

Two Automated Mechanisms to Create eLectures and to Videotape Regular Lectures

Nael Hirzallah

Faculty of Information Technology, Applied Science University
Amman, Jordan

Abstract

Many institutes are offering lectures that students may watch online. These lectures vary in type from true multimedia sophisticated documents to eLectures to videotaped regular lectures that may include auto generated closed caption. This paper studies some formats that can be easily composed by instructors. Whether adopting the eLectures format or videotaped lecture format, the paper proposes two automated mechanisms to create such lectures without the intervention of a camera or video mixer operator. The first mechanism depends on the absence of mouse motion events; while the second mechanism depends on the instructor's facial and hand detections.

Keywords: *Multimedia, eLearning, Open Course Ware, Content Development, Authoring tool*

1. Introduction

The major challenge in the implementation of eLearning is the content development process. From one point of view, it is very expensive. From another point, it is time consuming. One may either develop an eLearning platform or choose one from the market to install in the institute or company, but developing the content that is customized to specific audience is another story.

Initially, groups that include developers, designers, material specialists, and testers get together in order to develop an online course. Such a team may have to hold extensive meetings to mainly bridge the communication gap among the various poles of the team. This is because the material specialists are usually from the academic sector, while the others are from the industrial sector. This, by itself, is considered sometimes the first road-block that the project manager must overcome so all can talk the same technical language. The team leader may take sometimes expensive measures to overcome this problem and move forward. However, replacing the material specialist is usually the last option to take. Once the problem is overcome, a complete course would then be split into modules. Once the development and testing cycles are finished and during the integration phase, they will be recombined back. This cycle will be repeated for the second course and so on. However, for a different

course, the team may have to work with another material specialist, and the whole process would then be repeated from bridging the communication gap phase. This was found out to be a very expensive process in terms of money and in terms of time. Therefore, for these reasons, many institutes are ending up implementing an eLearning platform that is rich with features but poor in real content.

The other choice for these institutes is to train their material specialists on using content development tools. Based on the chosen tools and the backing up policies adopted, the decision may end up to be the right one or it would not reach the goal desired by the institute. Many authoring tools are available off-shelves, such as Adobe Captivate [1]. To select the right tool, one should consider the time and efforts the material specialists would need to use it well. It should also consider the time the specialists may need to develop a complete online course. In most cases, the easier to learn, the faster to use the tool is, but the fewer features would be used. Authoring-on-the-fly tools, AOF, [2,3] were developed to satisfy this purpose.

The third and last choice offered to these institutes is to videotape their regular lectures in order to produce online lectures using simple video format. A group of these online lectures will form a complete online course. Three different models are available for videotaping a lecture based on the number and type of cameras used. The different number and types of cameras are: more than one fixed-location camera with a mixing device, a single moving Camera, and a single fixed-location wide lens camera. In the first two models, a video editor/mixer person and a camera-man may be needed.

Due to the high cost of developing sophisticated multimedia lectures and the speed of the eLearning marathon among educational institutes, many are adopting the AOF and videotaping approaches discussed earlier for online lectures development. For AOF, one of its main motivations when compared with videotaping regular lectures is to minimize the bandwidth and storage capacity needed. In AOF, the output frame is usually composed of

many window regions (similar to Figure 1). These regions include images and hyperlinks along with one or two smaller size videos. It is considered to produce a little more sophisticated outputs than videotaped regular lectures. But, with better new video compression techniques and faster home internet speed, storage capacity and bandwidth do not anymore represent a problem. Thus, more and more universities are adopting the videotaping approach.



Figure 1 Slides and instructor view taken from [4]

This paper studies both approaches, namely the AOF and videotaping by presenting the cons and pros of each. It then proposes an automated simple mixed approach between the AOF, the more sophisticated approach, and that of the videotaping approach, in order to benefit from the cons of both approaches. It also presents the results of a survey performed on a group of students from various fields regarding their perception to the kind of video they believe would help them stay focused during an online lecture. Finally, the paper concludes with the last section.

