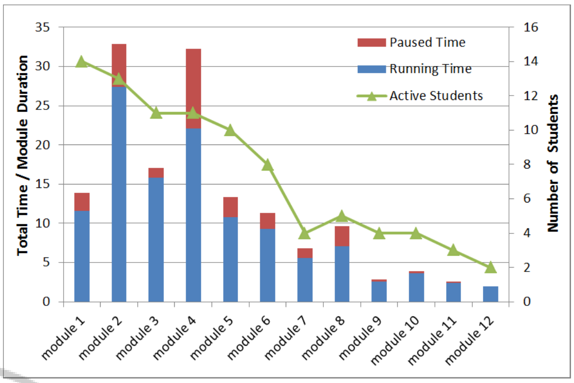


Figure 8: Presentation Modules Statistics.

Figure 9 summarizes the watching attendance of some modules. The horizontal axis is the number of seconds of each module (Presentation Space). The blue line represents the number of times the instant was watched by students, and the red line the number of different students that watched each instant.

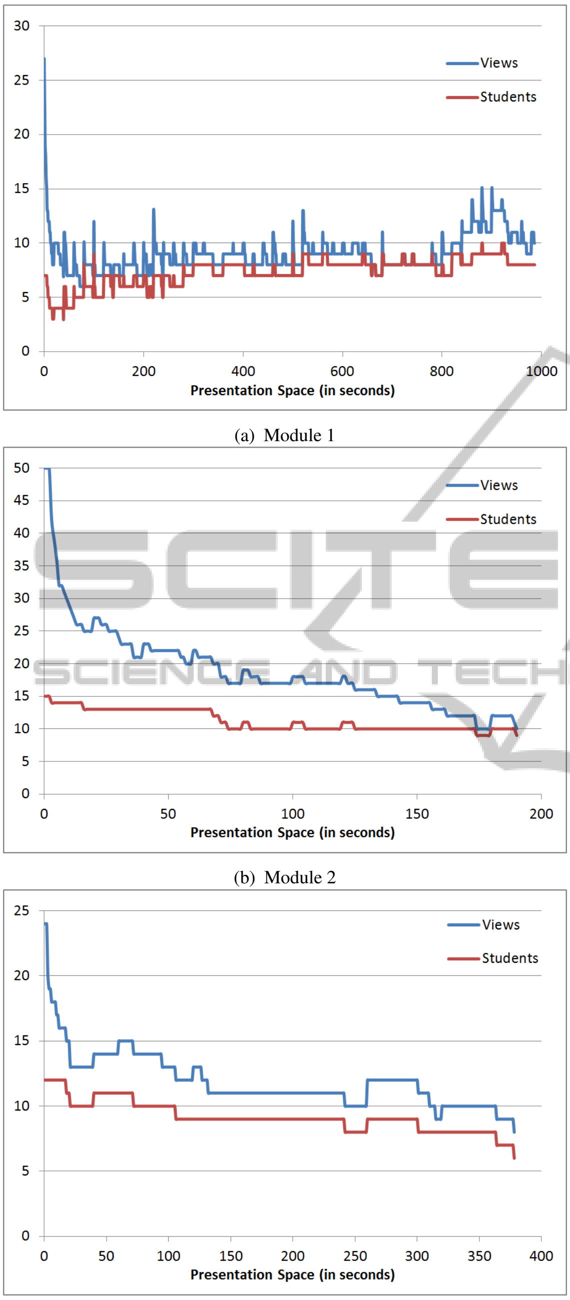
As the modules always start from second 0, it is natural that the attendance of the first seconds is bigger. The points where the blue line is above the red line mean that the moment was watched more than once by the same students. This graphic can be useful for lecturers to find out which parts of a lecture are more useful or important for the students, or even to identify points where students have difficult to understand. For instance, after the second 800 in Module 1 (Figure 9(a)), the blue line deviates from red line, it suggests that that segment of module 1 were watched more times by the students.

Given that the multimedia object has more than one video stream and that the students can choose which stream they wish to see as the main stream, the information of which stream is most selected as the main stream at each moment can be useful.

