CDEEP LECTURE VIDEO ADAPTATION TO MOBILE DEVICES

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree of
Master of Technology

Ganesh Narayana Murthy
under the guidance of
Prof. Sridhar Iyer

Department of Computer Science and Engineering,
Indian Institute of Technology, Bombay

Dissertation Approval Certificate

Department of Computer Science and Engineering

Indian Institute of Technology, Bombay

The dissertation entitled **"CDEEP Lecture Video Adaptation to Mobile Devices"**, submitted by **Ganesh Narayana Murthy** (Roll No: **08305027**) is approved for the degree of **Master of Technology** in **Computer Science and Engineering** from **Indian Institute of Technology, Bombay**.

Prof. Sridhar Iyer
CSE, IIT Bombay
Supervisor

Prof. Purushottam Kulkarni
CSE, IIT Bombay
Internal Examiner

Dr. Vijay Raisinghani
HoD NMIMS College, Mumbai
External Examiner

Prof. XXX
¡Dept¿, IIT Bombay
Chairperson

Place: IIT Bombay, Mumbai
Date: 24^{th} **June, 2010**

Abstract

Lecture videos access on mobile devices like mobile phones and PDAs, would prove beneficial to students, as it would provide quick and anywhere access of information. CDEEP lecture videos have a high video bit-rate that makes viewing of this video unsuitable, over low network bandwidth connections like GPRS, which is the network mobile devices generally use. Further, such networks charge the users by the amount of data transferred.

The aim of this thesis is to adapt lecture videos of "Centre for Distance Engineering and Education Programme"(CDEEP) of IIT Bombay, so that they can be viewed on a mobile device at low network bandwidth and low cost. Transcoding i.e. converting to another format is the most common method. Its efficiency at low-bit rates has been explored.

Content-Based Adaptation is a way to adapt videos based on the content present in the videos. A few examples of this methodology is discussed. A new method based on this adaptation methodology, called "Study-Element Based Adaptation" that focusses on dividing the video into study-elements, and then adapting the video based on each element, is defined and introduced. Its efficiency at low-bit rates in terms of total cost incurred by the user and the user experience is explored, and compared with the transcoding way of adaptation.

Identifying study-elements in the video requires tagging of the video. Image Processing based tagging that has been developed is explained, and its accuracy measured. Finally, an analysis of what future improvements can be made to the study-element based methodology has been discussed.

Acknowledgements

I would like to thank my guide, Prof. Sridhar Iyer for his continuos support and guidance. All this work and its successful completion would not have been possible without him. His advice on possible directions in my work, and his guidance regarding critical decisions has helped me in weighing the options wisely.

I would also like to thank Prof. Purushottam Kulkarni for his indepth reviews of my work. His anaylysis of the various aspects of the project and his suggestions during group meetings, gave me an idea of what more could be done in my work.

I would like to thank Prof. Sahana Murthy for the feedback given by her during my presentations. I would like to thank people of Informatics Lab especially Ms. Swathi Patil for helping me in hosting my work on the Oscar Website.

Chapter 1

Introduction

As part of distance learning initiatives by many universities all over the world, videos of course lectures are available to students for offline viewing on their computers, or as video-on-demand from the university websites. Mobile devices like mobile phones and PDAs are becoming more and more pervasive, especially among students. Hence, if these videos could be accessed on such devices, it will provide students with quick and anywhere access to educational content present in these videos.

However, current lecture videos are primarily authored for viewing on desktop computers, at high resolutions and occupying more space and more network bandwidth. These cannot be viewed on mobile devices because of device and network limitations, like low network bandwidth, cost and processing power to name a few, due to which they cannot be directly streamed to the mobile device or stored using less memory. Hence, these videos have to be adapted to suit the limitations of mobile devices.

The aim of this thesis is to adapt lecture videos of CDEEP(Center for Distance Education and Engineering Programme),IIT Bombay so that they can be viewed on mobile devices. The main focus of this thesis is to adapt the videos so that they can be viewed even at very low network bandwidths like that of cellphone GPRS connections and at low costs, without compromising on the user experience of the video. The other focus is to make the process adaptive, so that the user can be given control over the user experience he desires at higher network bandwidths. For example, if the mobile is connected via the college wifi LAN, then the network bandwidth would not be a limitation and hence the user should get better experience.

1.1 Limitations in Mobile Viewing

Primarily, there are two main deterrents in viewing videos on mobile devices:

1. **Network Bandwidth and Cost** - Currently, mobile devices use GPRS as their network connection. The drawback of GPRS is its very low network bandwidth, which is typically 50kbps. Since, the videos are authored at bit rates of 800-1400kbps, the available network bandwidth is very less and would cause a choppy reception even if the video is streamed to the mobile. Also, the users are charged for the amount of data transferred, in the range of 10paise for 10KB. Hence, for viewing the entire video that has a typical size of 600MB for a one hour lecture, the cost of viewing it would be enormous. Hence, any adaptation process, must make the network bandwidth requirement and size of the video to suit the available network bandwidth and desired cost.

2. **Content Visibility** - Lecture videos have written material, that may not be visible properly if viewed on small screen or if compressed too much. Hence, any adaptation process, must ensure that not only the network bandwidth and cost requirements are minimized, but also that the usability of the content is not compromised.

There are other limitations of mobile devices in specific, as given below:

1. **Processing power** - Most mobile devices have limited processing power. Even high end PDAs, do not have a processing power like that of a desktop computer. Currently, netbooks (low end laptops), are also becoming popular that also have limited processing power. Hence, the adapted video must not require high processing power to decode and play the video.