2. Overview

Content development is by far the most challenging task in eLearning. It is a time consuming and costly process. One may categorize the approaches used for this task based on how sophisticated, interactive, and whether it is a true multimedia or not the outcome is. A document is said to be a True Multimedia if it contains some kind of interaction that may change the presenting scenario as well as some animation. Thus, a traditional video cannot be called a true multimedia. True multimedia eLearning documents enjoy many advantages over other less sophisticated documents. Such advantages include being concise, attractive, and the topics covered are usually well presented. Besides, they can be compressed at a much higher ratio than regular videos due to the animations that

they include. However, due the high cost of developing a true multimedia document and the very lengthy process to produce one online course, they have forced many to focus on opportunities using AOF and videotaping approaches.

2.1. Videotaping

Videotaping is the easiest way to generate content for any eLearning platform. Traditional videotaping is as simple as putting one or more cameras in a classroom to record the lecture. During a traditional regular lecture, the instructor mainly does nothing extra. The instructor may either ignore the presence of the camera or deal with it as a yet another silent student. In a classroom, the instructor may do one or more of the following at the same time:

1. Reading or Explaining something written on the board or projected on the screen
2. Writing on the board
3. Speaking to students or Listening to their questions

When using one camera, there are three options to use in order to capture all the notes written by the instructor on the board and keep them readable. The first option is to use a small board that would represent the video frame boundaries. The other option is to hire a camera man during the lecture who knows what, how, and when to capture the notes (samples of such an approach can be spotted easily by MIT Open Course Ware system. Figure 2 shows a snapshot from a lecture conducted by Prof. Eric Grimsona and Prof. John Guttag from MIT-OCW).

2: Branching, Conditionals, and Iteration



Figure 2: MIT Open Course Ware system

While the last option is to use a camera with motion detection and ask the instructor to keep paying attention to it, so the instructor would stay next to the notes or area of the board being referred to. However, this may require the instructor to move back and forth often and quickly, thus, deviating from the main purpose of catching the notes clearly.

On the other hand, when using more than one camera, there are just two options. The first is to do an A/B mixing during the lecture, while the other option is do it after the lecture. In both cases, you will need some human intervention other than the instructor. Samples of such approach can be found within the NPTEL program [5].

The drawbacks of such an approach are two folds. Firstly, you need a human intervention in most of the scenarios discussed above. In other words, you need a camera man or a person to do the video mixing from the various cameras inputs. Secondly, the length and number of videotaped lectures to cover a complete course are usually large. This is due to the following reasons:

1. In Videotaped lectures, some of the topics scheduled to be covered in a certain lecture may get postponed whenever an attending student asks the instructor to repeat a certain topic.
2. The speed the information being delivered by the instructor at may vary based on the instructor's sense of the way the students perceive the information.
3. Unlike in eLectures, in videotaped regular lectures, the instructor usually takes some time to interact with the students as a kind of Keep-Alive messages, such as throwing questions and hearing answers.

In the case when the instructor uses Softnotes (notes being added or written on slides through a computer) rather than Hardnotes (notes written on a board), AOF becomes a better approach to use to generate online lectures. AOU for instant uses mainly Elluminate Live system [6] (kind of online conferencing system) to deliver and record their online classes.

2.2. Authoring on the Fly

Although videotaping a traditional lecture can also be called AOF, but AOF usually refers to a little more sophisticated output which may include more than one region within the output video frame. The simplest output that can be composed from the content of the videotaping approach is to assign one video region to the camera focused on the board, and another region to the camera with a wide shot showing the instructor movements and gesturing. Furthermore, certain positions of the video are then bookmarked and labeled according to the topic being discussed. Consequently, a region that displays the list of the bookmarks will also be included. Examples of such approach are those listed in [4]. Figure 1 shows a sample of such a lecture taken from [4] with the slide show screen used rather than a board.