Figure 10(a) and Figure 10(b) summarize which streams were most selected as the main stream in each moment of, respectively, module 1 and module 4. Each line represents how many times a stream was watched in a specific moment. The blue line refers to the slide projection capture (Figure 5(1)); the red refers to the camera focused on the conventional whiteboard (Figure 5(2)); the green camera focused on the slide presentation (Figure 5(3)); and the purple the wide-shot camera (Figure 5(4)).

According to Figure 10(a), the more watched streams were the slide presentation and the whiteboard camera. We can also note that the slide presentation is more watched near the moments when there

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS



(a) Module 1

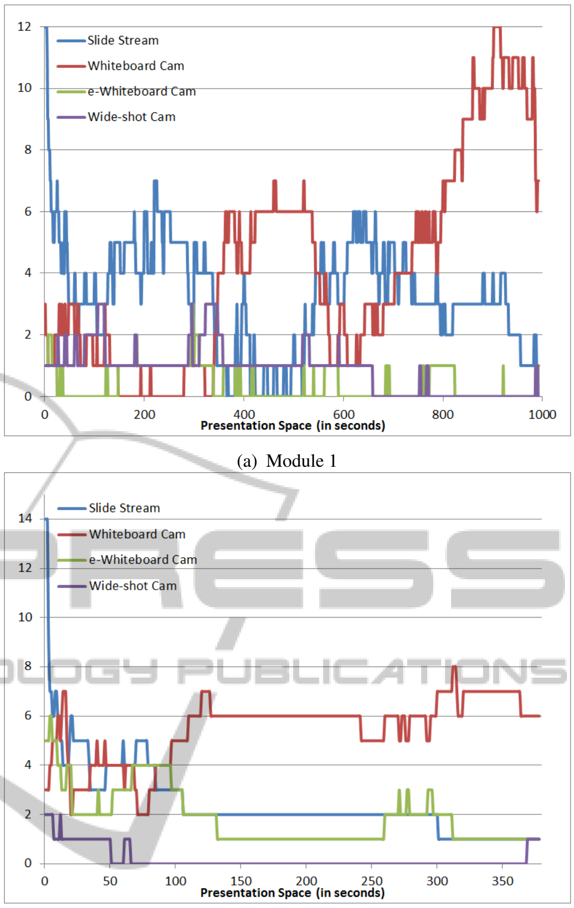
(b) Module 2

(c) Module 4

Figure 9: Modules Attendance.

are slide transitions in the module 1 (seconds 213 and 515). Figure 10(b) suggests that after the second 100 the predominant stream was the whiteboard camera stream.

Figure 11 illustrates the behavior of 2 students when interacting with the presentation for module 1, 2 and 4. The student P (blue line) and Q (red line) are the same students from the Figure 6. The horizontal axis is the playback timeline and the vertical axis is the presentation timeline (presentation space). Vertical straight lines represent a navigation that the student performed during playback and horizontal straight lines represents moments in which the student paused the presentation. These graphics allow to visualize how a student interact with the presentation in detail. For instance, we can observe in Figure 11(a) that student P starts watching from second 180 and performed some backward moves mainly in the end of the presentation. Student Q watched almost linearly until second 650 and then returned to the beginning of the presentation and watched it again until the end performing some pauses.



(a) Module 1

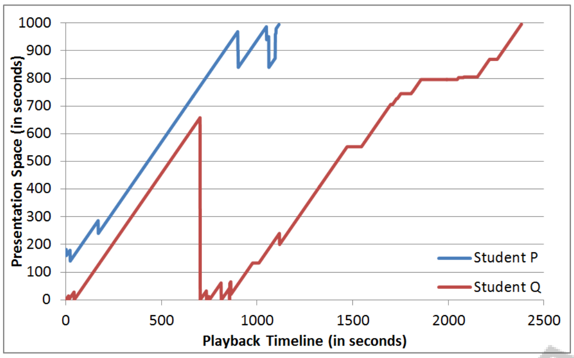
(b) Module 4

Figure 10: Streams View.

