Study Element Based Adaptation of Lecture Videos to Mobile Devices

Ganesh Narayana Murthy
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Powai, Mumbai - 400076
Email: nganesh@cse.iitb.ac.in

Sridhar Iyer
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Powai, Mumbai - 400076
Email: sri@cse.iitb.ac.in

Abstract—**With devices like mobile phones and PDAs becoming pervasive especially among students, there is a demand for viewing lecture videos on such devices, to get quick access to educational content. But, being traditionally designed for desktop computers, lecture videos have high video bit-rates, due to which viewing them on low network bandwidth connections like GPRS incurs long delays and high costs.**

In this paper, we propose a novel method for adapting lecture videos, to make them viewable on low bandwidth connections, at low cost. Our idea is to exploit redundancy present in lecture videos, such as non-changing presentation slides. We define study-elements within the video that have varying levels of redundancy and user-expectations. We send images at different intervals, such as one image every five seconds, according to the Study-Element involved along with continuous audio.

In our experiments, we have found that for a video of size 432MB, our method achieved 97% reduction in size, that was better than that achieved by standard codecs for that video. Additionally, our method can be customized according to the level of user experience desired.

I. INTRODUCTION

Many universities offer videos of their courses to their students. With mobile phones and PDAs becoming more and more pervasive among students, they could gain quick access to educational content, if these lecture videos could be accessed on their mobile devices. However, such lecture videos are traditionally designed for a desktop computer, and cannot be viewed on mobile devices, because of their limited capabilities like low network bandwidth, small screen size and less processing power.

Our aim is to adapt lecture videos to suit mobile device capabilities. There are two main challenges or problems that are encountered in this adaptation process:

1) **Network Bandwidth and Cost** - Network connections (like GPRS) that are used by mobile phones have a very low network bandwidth (40kbps). This is very low as compared to the bit rate of a video which is in the range of 400-1200kbps or even higher. Also, such networks usually charge the user, based on the amount of data transferred. Hence, the size of the video must be as small as possible to minimize cost.

2) **Usability of the content** - Lecture videos usually have written material, either as presentation slides or in the writings of the instructor. Hence, the adaptation process must ensure that understanding and visibility of the content must not be compromised.

One way to address these challenges is to use video transcoding[5]. Video transcoding is the process of converting the video from one compression format to another. In the process, video parameters like bit rate, frame rate and resolution can be changed to meet the target device requirements. However, compression formats like H.264[1] that achieve good video quality at low rates, have decoders that are computationally intensive requiring heavy processing power that is unavailable on mobile devices. Further, compressing to very low bit rates degrades the quality for any compression format, and can make the content incomprehensible.

Content aware video adaptation is the process of adapting the video based on its content, so that the content is visible clearly, even at lower bit rates. There have been attempts, to identify important segments of a video [2] and important objects or regions in a video frame sequence[3]. These are then encoded at a high quality and other portions are encoded at a low quality. These methods achieve a descent video quality (i.e. that looks as good as the original one) at moderate bit rates of 100kbps for non-lecture videos, but at very low bit rates of 40kbps and in the context of lecture videos, the quality and visibility of written content is doubtful, as the video quality might degrade further.

One promising method in the context of a lecture video has been presented in [4]. In this method, non-changing portions of a lecture video are identified and one image is extracted from each such region. The output is a slide-show of such images along with the audio stream of the whole lecture. This method basically aims to achieve reduction in bandwidth by eliminating redundancies in the video. We employ the same concept of displaying slide-show of images extracted from the video, but in a different way, as explained below.

We identify *Study Elements* within the video. We define "StudyElement" as the portion of video showing any medium of instruction, like a slide of presentation or an explanation written by a instructor on a white paper or the instructor herself explaining something. We then extract images from each Study-Element at a different intervals; for example one image every 5 secs. The output is the slide-show of the extracted images sent to the client according to the interval

at which they were extracted from the element, along with the audio.

The user can specify the user experience desired for each Study-Element and the network bandwidth available as inputs. The sending interval of images for each Study-Element is decided based on these inputs. To quantify user experience, network bandwidth and size of the adapted video, we have defined metrics for each of these parameters, and we have identified the relation between these parameters and the sending interval, by analyzing videos. Using this relation, we identify the sending interval of images for each Study-Element, with the constraints as user experience and network bandwidth. A higher user experience results in more amount of data being transferred and hence higher cost.

