Proceedings of the 5th WSEAS Int. Conf. on DISTANCE LEARNING AND WEB ENGINEERING, Corfu, Greece, August 23-25, 2005 (pp77-82)

Enhancing the Quality of Video in a Synchronous E-Learning Tool by Taking into Account the Concepts of Change-Blindness

ARUNA THAKUR and BENN LINES
Network Research Group
University of Plymouth
Plymouth
UNITED KINGDOM
aruna@jack.see.plymouth.ac.uk; b.lines@plymouth.ac.uk
PAUL REYNOLDS
Orange SA,
UNITED KINGDOM
paul.reynolds@orange.co.uk

Abstract: - Change blindness is the inability to note large changes in two similar photos or video sequences. In the last decade research carried out to understand this phenomenon has given a new understanding of human perceptual system. This understanding can help in solving some of the problems associated with the transmission of video in a synchronous e-learning tool. The paper describes (a) cinematic principles in recording lectures; (b) the concept of change blindness; (c) the ways its understanding can help in making the presentation of video more engrossing; (d) the guidelines it can provide to remove redundant data and enhance the quality of important sections of the video frame so as to improve the transmission of video; and (e) proposed scheme to stream video sequence having optimum perceptual quality by dynamically controlling frame rate and resolution, taking into considerations the concepts of change blindness.

Key-Words: - change blindness, cinematic principles, background, foreground, separation, perceptual quality, video, transmission

1 Introduction

The distance education is a cost-effective way of providing education to students in their own home or office. It is especially suited to students who are located in geographically distant places and who cannot attend classes due to distance or for other reasons. It is becoming popular among the people who are working and are in need of the further development of their skills but cannot take out time to attend regular classes.

The lectures can be classified into two different categories based upon the interaction between the instructor and students, these are *synchronous* and *asynchronous*. In synchronous e-learning the instruction involves real-time delivery of audio, video, and other media. In asynchronous e-learning educational materials are stored in server and students retrieve them at a time convenient to themselves. Due to relative ease in conducting the course work, asynchronous methods have become quite common, but with the improvements in the performance of videoconferencing tools, more and more colleges are experimenting with synchronous methods. [1].

In spite of its growing popularity, the use of video in synchronous e-learning suffers from a number of problems. Firstly, in most of the cases the instructor is forced to sit in front of camera and deliver lecture, hence, the students only see a head-and-shoulder presentation of their instructor. As a result, the instructor is handicapped because he is forced to present his class while seated in front of his camera and he cannot make use of the body language, like facial expressions, eye gaze and gestures to guide the students. Secondly, the poor quality video attracts attention to itself and distracts the students from the material being presented. It also leads to conscious / unconscious discomfort or distress. These deficiencies can result in none or at best limited interaction among participants resulting in boredom and disengagement in students.

To overcome these problems two important issues need to be resolved:

(a) Presentation issues. The instruction video should be prepared keeping in mind the principles of cinematography so as to excite the interest of students.

(b) Technical issues. This involves the capturing and transmission of video and audio in a quality comparable to TV quality.

2 Cinematic Principles in Recording Lectures

Students attending a classroom lecture experience a rich visual experience that is simultaneous, detailed and coherent. That is, if we take a panoramic view of the class we see a number of events at the same time, in high resolution (or high density of information) and all the sections of the image are combined in a proper jigsaw puzzle. But is it really so? Current research in human visual system shows that this is not the case. The human eye can only see one section of the scene at a time - either in detail or in a coherent (stable) way. We cannot see more than one object in detail and in a coherent way at the same time; and also we can focus attention only on one event at a time [1].

For many years film-makers have informally applied the knowledge of these limitations of human perception to make "continuous" and engaging movies. People watching those films do not complain about the limitations of the perception, on the contrary they find the presentation much more appealing than the real life. If that is the case, then the reason for dull e-learning video lectures is not the limitation due to the use of video for teaching but the actual limitation arises due to the style of presentation. Hence, efforts should be directed towards making video presentation according to the principles by which humans perceive the real world. By applying the knowledge of strength and weaknesses of human perception we can work towards making the lectures more exciting and captivating

Unfortunately, not many instructors are trained to be good presenters and video-graphers. Most often the instructor just sits in front of a monitor with a camera placed over it and delivers the whole lecture without any movement. Unfortunately this mere capturing of talking-head-and-shoulder video does not make an interesting video. What is more interesting than this is a "real lecture" video, where the lecturer has the freedom to move around, use electronic whiteboard and show real models as he / she would do in a real class room environment.