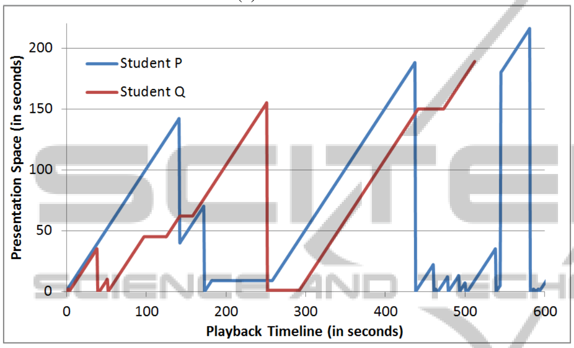
6 LESSONS LEARNED

The graphics were presented to the instructor. He analysed them taking into account the content of his presentation, how it was presented and which and how students interacted with it.

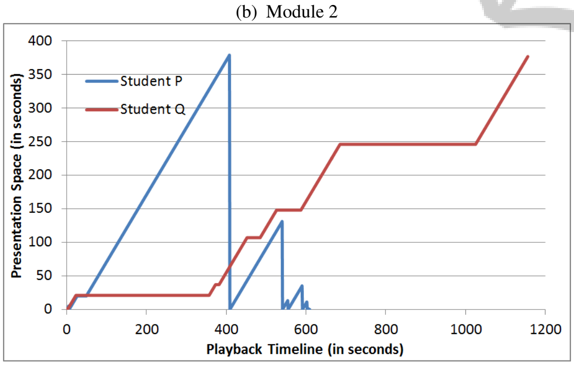
SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS



(a) Module 1



(b) Module 2



(c) Module 4

Figure 11: Students Navigation.

His first observation: “the graphics are very abstract for a teacher to analyse them by himself”. As as consequence, the remaining analysis was carried out then with the help of one of the authors. From now on, what is reported in this section is a combination of what the teacher observed and some conclusions of the authors.

Not all students interacted with the multimedia learning object, even knowing that it could have tips for the exam. Reasons for this may have been the commitments of students with other exams, the late release of the learning object (two days before the exam), and also its long duration, about one hour and twenty minutes (4800 s). As shown in Figure 8, several of the students only watched the first modules. Besides the reasons already mentioned, some of them may have found the presentation boring. A questionnaire with explicit questions could help understand this attitude.

Students were able to view the slides presented in two ways, watching the video of the instructor presenting (and maybe interacting with) the slide on the interactive whiteboard, or watching the slide captured directly from the output of the projector (best quality). The preference was for the latter, as shown in Figure 10(a) with the blue and green lines. It is likely that the type of the presentation, without many interactions with the interactive whiteboard, does not justify the view of the slide in lower quality.

The resulting multimedia learning object may consider context information. It can then guarantee that the focus of the presentation, at every moment, be automatically taken to the main window display. So, when the teacher uses the whiteboard, her or his video could be automatically selected to the main window. The same applies for the videos associated with the interactive whiteboard, the application captures, etc. However, we chose to force the student himself to perform all the video switching. Some students expressed frustration with such duty. The goal was to keep them alert to the presentation in order to make it less monotonous. The strategy worked. As shown in Figure 7, 60% of the interactions (975) were used for selecting the video to the main window. The effectiveness of the strategy in terms of learning, however, needs to be evaluated.

Figure 7 also shows the limited use of the navigation using Points of Interest. Students preferred to use the Timeline (14% of interactions) to control the presentation. Two reasons may be related to this: students are used to the paradigm of watching video in the Web; and the lesson has not encouraged or justified the need for this type of navigation. However, the navigation through the modules happened with a frequency (7%) corresponding to the one expected (and planned) by the teacher.

Figure 9(a) shows that an almost constant public watched module 1 (in terms of number of student). However, the blue line shows some peaks in visits to some parts of the presentation, in terms of times the segment was played. The moments around the 900th second are the evident ones. The analysis of the video in those moments, carried out by the teacher, indicates that the subject could be presented more clearly – that is, there is room for improvement in the way the presentation was made.

Modules 2 and 4 were the most popular, not the 1 as expected for being the first. The visiting time was normalized by the duration of the module in Figure 8.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

As the first module was the one with the longer duration, it may indicate that large modules are more verbose (which was confirmed by the teacher for module 1) and therefore tend to be somewhat repetitive. Moreover, this feature can be further studied since the content of module 1 was less complex than the others.