It is worth to mention that although the output in this example (Figure 1) may look like a complete video played by RealPlayer [7], but in fact, only the video region within the output frame is the real video; while the other regions may be images, scripts, or hyperlinks integrated together using a scripting language, such as SMIL [8]. The main reason for this was to reduce the bandwidth needed. For that, only discrete snapshots of the softnotes or hardnotes, were added. However, we believe that showing discrete or even significant shots of the notes being written by the instructor rather than continuous video of the notes while being added gradually is harder on the students to follow.

Another example of AOF is classroom 2000 [9]. Soft or hard notes written on the white board are being taken, power point slides are being copied, and both the video and audio from the class are being recorded. Such output includes three to four regions within the output video frame.

3. Merged Approach

In this paper, we will distinguish between the following three terms of online lectures: Videotaped lectures, AOF lectures, and eLectures. Videotaped lectures represent the output of the in-class cameras recording a complete lecture. AOF lectures, are the output of AOF tools when the lecture is delivered before some students (even if remotely, such as using Elluminate Live). In many cases, it is basically a Videotaped lecture that is integrated with notes generated by some form of video processing or electronic boards. While eLectures are those lectures that are generated by AOF tools when the lecture is delivered *outside* a class room, i.e. with the absence of students. In the next section, more elaboration on eLectures will be presented.

3.1. eLectures

eLectures generated outside the lecture rooms are supposed to be more concise than videotaped lectures. Storage and bandwidth needed for a complete course using videotaped lectures are magnitude more in size than if saved as eLectures. At our university for instance, it is found that a twenty minute eLecture is worth up to five hours of videotaped lectures. Also, when comparing with AOF lecture of same duration, although the AOF lectures need much less storage capacity and bandwidth (as mentioned earlier in paragraph 2 of section 2.2), but the same argument about not being concise if compared to eLectures still holds for AOF lectures as well.

For Distance Learning students, videotaped lectures or even AOF lectures may be more beneficent than

eLectures, especially if watched for the first time. For other students who have prior knowledge about the topic, or have watched the video before, it might be boring or time consuming to use them to review the topic or to search through for a particular subtopic.

While the overlap between the videotaping and AOF approaches are high, this paper describes a model that uses bits and pieces from both to generate eLectures. In this model, the instructor speaks before a camera with the absence of students. He/She uses a tablet PC, or a laptop with a pen-like mouse, to replace the white board or the screen. This model is used by some institutes, such as the some lectures from NPTEL [5], in which the video is being switched manually at certain times between a close-up view of the instructor and the laptop screen with the Softnotes being added. Figure 3 shows a snapshot of an eLecture by Dr. Chitralekha Mahanta from IIT. At some other instant, you may see a close-up view of Dr. Mahanta in Full-screen mode. Thus, the video shows either the screen of her laptop or a close up view of the instructor speaking to the camera.

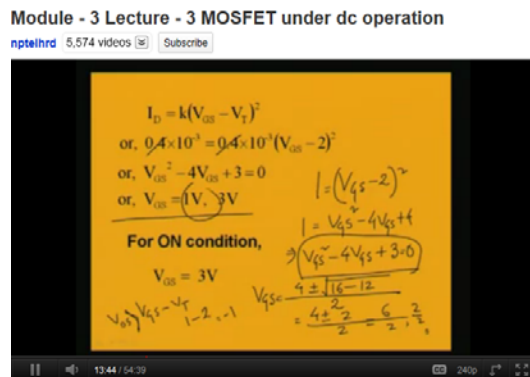


Figure 3: Lecture by Dr. Chitralekha Mahanta from IIT

Using this approach, we may have two different forms based on the location and size of the video window showing the instructor. The forms are as follows:

1. A Full-screen switch between both windows: Close-up view of the instructor and the slide show screen, as in the Figure 3.
2. A small window of the instructor continuously displayed at one of the corners of the screen.