2. **Screen Size** - Mobile phones and PDAs have considerably lesser screen sizes that a desktop or a laptop. But, traditional lecture videos, have high resolutions which cannot be displayed on such small screens.

3. **Memory** - The amount of RAM and permanent storage available in mobile devices is very less. Hence, the video must not occupy more storage space than is absolutely necessary, so that more videos can be stored.

1.2 Video Transcoding based adaptation

Video transcoding[14] is the process of converting videos from one video format to another. In the process, values of video parameters like bit rate, frame rate and resolution are altered to suit the target viewing requirements. For example, to stream a video onto a low bandwidth connection, it is transcoded by reducing bit rate and frame rate.

Video transcoding is used in mainly used in video delivery applications like video conferencing, HDTV, and for delivering content to heterogenous clients like mobile phones and PDAs. The idea is to use a single video and adapt it to meet the target client device requirements, and hence enabling uniform access. It is more efficient than just decoding the and re-encoding the video into a new format.

There are three parameters of video that could be adapted:

1. **Bit rate** - This represents the video playback rate i.e. the amount of data displayed per second. The network bandwidth must be atleast equal to or greater than this bit rate so that the video can be delivered without any delays. For example, if the video bit rate is 500kbps then the network bandwidth must be atleast 500kbps, as that is the rate at which the video is played.

2. **Frame rate** - This represents the number of frames displayed per second. It is measured in frames per second(fps). A high frame rate is generally used if the video has high motion content. Reducing frame rate can reduce the network bandwidth requirement of the video, as lesser frames are needed per second.

3. **Resolution** - It represents the size of a video frame. For applications like video conferencing, a high resolution video is necessary as the content will be viewed on a bigger screen, while for mobile devices lower resolution video is required. Reducing the resolution of a video, substantially reduces the network bandwidth requirement of the video.

The downside of video transcoding is the quality of video produced. The video quality depends on the format used and on how much the video parameters are reduced. For example, formats like MPEG-2 are designed for producing high quality video, and hence do not perform well when the video bit rate is set too small. Formats like h.264 require heavy processing power, that may not be available on mobile devices.

1.2.1 Video Codecs and Formats

MPEG-1

MPEG-1(format .mpg) is a lossy compression scheme developed by ISO/IEC (International Organization for Standardization/ International Electrotechnical Commission) for compressing videos to CD-ROMs. It achieves compression rates of 50:1 to 100:1 depending on the image sequence and the video quality desired. It is mainly used for compressing the video to suit a bit rate of 1.5Mbps or less [2], as this is the trasfer rate of CD-ROM players. It is also intended to be used with images of size 352x288 or less, at frames rates locked at 25fps(PAL) and 30fps(NTSC).

The MPEG-1 algorithm uses the Discrete Cosine Transform(DCT) algorithm, to convert each frame of the video to frequency domain, and then remove the least meaningful frequencies to achieve compression, without loss of perceptible quality.

MPEG-2

MPEG-2(format .mpg) standard was evolved to meet the needs of encoding video at a high quality. High quality video is required in applications like movies on DVDs and in digital broadcasts via satellite and cable. It is used to encode video at bit rates in the range of 5-8Mbps [4]. It is very similar in its compression techniques to MPEG-1 using DCT transforms, but it outperforms MPEG-1 at bit rates of 3Mbps and above. It is also the current standard for HDTV. It is also used in full broadcast TV systems, as it has better support for interlaced video.

The downside is that, it is not optimized for low bit rate applications, and its frame rate is also locked at 25fps(PAL) and 30fps(NTSC).

MPEG-4

The MPEG-4(format .mp4) file format is based on Apple's Quicktime technology and was developed as a standard for seamless delivery of high quality audio and video over the internet. It was designed to achieve two main goals - sending video content over low bandwidth networks like the internet and cell phone networks like GPRS, and to achieve better compression than MPEG-2 in broadcast TV systems.

It supports bit rates ranging from 64kbps to 1800Mbps, but it hasn't become popular in broadcast TV as its compression rate of 15% is not sufficient. MPEG-4 is a superset of MPEG-2 and hence all MPEG-4 players can play MPEG-2 content as well.

Format	Video Bit Rates	Resolutions	Frame Rates	Applications
MPEG-1	1.5Mbps or less	352x288	25fps and 30fps	CD-ROM videos
MPEG-2	5-8Mbps	High Resolutions	25fps and 30fps	HDTV, DVD and high quality broadcast
H.264/AVC	$\gg$ 40kbps	Variable	Variable	Internet streaming and video telephony
Flash Video (VP6)	65-200kbps	320x240	15fps	Internet streaming

Table 1.1: Comparison of Video Formats Usage

H.264/AVC

H.264/MPEG-4(format .mp4) AVC is a standard jointly developed by ITU-T Video Experts Group and ISO/IEC MPEG group. It is also called as MPEG-4 Part10 AVC(Advanced Video Compression). The main goal of H.264 was to provide a packet-based network friendly video representation so that it is suitable for both conversational(video telephony) and non-conversational(broadcast,streaming and storage) applications.

It provides excellent video quality over a wide range of bit rates from 40kbps to 10Mbps and above. It also incorporates a mechanism called "Scalable Video Coding" wherein video is authored as different layers. The lowest layer is of the lowest quality and hence consumes least network bandwidth, while each additional layer adds quality to the video. In this way, the video can be adapted to suit different network bandwidths.

The downside of H.264 is that it requires huge processing power, which may not be available in mobile devices like mobile phones and PDAs.