Some of these principles of film making can be adopted for making engaging lecture videos, they are:

(a) Use variety of shots from different angles and views. A long duration shot taken from the same angle and view is boring to the audience. Hence, shots of the same object should be taken from different locations and using different zoom. Also each shot should not be of less than 6 seconds and not more than 20 seconds.
(b) Follow a story line. The role of an instructor is to convey the material in an orderly manner. In a real classroom if this is not done then the instructor can be interrupted and any confusion addressed; on the other hand, such possibilities are far less in the case of e-learning. Hence, the sequence of presentation should be well thought out right at the beginning.
(c) Intelligent use of the tools to present the course material. In a real class, students can move around to get proper view of the demonstrations, but the students of e-learning have to rely on the instructor to show them the demonstrations clearly on the screen. Hence, it becomes the responsibility of the instructor to take shots from the appropriate angles and use zoom techniques so to make the object of demonstration clear to the students.
(d) The continuity of the scene. While it is important that the continuity of the scene be maintained, it has been observed that small continuity mistakes are often overlooked by the audiences.

3 Change Blindness

During the making of a movie different shots are taken at different times and settings. Hence, in spite of all the efforts some continuity mistakes do creep in. In fact there are web pages dedicated to listing these mistakes in popular movies. One of the sites lists the best three continuity mistakes as [2]:

(a) Commando: The yellow Porsche is totally wrecked on the left side, until Arnold drives it away, and it is fine;
(b) Spider-Man: In the scene where Mary Jane is being mugged by four men, Spider-Man throws two of the men into two windows behind Mary Jane. Then the camera goes back to Spider-Man beating up the other two guys. When the camera goes back to Mary Jane the two windows are intact.
(c) Terminator 3: Rise of the Machines: In the scene where John and Catherine are in the hangar at the runway, the Cessna's tail number is N3035C. When the plane is shown in the air, the number is N9373F. When they land, the tail number has changed back to N3035C.

These are quite glaring mistakes but how many viewers actually noticed them?

For long researchers have noted that people are quite poor at noticing such changes, but in recent years this subject has renewed considerable interest among researchers and has been researched as "change-blindness".

According to Simons and Rensink [3], *Change blindness is the striking failure to see large changes that normally would be noticed easily.* It is caused when the change is separated by a disturbance. At times even a large and repeatedly made change can go unnoticed for a long time. The disturbance between the scenes may be natural, like the eye movement (saccades) and eye blinks or induced artificially by image flicker, brief "splats" that do not actually cover the region of change. In the case of a movie sequence the disturbance can be caused by a sudden or gradual change in scene. Once the difference has been spotted the change, it becomes very obvious and eye catching [1,3,4].

The research carried out during the past decade suggests that there are a number of important factors that can cause change blindness.

(a) Attention: Change blindness results when the change is unable to attract attention towards itself. This happens more so when the attention is not paid during observation; but in spite of the attention change may often go unnoticed until unless the attention is directed towards the change, which can be aided by a number of ways like using strong or unusual colour, high contrast or verbal clues [3].

(b) Expectation: Change blindness is strong when the change is unexpected; this is because the attention is not directed towards the region when the change occurs [3].

(c) Background or foreground: Changes to the central item are detected more readily than the changes to the background even if the changes are of equal salience. This is because in general the central item receives more attention as compared to the background [4,5]

(d) Number of changes: Humans have the capability to catch only one change at a time; if more than one change occurs then the change will be noticed one after the other and not at the same time.

(e) Type of Change: It has been shown that the changes to location is more difficult to detect than the changes to the identity, where identity is defined as the objects features other than its position [6,7].

(f) Grouping of objects: It has been known for a long time that the elements are grouped together to form units. It is these units and not individual elements that are further processed for recognition, search, etc. As a result the individual elements lose their own identity [6]. Hence, the memory for the unit is quite good but memory for the individual elements that comprise the unit is poorer. Since the processing is done at the unit level and not at the element level, it has been observed that it is easier to detect changes in visual stimuli that are strongly grouped [6,7].

A number of models have been suggested to define human perceptual system and explain the existence of change-blindness. The complexity of the task means that none of these models are able to explain all the observations convincingly.