3.2. The Instructor in the Video

The most desired components of any output of both approaches mentioned in the previous section are as follows:

1. Clear audio (at least of the instructor)
2. Clear and continuous view of the Notes; whether soft or hard.

3. View of the instructor; either a close up or wide

While the first two components are agreed upon by all as a must for online lectures, you may find a group of people who would debate the third component. They would claim that seeing the instructor adds no value; or more politely, they would claim that the value added by seeing the instructor does not worth the extra storage capacity and bandwidth that they would need. Fortunately, with the cheaper and faster bandwidth and larger storage capacities, it is becoming easier for such a group to be convinced otherwise. Overcoming this obstacle, the issue of a close-up versus wide view of the instructor becomes debatable.

You may find a group of students who would prefer a close up view of the instructor, while others would prefer a wide view. In [4], a survey was performed on a group of students to find their preference towards the existence of such a window. The samples of eLectures that were displayed to them were A, B, and C as follows:

- A. Slides and notes only
- B. Slides and notes with a small window showing a wide view of the instructor at one corner
- C. Slides and notes with a small window showing a close-up view of the instructor at one corner

It was found that the students were better focused with the third sample where a close-up view of the instructor was displayed at all times at one corner of the video window, along with the slides and lecture notes. Figure 4 shows a snapshot of the third sample.



Figure 4 Notes, slides and close up view taken from [4]

Furthermore, we believe that switching the Full-screen mode between the screen and the instructor acts as a Keep-Alive message with the students. Thus, it helps them stay more focused. For that, we have adopted the mixed format between Full-screen switch form and the corner form. However, for this eLecture model, a camera man or a director who would decide when to do a video switch is needed. To overcome this problem, we have developed an automated tool for the instructor to use which will

generate the desired eLecture without the need for another person. The flow chart of the switching algorithm used by the tool is depicted in Figure 5.

As shown in the flow chart of Figure 5, once the system starts or upon a mouse idle time event for at least T_i (i for idle) seconds, the focus (or full-screen) will be granted to the window showing the instructor. The close-up view of the instructor should last for at least T_t (t for teacher) seconds. The focus will be granted to the slide-show whenever there is a mouse significant move, or a mouse button being click which indicates that a softnote is about to be written. The focus on the slide-show should last for at least T_s (s for screen) seconds, and so on. With this simple strategy, it is believed the students will stay focused and the softnotes will not be missed. Note that even when slide-show screen is in Full-screen mode, the close-up view of the instructor is placed in a small window on one corner of the full-screen.

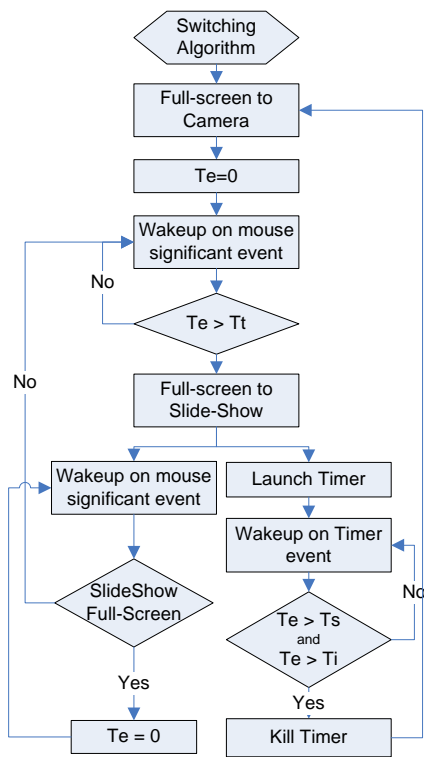


Figure 5 Flow chart of the mixing algorithm

3.3 Auto Videotaped Lectures

Many instructors are not used to give lectures with the absence of students. In fact, they prefer to produce long Videotaped lectures rather than composing more concise eLectures. Thus, after all what has been said on eLectures,

videotaped lectures are still considered a fast way to produce lectures online among the vast majority of instructors, as there is no need for much preparation. However, as mentioned earlier, the camera man inside the classroom must know what, how, and when to focus on the instructor or the board. In this section, we propose a way to auto zoom and auto pan the camera in response to the instructor's normal sign language captured by the camera.