1.2.2 Comparison of Transcoded Video Size

To determine the amount of compression achieved by the different codecs, an MPEG-1 lecture video was trascoded using different codecs to appropriate formats. The results of the achieved file sizes is shown in Table 1.2

It has to be observed that even though traditional codecs are able to achieve good compression of video size, the achieved file size is still high for GPRS, as the cost per MB that the user has to pay is high (In India, INR 3 per MB). Hence, a method is needed that can achieve further reduction in size.

Codec	Format	Original Video Size (MB)	Target Video Bitrate (kbps)	Target Audio Bitrate (kbps)	Resolution	Size (MB)
MPEG-1	mpg	432 (no audio)	40	no audio	320x240	26
MPEG-2	mpg	432 (no audio)	40	no audio	320x240	29.12
H.263	3gp	432 (no audio)	40	no audio	352x288	39
H.264	mp4	432 (no audio)	40	no audio	320x240	16.9
VP6	flv	432 (no audio)	40	no audio	320x240	20.5

Table 1.2: Comparison of Compressed Video Sizes

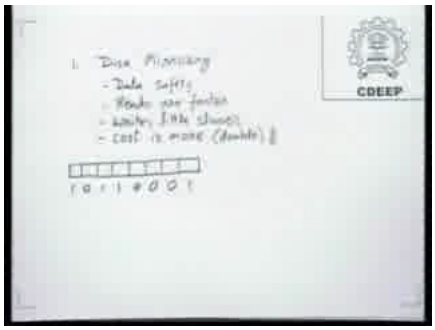
Format	Content Visibility at low bit rate	Remarks
MPEG-1	Poor	Unsuitable for lecture videos
MPEG-2	Very poor	Unsuitable for lecture videos
H.263	Poor	Unsuitable for lecture videos
H.264	Good	Content visibility is good. But, since decoding requires high processing power, it is unsuitable for mobile devices
VP6 (flv)	Good	Can be used

Table 1.3: Summary of Video Quality Comparison

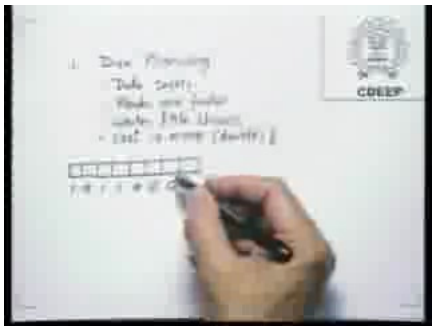
1.2.3 Comparison of Transcoded Video Quality

The major drawback of video transcoding is that the video quality at very low bit rates is poor. In the context of lecture videos, where content visibility and comprehension are important, just producing a video that meets the network bandwidth requirement, but in which the content is not visible clearly is useless. For networks like GPRS whose network bandwidth is in the range of 50kbps, the transcoded video would have very low quality.

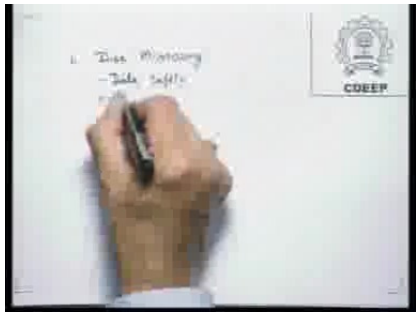
Figure 1.1 shows snapshots from transcoded videos of different formats. It can be seen from the figure, that the video quality for formats like MPEG-1 and MPEG-2, is very poor with the content being totally incomprehensible. H.264 produces a good quality video even at that low bit rate, but it has the drawback that it requires very high processing power [1], that may not be available in mobile devices. Flash Video achieves good quality even at low bit-rates but the compressed file size is still high from the cost perspective.



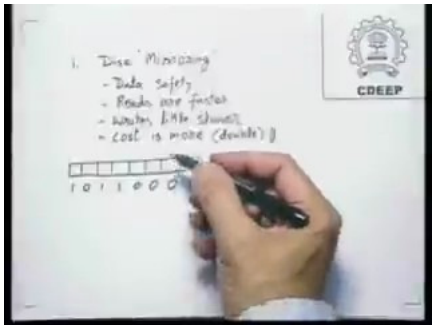
(a) MPEG-1



(b) MPEG-2



(c) H.263



(d) H.264

(e) VP6

Figure 1.1: Video Quality Comparison after transcoding. Trancoded Video Bit rate - 40kbps, Original Video Bit Rate - 1150kbps. Target Resolution: 320x240.

1.3 Content-aware Adaptation

Content-aware adaptation is the process of adapting the video, without compromising the visibility of the content, by taking into account the content in the adaptation process. In the context of the low-bit rate problem that we are aiming to solve, content-aware adaptation mainly involves using the available network bandwidth for sending the most important information from the video. There are many content-aware schemes in the literature, a few of which we discuss below.

1.3.1 Content-aware adaptation under low bit-rate constraint (Hsiao et.al.)

Hsiao et.al.[9] adaptation process basically identifies which regions of a video frame captures the attention of the user, and tries to achieve high quality for those regions, while other regions are encoded at low quality. The video is analyzed in the compressed domain i.e. without decoding, and such important regions are identifed using a visual-attention model based on brighteness, location and motion in the video frames. Region-weighted rate-distortion model is used for allocating bit rate to these regions.

This method is content-aware in the sense that it tries to identify important objects from the video, and use the available network bandwidth for improving the visual quality of these important objects. But, the drawback here is that the quality of the important region is still dependent on the available network bandwidth. If it is very low, then even the important objects, though they utilize the full bit rate, would still have low quality. This is a common problem for any method that tries to show a video at low bit-rate. The sample images shown in the paper, do not have satisfactory quality at a low-bitrate of 75kbps.