The rest of the paper is organized as follows. In Section II, we present and discuss the notion of Study-Elements. We then describe the main adaptation methodology in Section III. Finally, in Section III-I, we show that our method achieves better reduction in video size compared to standard codecs and is enabling viewing at low network bandwidths.

II. STUDY ELEMENTS

A. About CDEEP Lecture Videos

Center for Distance Education and Engineering Programme(CDEEP) is an distance learning initiative by Indian Institute of Technology Bombay. In addition to live broadcast of lectures, recorded videos of the lectures are also distributed as Video-On-Demand. Our aim was to adapt the lecture videos provided by CDEEP and make them viewable on mobile devices.

The instructors use mainly two modes of instruction namely presentation slides and writing on a white paper. There is a screen in the classroom on which the presentation or the white paper are projected. There are cameras recording the instructor and the screen. When the presentation slides or white paper are captured, the capture is precise to their boundaries and no classroom is shown. The final video is an interleaving of video of the instructor, video of the presentation slides and the video of the white paper.

B. Study Elements

As mentioned above, CDEEP lecture video contains an interleaving of video of presentation slides, video of white paper on which the instructor writes and explains material and video of instructor. We define each one of the above portions of the video as Study-Elements.

We have defined three types of Study-Elements:
1) **Presentation Element** - Portion of video that shows one slide of a presentation
2) **White Paper Element** - Portion of video that shows white paper on which instructor is writing something
3) **Instructor Element** - Portion of video that shows instructor talking

Sample images of the elements are shown in Figure 2

Study elements are portions of video that are different from one another in terms of the following properties:
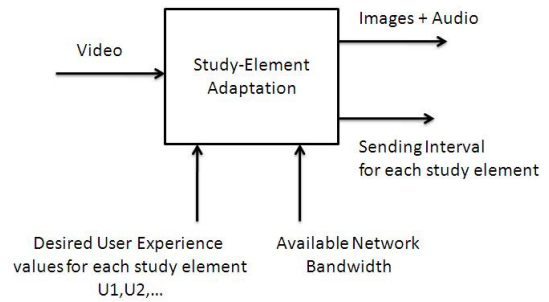

Fig. 1. Block diagram of System

- **Redundancy** - A presentation need not be shown as a video. Instead, it could be replaced by images of the individual slides. Hence, presentation elements that represent every slide of a presentation, have high amount of redundancy when shown as a video.
- **Viewing Requirements** - Portions of the video showing the instructor talking might be replaced by just the audio of the instructor's voice, as what the instructor says is more important. In this case, the viewing requirement of instructor element is just audio and video is redundant. Like this, different Study-Element may have different viewing requirements.

The above two properties of Study-Elements could be exploited to reduce the size of the video and its network bandwidth requirement.

III. STUDY ELEMENT BASED ADAPTATION

Figure 1 shows the block diagram of the system. The adaptation methodology accepts the video as input. It further accepts the desired user experience and the network bandwidth available, as parameters from the user. The desired user experience value for every Study-Element is accepted. For example, the desired user experience values for presentation element and instructor element are taken as input. The output consists of a set of images, sending interval for each Study-Element and continuous audio.

Images are sent to the client device according to the sending interval. For example, if the sending interval is five seconds, then one image is sent every five seconds. The audio is continuously streamed to the client. The user, hence, would see a slide-show of images while hearing audio.
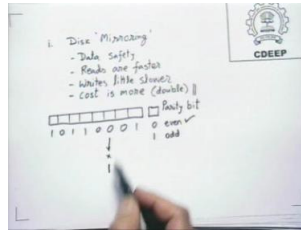
A. Sending Interval

We define it as the time gap between any two images sent consequitively. Images are sent periodically every "sending interval" seconds. For example, if sending interval is five seconds, then one image is sent every five seconds.
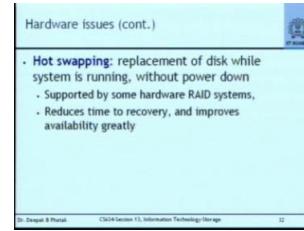
The required network bandwidth depends on the sending interval. If the images are sent at a higher interval, say one image every ten seconds, rather than five seconds, more time is available to send one image size of data, and hence the required network bandwidth is less.