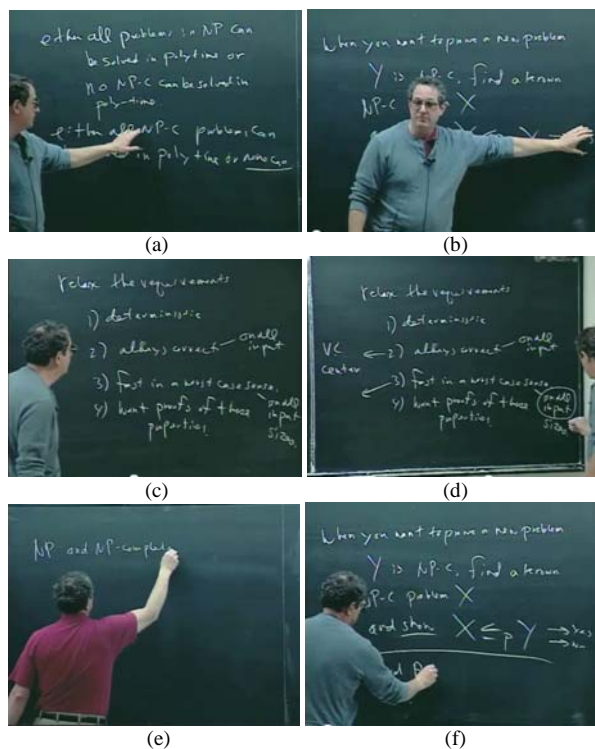


Figure 6 Writing and Explaining

To understand the proposed algorithm, let's first classify the behaviors of the instructor. The Reading, Writing and Speaking behaviors of the instructors mentioned in section 2.1 are illustrated in Figure 6 a to f which shows Prof. Dan Gusfield from UC Davis in some Youtube videos produced by UC Davis [10]. Figures a to c, shows the instructor when Reading from the board in various positions including: sideways, facing audience, and facing the board, respectively. In both figures a and b, the instructor is using his hand to point at the text he is referring to.



Figure 7 Speaking

Figures d to f, show the instructor while writing on the board. Finally, Figure 7a and 7.b show the instructor while speaking to the audience. The major signs to look for in these figures are as follows:

1. An extended hand
2. A face
3. The absence of motion (instructor disappears from the frame)

To detect these from the video, edge and facial detection algorithms are applied. For instance, for edge detection, there are many ways to do so. However, the most may be grouped into two categories: gradient and Laplacian.

The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. A pixel location is On or declared an edge location if the value of the gradient exceeds some threshold. Edges will have higher pixel intensity values than those surrounding it. So, once a threshold is set, you can compare the gradient value to the threshold value and detect an edge whenever the threshold is exceeded.

Furthermore, when the first derivative is at a maximum, the second derivative is zero. As a result, another alternative to finding the location of an edge is to locate the zeros in the second derivative. This method is known as the Laplacian. A zero-crossing edge operator was originally proposed by Marr and Hildreth [11]. A method based on the gradient method could be found in [12] while another on the Laplacian method in [11]. Hand detection is based on both edge and motion detection. Hand is simply considered as the non-vertical block that is on high motion. Yet, an overview on existing Facial and Hand detection algorithms is beyond the scope of this paper.

The ideal scenario to record a lecture is to use a wide board that may be split by vertical straight lines into segments of around 1.5 meters wide each, as shown in Figure 9. The camera is mounted on a motor and calibrated based on these straight lines. The maximum zoom is set to cover one complete segment. While the minimum zoom is set to cover the whole board.