1.3.2 Real-time Content-Based Adaptive Streaming of Sports Videos (Chang et.al)

In Chang et.al.[8], the adaptation is in the context of sports videos. Here, important segments are certain events occuring in the video. An event is happening of an action such as serves in tennis, pitches in baseball, commercials etc. Such events typically occur when the camera view changes, and can be identified using shot-detection techniques in image-processing.

Some events are important like a serve in tennis, pitch in baseball, scoring of goal in footaball. Such important events are shown as video encoded at highest possible rate according to the available network bandwidth. Unimportant events like commercial breaks are encoded as slideshows by showing only keyframes thereby reducing their bandwidth requirement. They try to solve the limited network bandwidth problem through adaptive streaming by varying the bit rate of the streamed video, so that it doesnot exceed the client bandwidth and buffers.

This method achieves network bandwidth reduction by encoding non-important sections as slideshows, thereby distributing their bandwidth among important videos. It can be observed that by showing slideshow of images, network bandwidth can be reduced considerably, and thereby meet the low-bit rate constraint.

1.3.3 Characteristics-Based Bandwidth Reduction Technique (Tavanapong et.al)

In Tavanapong et.al[13], the above idea of showing the video as a slideshow of images is used. It is done in the context of pre-recorded videos. Non-changing portions of the video are identified, and one image from that portion is taken. The adapted video is basically a slideshow of these images along with the audio. Portions of video that do not have much change in the content are identified by using inter-frame differences.

This is a promising method in the sense that it achieves both reduction in required network bandwidth and it achieves huge savings in total amount of data transferred (about 95%). This is important for networks like GPRS where the user is charged for the amount of data transferred, the adaptation process must ensure that the size of the adapted video is small apart from meeting the low-bitrate constraint.

1.4 Proposed Adaptation Methodology

As observed above, showing sections of video as slideshow of images extracted from it, results in huge reduction in the required network bandwidth for showing the video, and a reduction in overall size of the video, and hence is a promising method to enable viewing of the information in the video over networks with low network bandwidth and cost constraints. We use the same method, in our proposed method called “Study-Element based adaptation”.

Method Name	Adaptation Mechanism	Video Quality	Remarks
Hsiao et.al	Identify visual attention regions in a frame. Encode them at high quality	Poor	Quality of important objects still depends on network bandwidth
Chang et.al	Identify important events like serves in tennis videos, goals in football videos etc. Encode these at high quality.Encode non-important events like commercials as slideshow of images.	Good	Showing slideshow of images reduces required network bandwidth, enabling important regions to be encoded at high quality. It also reduces the size of the adapted video.
Tavanapong et.al	Identify non-changing portions of video. Extract one image from each such portion. Output is a slideshow of images and audio.	Good	Exploits redundancy in video by eliminating redundant information in non-changing portions of the video.

Table 1.4: Comparison of Content-Aware Methods

In our method, "Study-Elements" are identified within the video. Study-Elements are portions of the video representing information presented on a specific medium of instruction. For example, the portion of video showing a presentation slide is a study element. Three types of study-elements are used - "Presentation Element" representing one slide of presentation, "White paper element" representing the portion where the instructor is explaining a topic on a white paper, and "Instructor element" which shows the instructor explaining to the class.

Each study element, is then presented as a slideshow of images along with the audio. These images are sent at an interval called the "sending interval" along with the audio. For example, a sending interval of five, implies that one image is sent every five seconds. To quantify parameters like user-experience, network bandwidth requirement and resulting size of the adapted video, metrics have been defined for the same.

A general relation is then found between the sending interval and the above parameters. This relation makes our method adaptive and customizable, as the user can be given control over the level of user-experience he desires, at a certain level of network-bandwidth. This is possible as more the sending interval more will be user-experience. Hence, given the user-experience and network bandwidth the sending interval to be used for each study element can be foound.

Hence, our method not only adapts the video to low-network bandwidths, but is also

customizable with the user given control decide the level of user-experience he desires, based on the network bandwidth. The process is elaborated in the following chapters.

Chapter 2

Study Element Based Adaptation

2.1 About CDEEP Lecture Videos

Center for Distance Education and Engineering Programme(CDEEP) is an distance learning initiative by Indian Institute of Technology Bombay. Videos of lectures are broadcasted live to remote centres and are also available as Video-On-Demand from the CDEEP website. Students can download these videos free of cost.

The instructors use mainly two modes of instruction namely presentation slides and explaining on a white paper. Three different cameras capture the presentation slides, the white paper and the instructor separately. During the lecture, the desktop screen on which the presentation slides are shown, is projected onto a screen present in the classroom. When the instructor uses the white paper to explain, input is taken from the whitepaper camera, and the is projected on the screen.

The switch from one camera to another is performed by the technical staff present in the classroom. The final video contains the video that was projected on the screen, with intermediate switching to whitepaper or the instructor as wad done during the lecture.Hence, the final video is an interleaving of video of the instructor, video of the presentation slides and the video of the white paper.

2.2 Study Elements Definition

As mentioned above, CDEEP lecture video contains an interleaving of video of presentation slides, video of white paper on which the instructor explains something and video of instructor. We define each one of the above portions of the video as study elements.