While change blindness has attracted considerable interest among the psychologists, it has gone largely unnoticed by computer scientists. In fact it opens a large unexploited domain of study. In this paper we have tried to explain how the understanding of change blindness can help in designing e-learning courses and the work that we are carrying out to use the understanding of change blindness to develop an efficient system for transmission of video.

4 Lessons for e-Learning

The research on change-blindness raises some interesting prospects for improving the quality of e-learning video both in terms of presentation and transmission.

4.1 Presentation

In a lecture, instructor has to continuously work hard to hold the interest of students. This issue becomes even more important in case of e-learning where the options available to instructor are limited. The knowledge of change-blindness can help the instructor in designing the lecture so that he/she is able to hold the attention of students and make sure that they do not lose track of the subject.

Change blindness can play a positive or a negative role in the students' attention. On one hand, due to small attention span, the students' attention can get momentarily diverted away from the object of interest and then due to change-blindness they may forget the earlier scene, thereby resulting in the loss of context. In such a situation the teacher should try to direct the attention of students towards the object of interest by making gestures and sending video sequence taken at the correct angle and zoom of the appropriate object at appropriate time. On the other hand, it can be used purposefully, as done by the film makers and conjurers for a long time, so as to make a well connected presentation and prevent attention getting focused on objects of no interest.

Unfortunately, it is not possible for every instructor to be good film-makers. Hence, there is a need for automating the "art" of film-making and this is possible only within a limited domain. The first important step to make the presentation captivating is to break out of the head-and-shoulder mould and present the lecture in a "real" classroom setting but without any interference. In this condition, multiple camera system can take shots from different angles and using different zooms. A combination of fixed and Pan/Tilt/Zoom (PTZ) cameras can cover the lecture dais quite effectively. Once the shots from different angles have been obtained, the second step is to automatically transmit appropriate shot at the appropriate time so as to keep the attention of students focused on the course material and prevent it from wandering away. The transition between the cameras can be done by a number of ways, like manual use of switches, use of gestures, etc. While changing the scene, the transition should be as smooth as possible.

The presentation can be improved further by using a few camera tricks like:

- Use of a very short different shot between two sequences to help in separating different but similar scenes. The very short duration shot does not register in the brain, but it helps to erase the previous scene from the mind, almost like the flicker paradigm. This will help in focusing the attention on two different objects without interference.
- Once an object has been demonstrated it should be removed from the stage while the camera is not focused on it. Change blindness will cause "out of sight, out of mind", and its removal will eliminate an object that could potentially be a cause for wandering of attention.

The concept of change blindness can be used in many such ways to improve the quality of presentation.

4.2 Transmission

In e-learning the students often attend the class sitting at their home and are most often linked with very limited bandwidth of up to 60 kbps. Hence, the challenge is to send an acceptable quality of video within that limitation. Significant amount of work has been done in this direction, but inclusion of the concepts of human perceptual model provides a sound basis for further improvements.

When we look at our surrounding we "feel" as if we see a detailed coherent world, and we try to represent this in our video. In actual life this is not so. At a particular instance, we only see either the gist of the panorama or the detailed view of a small part of it. At no instance of time do we have detailed representation of the whole scene. We can apply the same principles for transmission of video.

The quality of video transmission can be significantly improved by reducing the information of the background and enhancing the information of the foreground. Extensive amount of work is being done in separation of foreground and background. This feature is required for various purposes like, tracking of moving object, face recognition, security surveillance, film making, etc. All these applications have different requirements on the quality of separation. In case of transmission, this feature is used to reduce the load on network and has good potential for the use in the transmission of e-learning video, since the students normally have access to limited bandwidth.

In e-learning the separation can be guided by the knowledge of change blindness. Based on the earlier discussion, it is reasonable to expect that the attention of the students would normally be directed towards:

(a) instructor,
(b) blackboard / whiteboard / projection, and
(c) any model or object of interest.

Hence, the segments of the frame covering these objects should be sent preferentially over other parts of the frame. Object of low interest or ungrouped objects like – pointers, etc need not be given much attention and the slight discrepancies in the position of these objects would go unnoticed.

The transmission of important and unimportant sections can be done in various ways, for example:

- The foreground can be sent at high resolution and the background at low resolution;
- The foreground can be sent more frequently as compared to the background;
- In case of panorama shots, high resolution image of the dais can be sent right at the start of the lecture; later on only the foreground can be updated.