(a) Instructor Element   (b) White Paper Element   (c) Presentation Element

Fig. 2.   Sample Study Elements

User Experience also depends on the sending interval. If the sending interval is low, say one image every five seconds, then the user would see updates faster. Hence, the delay experienced by the user in between the updates is less and the user experience is more and vice-versa.

B. User Experience

We define "User Experience" as a value between 0 and 1 that represents the desired level of user experience. This is given as input by the user to the system. We choose delay experienced by the user as the parameter for defining user experience since we feel that in our system, since images are seen with some delay, the user would be more satisfied if that delay is lesser. Hence, a zero value means that the delay experienced is large (to be chosen by system administrator), and a value of one means that the delay experienced is equal to one second.

The delay to be considered for different Study-Elements is different. For example, for a presentation element(slide), the user would expect the first image of slide quickly. Once it appears, it does not matter if he sees any more image or not, as the slide does not change. Hence the delay in the start of the slide contributes to user experience. But, for a white paper element where the content is changing, the updates have to be faster. Therefore in this case, the delay between any two images is contributes to the user experience.

The user experience value also controls the cost to be beared by the user. As already explained, for user experience to be high, the sending interval has to be low i.e. more number of images have to be sent in a given time. Typically, sending more images means sending more data and hence higher cost in networks like GPRS where the user is charged for the amount of data transferred.

However, we felt that the user should have control over what elements he wants better user experience and hence is ready to bear higher cost, and for what elements a lower experience is acceptable to him and hence expects a lesser cost. To enable this, the system accepts user experience value for each Study-Element.

C. Adaptation Steps

The Study-Element based adaptation method has the following three basic steps:

- Define the boundaries of Study-Elements in the video

- Decide sending interval of the images for all Study-Elements, based on the desired user experience and the available network bandwidth.
- Extract images and audio from the video, at an interval equal to the sending interval. Output the images, audio and the sending interval of images for each Study-Element.

D. Defining Study Element Boundaries

We use manual tagging to define the Study-Element boundaries. The tagging involves defining the following parameters for all instances of a Study-Element:

1) Starting time.
2) Ending time.
3) Type of Study-Element.

For example, for a presentation element, that starts at time 00:05:00 and ends at 00:07:00, the starting time and ending time would be correspondingly entered, and the type would be entered as "Presentation". These details are stored in an XML file.

E. Calculating Sending Intervals

To calculate sending interval, a relation is required between the sending interval, the network bandwidth required to support it, the total size of the images if that sending interval were used, and the user experience at that sending interval. Further, this has to be done for each Study-Element separately. For this, we define metrics that quantify each of the above parameters and which are named User Experience (U), Network Overhead(NO) and Size Overhead(SO).

We extract images from the video at different intervals, called interval of extraction. Then, we find the relation between sending interval, which is same as interval of extraction, and the above parameters by examine a set of ten videos which contain videos from all departments, as explained below.

F. User Experience of Study-Elements

1) Presentation Element: As explained already, for a presentation element (a slide), the delay in the start of slide contributes to user experience as the user expects to see the slide quickly once the instructor has started explaining the slide.

We fix a sending interval 'r' and then find the delay in the start of slide experienced by the user. For this we choose five instances of presentation element (i.e. wherever there is
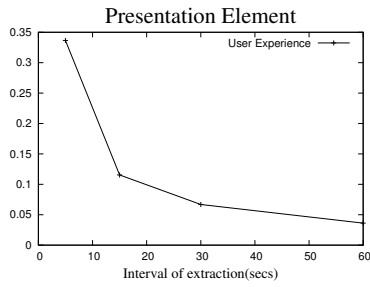
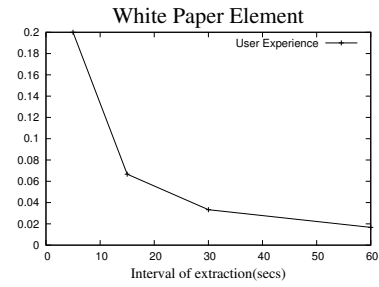Fig. 3.   User Experience of Presentation Element