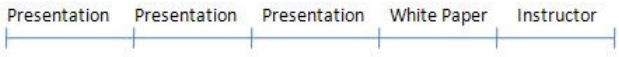
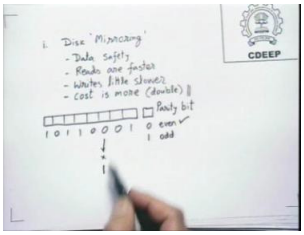
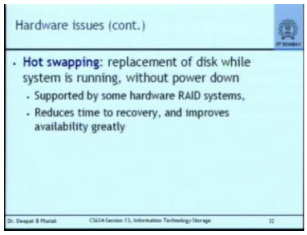


Figure 2.1: Study Elements in a video

(a) Instructor Element



(b) White Paper Element



(c) Presentation Element

Figure 2.2: Sample Study Elements

Figure 2.1 shows the study elements in a sample lecture video.

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
3. **Instructor Element** - Portion of video that shows instructor talking

Sample images of the elements are shown in Figure 2.2

2.2.1 Motivation of study elements

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

1. **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.
2. **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

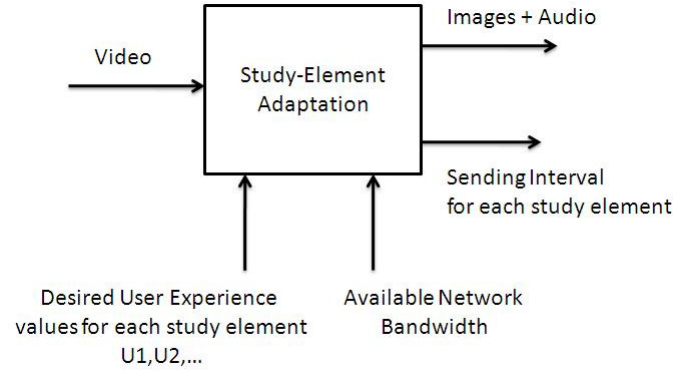


Figure 2.3: Block diagram of system

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.

2.3 System Overview

Figure 2.3 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 is taken for each study element. For example, the desired user experience values for presentation element and instructor element are taken as input. The output is a set of images, image sending intervals for each study element and the audio.

The 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 slideshow of images while hearing audio. It should be noted that the sending interval can be different for different study elements, thereby achieving more user experience for some study elements while less for others. This is significant as, for example, for white paper element, a lower sending interval might be preferred for seeing the changes faster while a higher sending interval may be preferred for a presentation element.

The desired user experience value also controls the amount of data transferred and hence the cost. When the desired user experience is more, the amount of data sent and hence the total size, would be more. Hence, to control the cost, the user has to reduce

the desired user experience value.

2.4 Adaptation Procedure

The study element based adaptation method has the following three basic steps:

1. **Tagging** - In this step, the starting and ending points in time are identified in the video for each type of study element.

2. **Building User Experience Index** - User Experience Index is a relation between the sending interval, and the user experience that would be achieved at that sending interval, and the network bandwidth that will be required to support it.

3. **Finding Sending Intervals** - In this step,the sending intervals of each study element are calculated by using the user experience index, with the network bandwidth and desired user experience values as inputs from the user.

4. **Ouput the Adapted Video** - In this step, images are extracted from the video, at an interval equal to the sending interval. Also, we extract audio from the video. Output the images, audio and the sending interval of images for each study element. The output images and audio form the adapted video.

The steps are described in the following sections.

2.4.1 Tagging

Feature based tagging, as described in Chapter 3, is used to define the study element boundaries. The tagging involves defining the following parameters for all instances of study elements:

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.

2.4.2 Building User Experience Index

Sending interval of images is related to the user experience, network bandwidth required and the total size of the adapted video.

For example, if images are sent at a low interval, for example, one image every one sec, the user would see the updates very fast and hence the user experience would be high. Simultaneously, the network bandwidth required would be high. The total size would also be high, because one image is sent every one second and hence many images have to be transmitted. Similarily, for a higher sending interval, the user experience would be low, but the network bandwidth required would also be low.

Hence, a relation is required between the sending interval and each of the above parameters. For this, we define metrics that quantify each of the above parameters of user experience, network bandwidth and size, 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 examining a set of ten videos which contain videos from all departments. This is called as the "User Experience Index". We assume the minimum value for sending interval as one second and maximum value as one minute.

Finding User Experience

When images are sent to the client as a slideshow along with the audio, there will be certain delay experienced by the user, between the time of hearing and time of seeing. For example, if one image comes every five seconds, then the user might hear about some concept now, but will be able to see the instructor's writing explaining the concept or the presentation of the concept, only five seconds later . For each study element, the delay to be measured is different. This delay experienced can be compared to the delay that will occur if a video is streamed. This ratio would give the user experience value for the study element.

1. Presentation Element As mentioned already, a presentation element represents one slide of a presentation. The delay that is important here is the delay in the starting of the slide. Once the slide is visible, no other image of the slide is necessary, and hence can be either discarded or send less number of images by choosing a high sending interval.

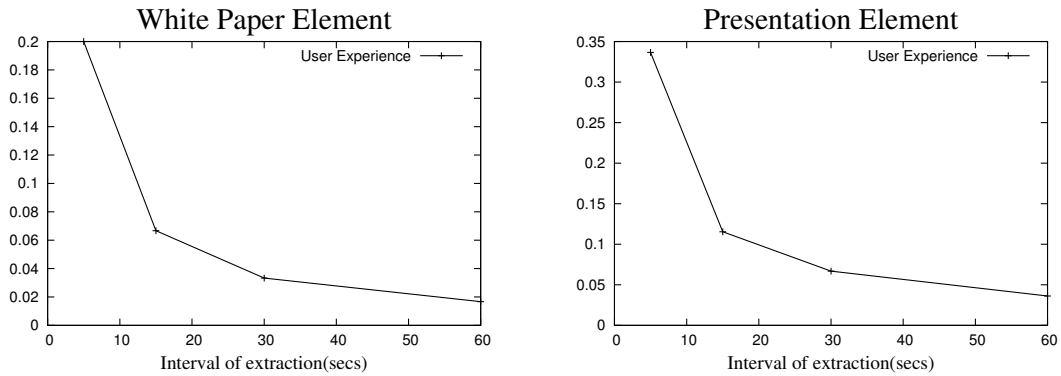
Thus, we fix a sending interval 'r' and then define the delay experienced by the user, and the user experience based on this delay, as given below. The user experience is just a comparison of a delay at any sending interval 'r' with the delay achieved at a sending interval of one second. It is assumed to that a delay of one second that the user perceives at r=1 is acceptable to the user. A sending interval of one, requires apprx 96kbps of network bandwidth.