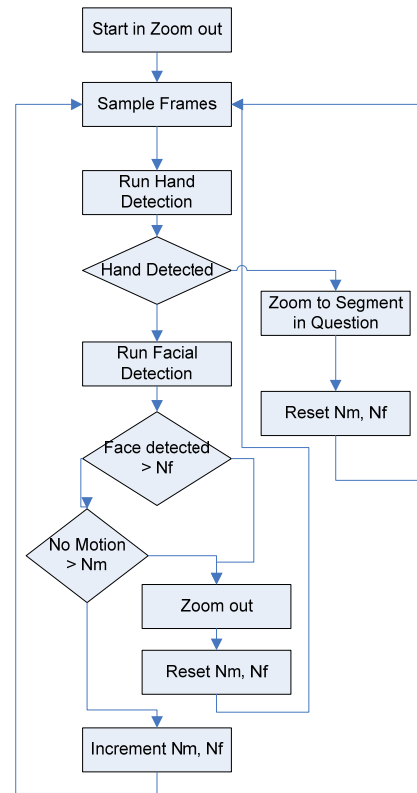


Figure 8 Second Mechanism Flow Chart

The basic camera actions to consider are zooming and panning. To understand the signs that trigger these actions, consider the follow chart depicted in Figure 8.

Frames that are periodically sampled from the video stream are processed for all three signs: Hand detection, Facial Detection and absence of motion. The Zoom out action occurs when either the instructor disappears (absence of motion) for duration $> N_m$ or the face of the instructor shows up continuously for a duration $> N_f$. These parameters refer to the number of sampled frames.

The zoom and panning actions occur when the hand sign appears focusing on the segment of which the hand is pointing at. Thus, with the existence of an extended hand, the facial detection becomes secondary.

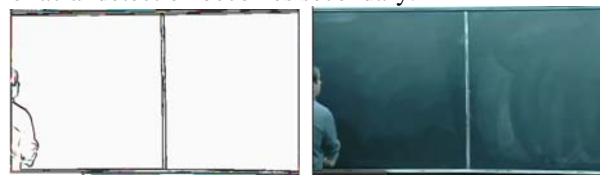


Figure 9 Segmenting the Wide Board by Lines

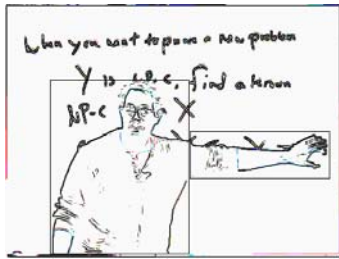


Figure 10 Hand Detection

4. Analysis And Survey

Few regular lectures were videotaped and in most cases the camera motor responded to the correct signs. Unlike the Facial and Motion detection algorithms used, the Hand detection was the weakest. Figure 10 shows a successful hand detection illustration to that of Figure 6.b. Cases in which the hand detection algorithm missed were mostly similar to that represented in Figure 6.d. Fortunately, the proposed mechanism does not zoom out or pan easily. Missing hand detection does not affect as much as detecting false ones; and these are rare. Enhancing Hands detection algorithm is left for future work.

A different, yet more focused survey similar to that in [4] was performed. A group of students from different fields, namely Business, and IT, were chosen to select which among the three different models of online lectures would help them to:

1. Stay more focused when watching the lecture for the first time
2. Watch the lecture again for the purpose of reviewing the material or searching through for a specific subtopic.

The three models presented to the students were: a videotaped lecture similar to that of MIT OCW, and second and third models are eLectures similar to that of Dr. Chitralkha Mahanta [5] that switches between the instructor view and the slide-show screen. Both the second and the third differ in the instructor view: the second with a close-up view of the instructor while the third with a wide view. The number of randomly selected students was ten from each field. Each group was given two samples for each model. The first was in the same domain as the students', while the other sample was in a different domain than theirs. The duration of each lecture was about 15 minutes long (the VideoTaped was a little longer). In summary, the results were biased towards eLectures with the close-up view of the instructor.