Fig. 4.   User Experience of white paper element

a slide in the video) and take the average. Also, we assume a minimum sending interval of one second, and hence the minimum delay that can be incurred is one second. Hence, the user experience is calculated as the delay experienced compared to a one second delay.

$$
\begin{aligned}
DelayExperienced(D2) &= \text{Time that user sees the image} \\
&\quad\ \text{of the slide} \\
&- \text{Time that the slide actually} \\
&\quad\ \text{started in the original video} \\
UserExperience(U2) &= \frac{1\,Sec}{Delay\,Experienced}
\end{aligned}
$$

The graph of the relation between U2 and sending interval for presentation element is shown in Figure 3. It is evident that the user experience decreases with increase in the interval of extraction (same as sending interval).

It can be observed that, there is an increase of 83% in user experience when the interval is varied from sixty to thirty, while there is an increase of 192% when it's varied from fifteen to five seconds. This means that at smaller network bandwidths (2 kbps from Fig 5) i.e. larger intervals, the user experience changes slowly for changes in network bandwidth as compared to higher network bandwidths (25 kbps).

Also, notice that the delay experienced at any interval is less than the interval. For example at five seconds, the delay is 1/0.34 = 2.94. This shows that elements are not starting too far from the interval boundary. Hence, we might actually choose a sending interval higher than five seconds, to get a delay of five seconds and hence lower the cost.

2) White Paper Element: As explained already, in a white paper element that shows instructor writing on a white paper, the user expects to see updates faster and hence delay between any two successive images contributes to user experience.

$$
\begin{aligned}
Delay\,Experienced(D1) &= Sending\,interval \\
User\,Experience(U1) &= \frac{1\,Sec}{Delay\,Experienced}
\end{aligned}
$$

The graph of the relation between U1 and sending interval for white paper element is given in Figure 4. The user experience increases by 200% when the interval is changed from sixty to thirty while increases by 232% when the interval

is varied from fifteen to five. Hence, at higher network bandwidths (16 kbps from Fig 5) i.e. lower intervals, small changes in the sending interval gives large changes in user experience, than at lower network bandwidths (2kbps).

3) Instructor Element: For instructor element, we assume that the user would not expect to see images of instructor movement. Hence, we do not send any images from the instructor element. Hence user experience is assumed to be one.

G. Calculating network bandwidth and size

$$
\begin{aligned}
Network\,Overhead(NO) &= \frac{Image\,Size}{Sending\,Rate} \\
Size\,Overhead(SO) &= Total\,Size\,Of\,The\,Images \\
&\quad\ Extracted\,From\,The\,Element
\end{aligned}
$$

Here 'Image Size' refers to the maximum image size of all the images generated from all instances of the Study-Element in the video. For example, out of all the images of white paper element, the maximum image size is considered. SO is found by adding the size of all images that represent a type of element. For example, SO of presentation element for a sending interval 'r' is the total size of all images that show a slide, when images are extracted from the video at that interval 'r'.

The graph of the relation is shown in Figure 5. It can be seen that higher the sending interval, lower the network bandwidth required and lower the size of the adapted video. For white paper element, the network bandwidth required increases by 100% when the interval is changed from sixty to 30 and by 200% when it is increased from fifteen to five. The total size increases by 150% from sixty to thirty and 210% when the interval is changed from fifteen to five. Similar is the behavior of presentation element. Hence, at lower intervals, small changes in interval leads to large changes in network bandwidth and overall size.

H. Finding Output Sending Interval

Given the user experience value for each element and the network bandwidth available to the user as input, finding the appropriate sending interval is straightforward from the graphs of NO vs sending interval and U vs sending interval. However, the available network bandwidth puts an upper limit on the
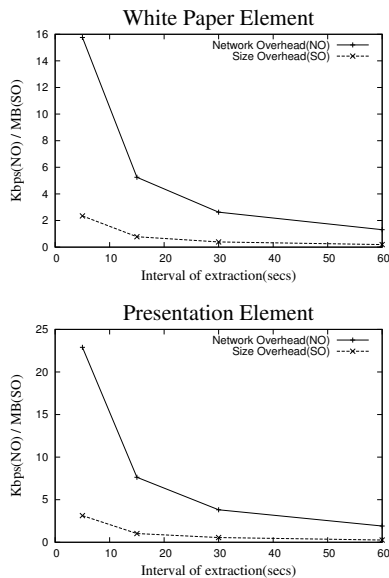
Fig. 5. Network Bandwidth and Size overheads of study-elements