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

Now, for the sending interval 'r', there will be some amount of network bandwidth required to support it. Typically, since one image is sent every 'r' seconds, it would be the ratio of average size of an image extracted from a presentation element and the sending interval 'r'. This is called as "Network Overhead". Correspondingly, at this sending interval, the total amount of data transferred as images, called as "Size Overhead" would be total size of all images of all presentation elements in the video.

$$\begin{aligned} NetworkOverhead(NO2) &= \frac{Average\ Image\ Size}{Sending\ Rate\ 'r'} \\ SizeOverhead(SO2) &= Total\ Size\ of\ all\ images\ sent\ at\ interval\ 'r' \\ &\quad from\ all\ presentation\ elements\ in\ the\ video \end{aligned}$$

Experiment - A set of ten videos of courses of different departments are considered. For each video, firstly, a sending sending 'r is chosen, and the average image size of images extracted from presentation elements in the video at this interval 'r', is found. Then, the parameter D2 is found for each video by examining a sample of five presentation elements in the video. This is followed by finding NO2,BO2 and U2 for the video. Finally, the average of all these parameters is taken to generate a relation between the sending interval 'r' and the U2,NO2,BO2. This is repeated for various values of 'r'.



(a) White Paper Element (b) Presentation Element

Figure 2.4: User Experience values

Result - From the above experiment, a relation was obtained between the sending interval 'r' and the User Experience(U2). This relation is shown as graph in Figure 2.4. Another relation between the sending interval 'r' and Network and Size Overheads (NO2,SO2) was obtained. This is shown in Figure 2.5. From the graphs, it is evident that by choosing appropriate sending interval 'r', even low network bandwidths of 16kbps can be supported without compromising user experience.

2. White Paper Element As mentioned already, a white paper element represents the portion of the video, where the instructor is writing something on a white paper, as shown in Figure 2.1. In a white paper element, the instructor writes or explains written material, and hence the delay between any two images is noticeable to the user. So, the delay that is important here is the delay between any two images, which is same as the sending interval.

So, we fix a sending interval 'r' and define the Delay Experienced and the user experience as given below. Here again, the user experience is a ratio of the delay at the interval 'r' with the delay at interval '1', which is assumed to be acceptable to the user.

$$\begin{aligned} Delay\ Experienced(D1) &= Sending\ Rate\ 'r' \\ User\ Experienced(U1) &= \frac{1\ sec}{D1} \end{aligned}$$

Similar to presentation element, the network and size overheads for the sending interval 'r' for a white paper element, is calculated as:

$$\begin{aligned} NetworkOverhead(NO1) &= \frac{Average\ Image\ Size}{Sending\ Rate\ 'r'} \\ SizeOverhead(SO1) &= Total\ Size\ of\ all\ images\ sent\ at\ interval\ 'r' \\ & \quad from\ all\ white\ paper\ elements\ in\ the\ video \end{aligned}$$

Experiment - A set of ten videos of courses of different departments are considered. For each video, firstly, a sending sending 'r is chosen, and the average image size of images extracted from white paper elements in the video at this interval 'r', is found. Then, the parameter D1 is found for each video by examining a sample of five presentation elements in the video. This is followed by finding NO1,BO1 and U1 for the video. Finally, the average of all these parameters is taken to generate a relation between the sending interval 'r' and the U1,NO1,BO1. This is repeated for various values of 'r'.

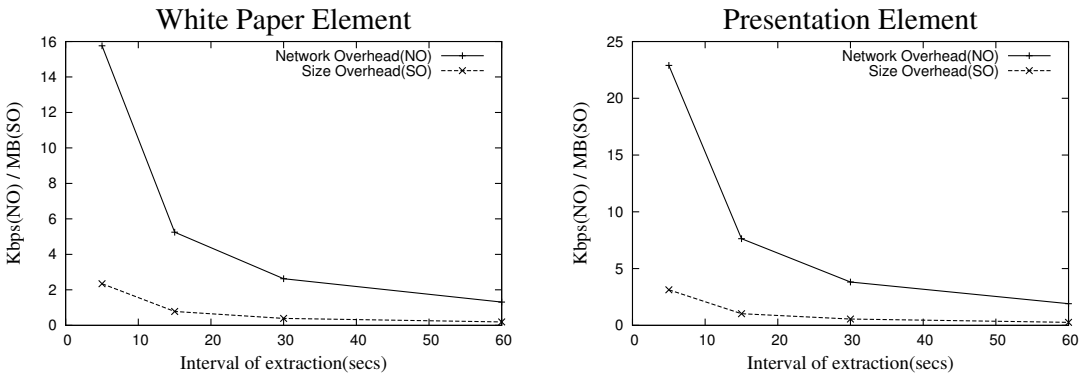
Result - From the above experiment, a relation was obtained between the sending interval 'r' and the User Experience(U1). This relation is shown as graph in Figure 2.4. Another relation between the sending interval 'r' and Network and Size Overheads (NO1,SO1) was obtained. This is shown in Figure 2.5. From the graphs, it is evident that by choosing appropriate sending interval 'r', even low network bandwidths of 16kbps can be supported without compromising user experience.