5. Aim and Scope of Our Work

In our project we are developing a system for automating the production and transmission of real time video for synchronous e-learning in an interesting and cost-effective manner. In this paper we have only presented the concepts of our ongoing research work. It has not been completely implemented and we are aware that modifications may be required as we progress with its development. There are basically three parts of our work:

(a) Use of multiple cameras in order to break the monotony of head-and-shoulder shots and to show the object from the appropriate angle and zoom.
(b) Save on bandwidth by removing the redundant features of the frames.
(c) Transmit video within the available bandwidth in the perceptually optimal way.

5.1 Use of Multiple Cameras and Automatic Selection of Shots

For the first part, we have proposed a three camera system; it is easy to add more cameras on the same principle. Camera 1 is a fixed camera covering the panoramic view that shows the entire lecture dais from which the instructor presents his lectures. Camera 2 is a tracking PTZ (Pan/Tilt/Zoom) camera to take close-up shot of the instructor's head and shoulder. Camera 3 is a tracking PTZ camera to take mid-shot so as to show the instructor and the region of interest next to the instructor, for example the white board, model, etc. [8].

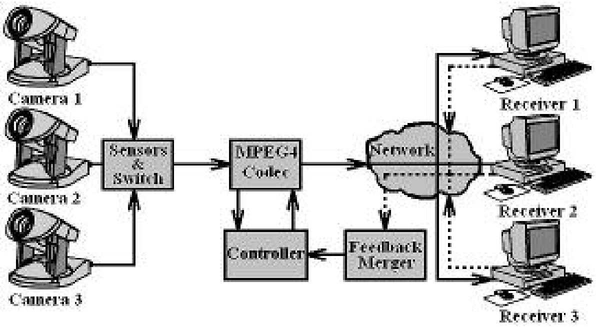

Figure 1. Block diagram of the architecture.

5.2 Separation of Foreground

As discussed earlier there are a number of aspects of background separation. A single algorithm cannot cover all the aspects. What is required is an integrated system to cover all the three cases. We are currently working on the development of this system. Primarily it involves the following:

(a) Separation of foreground, where there is a relative movement between the foreground and background. This has two sub-cases: (i) Where the foreground is moving and the background is stationary; and (ii) Where both foreground and background are moving. In the first case, first the background is grabbed without any foreground. In order to account for the slight variations among different frames arising due to lighting, camera operation, etc. mean and standard deviation of every pixel is taken for a number of frames. The variations normally follow the standard Gaussian curve. A threshold of variations is then decided. When the lecture begins, the ongoing frame is compared pixel by pixel against the mean background frame, and a pixel whose value is out of the set threshold is taken as the foreground. The second case in more complex and a few algorithms have been developed to separate the two. Further work needs to be done in this case.
(b) Foreground as represented by objects like blackboard. In this case the region of foreground is fixed and there may or may not be changes to it. In the first step, the outline of the object is detected using the edge detection algorithms and then the region of foreground is decided.

At the beginning of the session, these background images will be stored both at the sender and receivers ends. During the lecture, the video will be grabbed and the foreground separated from the frames. The foregrounds so generated will be transmitted to the receivers. At the receiver end the foregrounds will be merged with appropriate backgrounds and shown to the receiver. Due to change blindness the receivers may not be able to catch slight discrepancies arising out of the combination of foreground with a general background.

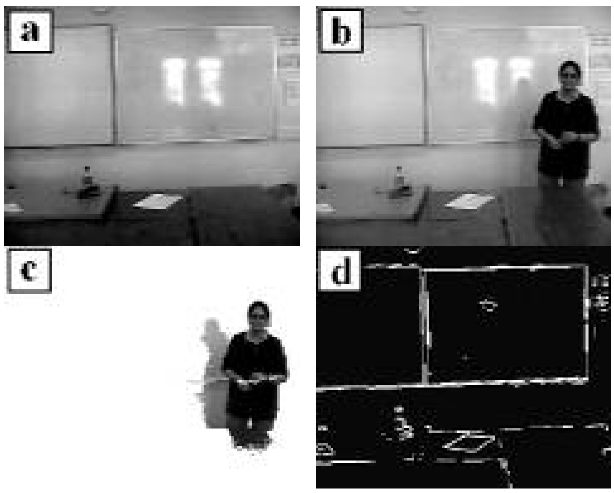

Figure 2. (a) Background; (b) Panorama view of a class; (c) Separation of foreground from background; and (d) Identification of edges using Sobel filter.