The navigation patterns, illustrated in Figure 6, show different behaviors by the students. There are students who simply "watch" the presentation and do not perform any interaction at all, even to change the video in the main window, as was the case of the students G and O. These, probably thinking they would be evaluated by their performing in viewing the presentation, let the presentation run without perhaps give attention to it. Others, such as students L and P, watched all the presentation and performed many interactions. There are also students, as Q and R, who, besides interacting a lot, also watched repeatedly parts of the presentation, nearly doubling the original time of the presentation. The figures show that the number of interactions was proportional to the time in which the presentation was watched, which indicates a similar degree of interactivity between the students in the class. Another interesting observation about the behavior of students was made by the teacher: "one of the students who watched and interacted the most with the multimedia learning object, the student N, usually shows a very apathetic behavior in the classroom". This may indicate that interactive multimedia learning objects, generated by capturing multimodal and multi-device presentations, may be a good option for students who like to be in control of what they pay attention to.

7 FINAL REMARKS

Extra-class material may be offered to students in the form of multimedia objects that integrates synchronized text, image, audio and video explanations on the studied subject. A learning object like this can be produced in studios, with support of various professionals. Alternatively, as is the case presented in this paper, the multimedia object can be automatically generated from the ubiquitous capture of a traditional lecture in the classroom. The lecture can be delivered to a group of students, or be delivered to an empty classroom just for capture purposes. Context information informing moments of interest such as slide transitions can be included in the multimedia object to provide students with semantic navigation.

The multimedia object should be instrumented to log the navigation performed by students so that, besides acting as extra-class material, they can be effective as tool that provides feedback which contributes to improve its own content. In the situation presented in this paper, it is the instructor who receives the feedback, which she can analyse to identify improvements not only in terms of the content itself but also in terms of how the exposition was made at the time of capture.

The case study presented suggests how similar analyses that can be performed in other presentations, even though only a portion of the logged information was used. As a result, the analysis is useful both as a reference for the preparation of presentations used in research involving interactive multimedia objects, and in the research in Education.

Regarding future work, we plan to investigate alternatives for: (a) the enrichment of the graphic interface of the multimedia object so as to improve interactivity; (b) the capture of more contextual information during the presentation toward providing novel navigation facilities; (c) the development of visualization tools for the instructor to analyse the information captured while the students interacted with the multimedia object. The aim is to built a general infrastructure that helps building similar capture-based applications (Pimentel et al., 2007).

We also to conduct interdisciplinary research toward better understanding the impact, on education, of the use of multimedia learning objects built from the capture of multimodal and multi-device presentations.

The teacher also noted a relationship between student's performance on assessment on the subject of the presentation and the time each one spent with the multimedia learning object. Most who watched and interacted with all modules of the presentation performed well. The individual analysis of each student can be performed using graphs similar to those shown in Figure 11, for instance.

REFERENCES

ABNT (2007). Associação Brasileira de Normas Técnicas. 2007. Digital Terrestrial Television Standard 06: Data Codification and Transmission Specifications for Digital Broadcasting. Technical report, Part 2–GINGA-NCL: XML Application Language for Application Coding São Paulo, SP, Brazil.

Abowd, G., Pimentel, M. d. G. C., Kerimbaev, B., Ishiguro, Y., and Guzdial, M. (1999). Anchoring discussions in lecture: an approach to collaboratively extending classroom digital media. In *Proc. Conference on Computer support for Collaborative Learning*, CSCL '99. International Society of the Learning Sciences.

Bianchi, M. (2004). Automatic video production of lectures using an intelligent and aware environment. In *Proc.*

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

International Conference on Mobile and Ubiquitous Multimedia, MUM '04, pages 117–123. ACM.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brotherton, J. A. and Abowd, G. D. (2004). Lessons learned from eClass: Assessing automated capture and access in the classroom. *ACM Trans. Comput.-Hum. Interact.*, 11(2):121–155.

Cattelan, R. G., Baldochi, L. A., and Pimentel, M. D. G. (2003). Experiences on building capture and access applications. In *Proc. Brazilian Symposium on Multimedia and Hypermedia Systems*, pages 112–127.