3. Instructor Element We donot send any images from instructor element, as it is assumed that it does not present any information content. Hence, we can consider that the user experience value for instructor element at any sending interval 'r' is zero and the network and size overheads are also zero.

2.4.3 Finding Sending Intervals

Given that the relation between sending interval and user experience(U), network overhead(NO) and size overhead(SO) has been found, we can easily calculate the sending interval to be used, according to the choice of user experience value and available network bandwidth. For each study element do the following:

1. **Find minimum sending interval** - From the user experience index, find the sending interval based on the network bandwidth availabe, supplied by the user.



(a) White Paper Element (b) Presentation Element

Figure 2.5: Network and Size Overheads

This sending interval is a lower bound, and the images cannot be sent at a interval less than this. But, to reduce cost, we could send images at a higher interval, so that lesser images are sent, but correspondingly user experience would be lower.

2. **Find the actual sending interval** - Now, from the sending intervals greater than the minimum sending interval, find the one that corresponds to the supplied user experience value.

The choice of desired user experience values available to the user is decided by the system administrator. Once defined, the choices entered by the user can be looked up in the user experience index i.e. the relation graphs and the corresponding sending interval can be decided, as described above.

2.4.4 Ouput the Adapted Video

Adapting the video involves extracting images according to the sending interval, and extracting audio from the video that is to be streamed. It is assumed that in a preprocessing step, images have already been extracted at a interval of 1 image every 1 second. We can find the target images and audio as follows:

1. **Extract images according to tagging** - Suppose if a presentation element occurs from t=2s to t=15s in the video, and the sending interval for presentation element is decided to be 1 image every 3 seconds, then we choose the images at t=3,t=6,t=9,..,t=15 for this element. This is done for each study element according to its chosen sending interval.

2. **Extract audio** - We use mp3 codec to compress audio from the video. The system administrator can decide the parameters to used like bit rate so that different versions of the audio file can be used for different network bandwidths to get better experience.

2.5 Results

2.5.1 Video Size at GPRS Bit Rates

GPRS bit rate is about 40kbps, so a video was adapted to suit this network bandwidth by setting appropriate values of network bandwidth. Same user experience was set for all elements, and the total size of all the images received was measured. The results are given in Table 2.1. It can be seen that the reduction in size achieved is near to 100%.

Original Video Size	**Adapted Video Size**	**Percentage Reduction**
432 MB	10 MB	98%

Table 2.1: Adapted Video Size

Format	MPEG-1
Bit-Rate	1150 Kbps
Frame Rate	25 fps
Audio	No audio

Table 2.2: Original Video Parameters

Bandwidth	40 kbps and above
UE White Paper	0.2
UE Presentation	0.33
Audio	No Audio

Table 2.3: Adapted Video Parameters

2.5.2 Transcoding Vs Study-Element Method

A comparison of transcoding method of adapting a video to the study-element method, by using the same video described above. The target network bandwidth was 40 kbps,

Original Video Size	Adapted Video Size	
	Transcoding(H.264)	Study-Element
432	17 MB	10 MB

Table 2.4: Transcoding Vs Study-Element

and the adaptation parameters of study-element method was same as those given in Table 2.3. The results of the adaptation has been shown in Table 2.4.

It can be seen that by using the study-element method of adaptation, the video size achieved is lower and hence lower is the cost. The cost will be further reduced if instructor images are not sent at all.

2.5.3 Customize Cost with Same User Experience

Figure 2.6 shows the sending intervals of the two types of elements and the corresponding cost(size), user experience and the network bandwidth required to support that sending interval.

It can be observed that by reducing the user experience of the presentation element, the overall cost incurred can be reduced yet keeping the user experience of white paper elements the same. This shows how the adaptation methodology provides individual control over each element's cost and user experience.