5.3 Optimum Transmission of Video

During the transmission of video the fluctuation in the quality of video will be mostly due to: (a) changes in video scenes, and (b) random variations in the network conditions. Hence, there is a need for a dynamic adaptive procedure so as to provide consistently good quality video in real-time.

We have developed a controller to dynamically adapt to the variations by adjusting the frame rate and spatial resolution [8]. In our architecture (Figure 1), the video signals from cameras are grabbed and encoded using an open source-code MPEG4 codec on Linux platform. The encoded video is then multicast via RTP/UDP/IP stack to establish on-line interactions between server (instructor) and clients (students). A stream from encoder is sent to a feed-forward controller to [8]:

(a) Decide the required bitrate using the information from feedback merger that collects RTCP feedbacks.
(b) Dynamically characterize the nature of video sequence (Temporal Information and Spatial Information).
(c) Use linear prediction for predicting the complexity of the incoming movie sequence.
(d) Calculate the options available (frame rate and resolution) that would satisfy the availability of the bandwidth.
(e) Decide the optimum way of configuration (frame rate and resolution) of the codec for the incoming video sequence based on the correlation between the Bandwidth – Frame rate – Resolution – Perceptual quality [9].
(f) Send the optimized values of configuration parameters to the encoder to configure the codec for coming scene.

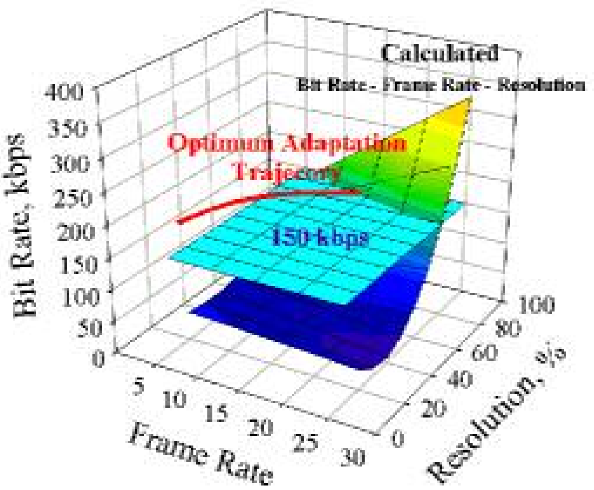

Figure 3. An example of the working of the controller. For 150 kbps it is possible to have 25 options. Out of these 25 options there is one condition Frame Rate = 14 fps, Resolution = 89% which will give the best perceptual quality.

8. Conclusions and Future Work

In this paper we have discussed the concept of change blindness and how its understanding can help in improving the presentation and transmission of video used for synchronous e-learning. We have also outlined our approach to implement these concepts in the development of e-learning tool.

References:

[1] R. A. Rensink, Internal vs. External Information in Visual Perception, *Symp. on Smart Graphics*, June 11-13, 2002, Hawthorne, NY, USA
[2] http://www.moviemistakes.com/best_continuity.php
[3] D. J. Simons and R. A. Rensink, Change Blindness: Past, Present, and Future, *Trends in Cognitive Sciences*, Vol.9, No.1, 2005, pp.
[4] M. Turatto, A. Angrilli, V. Mazza, C. Umilta and J. Driver, Looking Without Seeing the Background Change: Electrophysiological Correlates of Change Detection versus Change Blindness, *Cognition*, Vol. 84, 2002, pp. B1–B10.
[5] S. Werner and B. Thies, Is "Change Blindness" Attenuated by Domain-specific Expertise? An Expert–Novices Comparison of Change Detection in Football Images, *Visual Cognition*, 2000, Vol. 7, No. 1/2/3, pp. 163–173.
[6] A. Rich and B. Gillam, Failure to Detect Changes in Color for Lines Rotating in Depth: The Effects of Grouping and Type of Color Change, *Vision Research*, Vol. 40, 2000, pp. 1377–1384.
[7] D. J. Simons and D. T. Levin, Change Blindness, *Trends in Cognitive Sciences*, 1997, , No. 7, pp. 261-267.
[8] A. Thakur, B. Lines and P. Reynolds, Dynamically Adapted Streaming of Video for a Real-Time Multimedia Application, (Submitted).
[9] N. Cranley, L. Murphy and P. Perry, User-perceived quality-aware adaptive delivery of MPEG-4 content, *NOSSDAV'03*, 2003, Monterey, California, 1-3 June, 2003.