5. Conclusion

Content development process has become the most challenging task in implementing eLearning. The process outcome ranges from being very sophisticated and expensive, such as producing interactive multimedia documents where animation, quizzes, and multiple scenarios exist, to simple videotaped lectures. This paper argues that with the faster home internet speed and the availability of higher bandwidth and storage, videotaped lectures are becoming more popular. However, the drawbacks of these lectures were discussed in this paper. They included their lengths and the need for a camera man and/or a director to do video editing and mixing to produce them. In simple words, a 45 hours course will need about 45 hours of video. Consequently, the paper presented an approach similar to that being used by some institutes to produce eLectures that is more concise in length. A course with 45 contact hours may need less than 4 hours of eLectures. Moreover, the paper proposes a simple way to automate the video mixing process needed for this approach to keep the students more focused, as being claimed by the results of the survey being performed on a group of students. The simplicity is the main feature of this proposal. Therefore, by adopting this approach, more eLectures will be available in less time at a lower price than what most institutes would expect.

Acknowledgments

The author wishes to acknowledge the financial support received from Applied Science University for publishing this paper.

References

- [1] Adobe Captivate 5.5 www.adobe.com/Captivate
- [2] Ottmann, Th. & Müller, R. (2000). The "Authoring on the Fly"-System for Automated Recording and Replay of (Tele)presentations. ACM/Springer Multimedia Systems Journal, Special Issue on "Multimedia Authoring and Presentation Techniques", Vol. 8, No. 3.
- [3] N Hirzallah, "Developing eLearning Content Considering Various Video Scenarios" The Asian Conference on Education 2009 Osaka Japan, October 24-25 2009.
- [4] N Hirzallah, W Albalawi, A. Kayed, S. Nusair "A simple algorithm to enrich eLectures with instructor notes" ACM Multimedia Tools and Applications: Volume 49, Issue 2 (August 2010), Page 259.
- [5] National Programme on Technology Enhanced Learning <http://nptel.iitm.ac.in/>
- [6] Elluminate Live , blackboard Collaborate™, <http://www.blackboard.com/Platforms/Collaborate/Overview.aspx>
- [7] RealPlayer application, www.real.com

- [8] Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation, 15-June-1998 (<http://www.w3.org/TR/REC-smil/>)
- [9] Gregory D. Abowd: "Classroom 2000 - An Experiment with the Instrumentation of a Living Educational Environment", IBM Systems Journal, Special issue on Pervasive Computing, Volume 38, Number 4, pp. 508-530, October 1999.
- [10] professor Dan Gusfield from UC Davis computer science on Youtube http://www.youtube.com/watch?v=MZSMcspzj_E
- [11] Marr, D., and Hildreth, E. "Theory of Edge Detection," Proceedings of the Royal Society London 207 (1980) 187-217. A.B. Smith, C.D. Jones, and E.F. Roberts, "Article Title", Journal, Publisher, Location, Date, pp. 1-10.
- [12] [12] A. K. Cherri and M. A. Karim, "Optical symbolic substitution: edge detection using Prewitt, Sobel, and Roberts operators," Appl. Opt. 28, 4644- (1989)

Nael Hirzallah Dr. Nael Hirzallah has finished his Masters and PhD Degrees from the University of Ottawa, Canada, in 1993 and 1997, respectively. He then worked in the industry in leading software and semiconductor companies in both Canada and USA. He joined Applied Science University in Jordan in 2002 to work as an Assistant Professor in the faculty of IT. In 2007 he was promoted to Associate. He became its Computer Center Director in 2003. For three years starting 2008, he worked as a department chairman and Acting Dean in Fahad Bin Sultan University in Tabuk, KSA. Since 2011, he returned to Applied Science University to assume the position of the Faculty of Information Technology Dean. His research interests are in eLearning, eCommerce, Multimedia and image processing.