Cattelan, R. G., Teixeira, C., Goularte, R., and Pimentel, M. D. G. C. (2008). Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Trans. Multimedia Comput. Commun. Appl.*, 4(4):28:1–28:24.

Chou, H.-P., Wang, J.-M., Fuh, C.-S., Lin, S.-C., and Chen, S.-W. (2010). Automated lecture recording system. In *Proc. International Conference on System Science and Engineering (ICSSE)*, pages 167 –172.

Dickson, P. E., Arbour, D. T., Adrion, W. R., and Gentzel, A. (2010). Evaluation of automatic classroom capture for computer science education. In *Proc. Annual Conference on Innovation and Technology in Computer Science Education*, ITiCSE '10, pages 88–92. ACM.

Dickson, P. E., Warshow, D. I., Goebel, A. C., Roache, C. C., and Adrion, W. R. (2012). Student reactions to classroom lecture capture. In *Proc. ACM Annual Conference on Innovation and Technology in Computer Science Education*, ITiCSE '12, pages 144–149. ACM.

H.761, R. I.-T. (2009). Nested context language (NCL) and ginga-NCL for IPTV services. Technical report.

Halawa, S., Pang, D., Cheung, N.-M., and Girod, B. (2011). ClassX: an open source interactive lecture streaming system. In *Proc. ACM International Conference on Multimedia*, MM '11, pages 719–722. ACM.

Lampi, F., Kopf, S., and Effelsberg, W. (2008). Automatic lecture recording. In *Proc. ACM International Conference on Multimedia*, MM '08, pages 1103–1104. ACM.

Liu, T. and Kender, J. (2004). Lecture videos for e-learning: current research and challenges. In *Proc. International Symposium on Multimedia Software Engineering*, pages 574 – 578.

Melo, E. L., Viel, C. C., Teixeira, C. A. C., Rondon, A. C., Silva, D. d. P., Rodrigues, D. G., and Silva, E. C. (2012). WebNCL: a web-based presentation machine for multimedia documents. In *Proc. Brazilian symposium on Multimedia and the web*, WebMedia '12, pages 403–410. ACM.

Müller, R. and Ottmann, T. (2000). The "authoring on the fly" system for automated recording and replay of (tele)presentations. *Multimedia Systems*, 8(3):158–176.

Nagai, T. (2009). Automated lecture recording system with avchd camcorder and microserver. In *Proc. Annual ACM SIGUCCS Fall Conference*, SIGUCCS '09, pages 47–54. ACM.

Pang, D., Halawa, S., Cheung, N.-M., and Girod, B. (2011). ClassX mobile: region-of-interest video streaming to mobile devices with multi-touch interaction. In *Proc. ACM International Conference on Multimedia*, MM '11, pages 787–788. ACM.

Pimentel, M., Abowd, G. D., and Ishiguro, Y. (2000). Linking by interacting: a paradigm for authoring hypertext. In *Proc. ACM on Hypertext and Hypermedia*, HYPERTEXT '00, pages 39–48. ACM.

Pimentel, M., Baldochi Jr., L. A., and Cattelan, R. G. (2007). Prototyping applications to document human experiences. *IEEE Pervasive Computing*, 6(2):93–100.

Ross, G. M. (2011). What's the use of lectures? - forty years on. *Discourse*, 10(3):23–41.

Schulte, O. A., Wunden, T., and Brunner, A. (2008). REPLAY: an integrated and open solution to produce, handle, and distributeaudio-visual (lecture) recordings. In *Proc. Annual ACM SIGUCCS Fall Conference: moving mountains, blazing trails*, SIGUCCS '08, pages 195–198. ACM.

Schwerdt, G. and Wuppermann, A. C. (2011). Sage on the stage: Is lecturing really all that bad? *Education Next*, 11(3):62–67.

Viel, C. C., Melo, E. L., Pimentel, M. d. G., and Teixeira, C. A. C. (2013). Presentations preserved as interactive multi-video objects preserved as interactive multi-video objects. In *Proc. Workshop on Analytics on Video-Based Learning*.