Original Video	
Size(MB)	432
Bit Rate (kbps)	1150
Resolution	352x288
Audio Present	No
Target Video (Study-Element Method)	
Images Size (MB)	2.85
Audio Present	No
Supported Bit Rate (kbps)	20 and above
Image Resolution	320x240
U1	0.2
U2	0.38
Size Reduction	97%
Target Video (Standard Codecs)	
Codec	**Size(MB)**
H.263(3gp)	39
H.264(mp4)	16.9
VP6(flv)	20.5
Image Resolution	320x240
Bit Rate (kbps)	40

Fig. 6. Comparison of Size Reduction

level of user experience achievable. Hence, when there is a restriction on network bandwidth the user should be allowed to enter only those values of user-experience, that are below this limit. This limit can be calculated from the graphs of sending interval vs NO and sending-interval vs U.

So, we propose to implement it as follows. Firstly, from the graphs of NO vs sending interval, we find the minimum sending interval possible for the given network bandwidth. We find the maximum U possible corresponding to this sending interval from the graph of U vs sending interval.

Then the system administrator chooses four values of user experience below this maximum limit, from which the user can choose one according to the cost he is ready to bear. Then, from the value of U entered by the user the sending interval is found from the graph of U vs sending interval. This is done for each element separately.

I. Video Size Comparison

To calculate the typical size of video achieved in a mobile device connection like GPRS, one video was considered which had all the three Study-Elements. Images are extracted from all elements at the same interval of extraction of five seconds, at mobile device resolution. Figure 6 shows the results.

U1 and U2 are user experience values for white paper and presentation elements respectively. From the result it can be observed that our method achieves much higher size reduction than the traditional codecs and is enabling viewing at low network bandwidths of 40kbps (accounting for the audio). The user experience values are inverse of delays experienced and hence the average delay incurred by the user is only five seconds for a presentation element and only 2.5 seconds for a white paper element, which we assume to be acceptable considering the size reduction achieved.

IV. CONCLUSION AND FUTURE WORK

Slide show based adaptation achieves far greater bandwidth reduction than other techniques. It can be used to adapt lecture videos so that they can be viewed on low bandwidth network connections. It should be observed that we assume that video of slides is shown because this is important for offline viewing of the lecture.

Currently, for our method, tagging the boundaries of Study-Elements is assumed done manually typically at the time of production of video. We feel that this process could be automated by using a technique called shot detection, that identifies the places in the video, where major changes occur.

We also plan to implement the idea as a working system, and take the user experience of actual users. This could be then compared with the values obtained from the metrics.

V. ACKNOWLEDGEMENTS

REFERENCES

[1] H.264. ''http://ati.amd.com/products/pdf/h264_whitepaper.pdf''.
[2] Shih-Fu Chang, Di Zhong, and Raj Kumar. Real-time content-based adaptive streaming of sports videos. In *CBAIVL '01: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01)*, page 139, Washington, DC, USA, 2001. IEEE Computer Society.
[3] Ming-Ho Hsiao, Yi-Wen Chen, Hua-Tsung Chen, Kuan-Hung Chou, and Suh-Yin Lee. Content-aware video adaptation under low-bitrate constraint. *EURASIP J. Adv. Signal Process*, 2007(2):27–27, 2007.
[4] W. Tavanapong. A characteristics-based bandwidth reduction technique for pre-recorded videos. In *IEEE International Conference on Multimedia and Exposition*, page WP11, 2000.
[5] J. Xin and C-W Sun M-T. Lin. Digital video transcoding. Technical report, Mitsubishi Electric Research Laboratories, January 2005. Digital Video Transcoding, Proceedings of the IEEE, Vol. 93, Issue 1, pp. 84-97, January 2005, IEEE Xplore (http://ieeexplore.ieee.org/Xplore),.