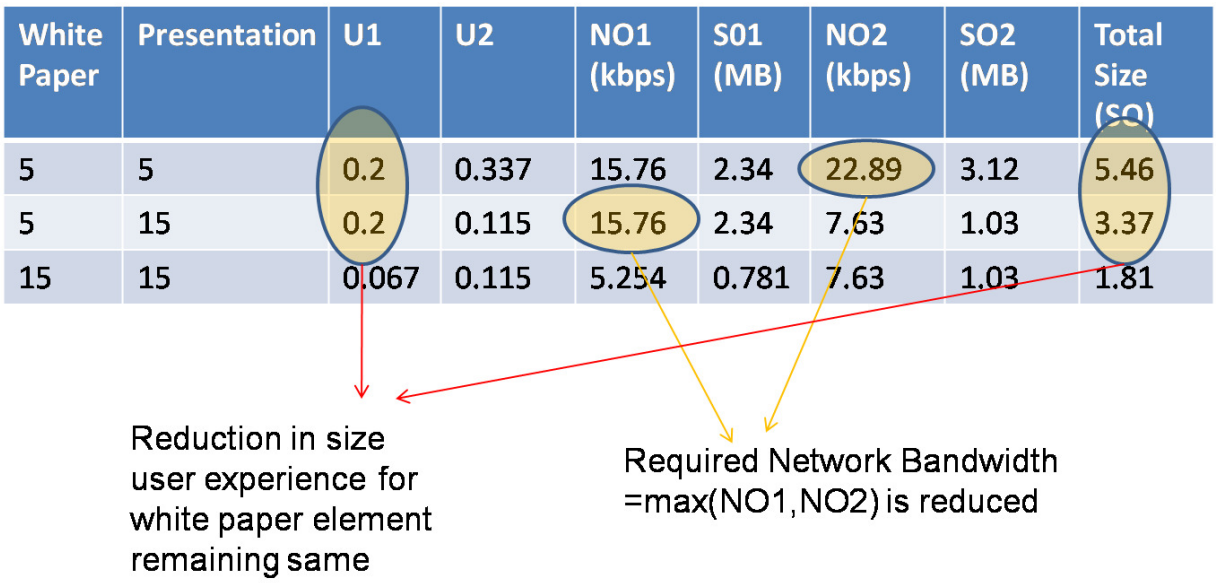


Figure 2.6: Reducing Cost

2.5.4 User Experience at different network bandwidths

To see how our method works at different network bandwidths in terms of the user experience, we found the user experience of our method at different bitrates ranging from 16kbps to 500kbps, for a white paper element. For this we consider an average image size of 11.8KB and find the lowest sending interval possible at any network bandwidth. We take this as Delay (D) and calculate the user experience.

$$Delay\ Experienced(D) = \frac{Image\ Size}{NetworkBandwidth}$$

$$User\ Experience(U) = \frac{1\ sec}{D}$$

From the graph, it can be seen that the user experience of our method increases and the delay experienced by the user decreases. This is significant as the user expects better performance if available network bandwidth is higher. So, a user who has high network bandwidth connection like WiFi, can expect the highest network bandwidth. A user with EDGE with GPRS, that gives bit-rates close to 80kbps, can expect higher user experience than a user working on GPRS alone.

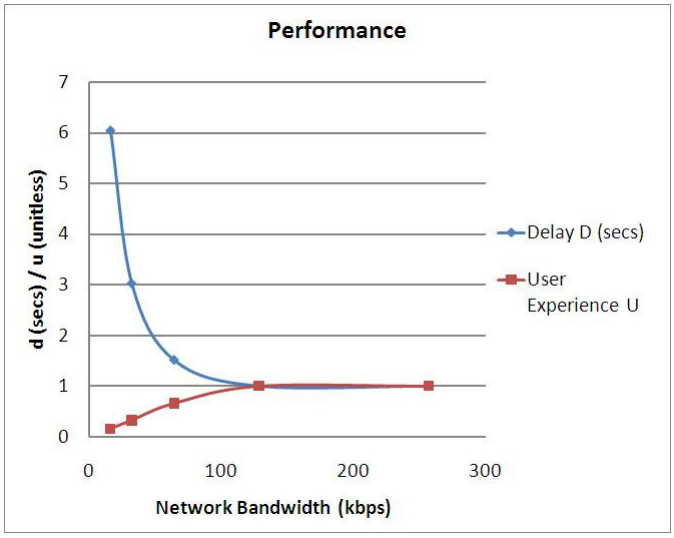


Figure 2.7: Performance at Different Network Bandwidths for White Paper Element

Chapter 3

Video Tagging

In the adaptation scheme discussed so far, one necessary component is the identification of the time boundaries of each type of element. In order to change the user experience of a specific element (say presentation), we need to know in what portions(in time) does this element occur.

Video Tagging is the process of identifying the time boundaries of each of each element within the video. Shown in Figure 3.1 is a timeline of a video and in the timeline where each element occurs. After this video is tagged, the result would be an xml file containing the tags as shown in Figure 3.1

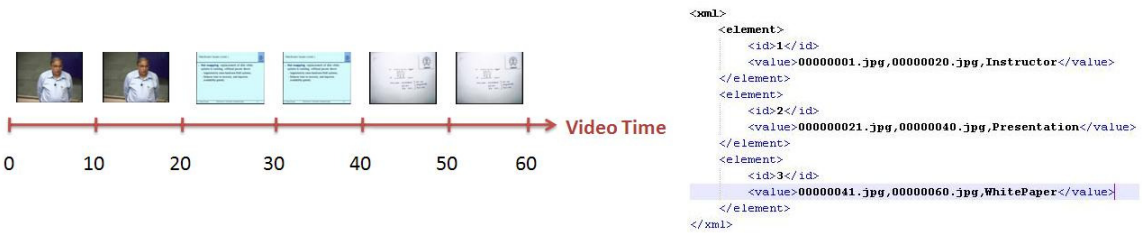


Figure 3.1: Study-Elements in Video

3.0.5 Tagging Methodologies

There are different options for tagging the element boundaries in the video.

Collaborative Tagging

In this form of tagging, the users who watch the orginal video(i.e. before adaptation), can tag regions as containing whitepaper, presentation slides or instructor. An attempt for collaborative tagging in e-learning can be found in [7].

Advantages

1. Certain element time boundaries will be available because of the tagging by users. Hence using this other similar elements can be identified and the accuracy of boundaries of other elements can be improved.

2. Important and commonly visited elements can be identified.

Disdadvantages

1. Manually tagging all the instances of elements in the video cannot be expected, as it is time consuming. For example, users cannot be asked to manually tag all regions in the video where a white paper occurs.

2. This process is not automatic.

Frame Differences Method

In this form of tagging, image processing is used. Every frame of the video is compared with the previous frame, and if the difference is greater than a threshold then it means there is a substantial change in the video at that portion. The parameter used could be RGB values or other custom defined parameters and it varies from implementation to implementation. This method is used in [13] to find non-changing portions of the video.

Advantages

1. It can find out non-changing portions of the video.

Disadvantages

1. It is not content based i.e. it cannot identify user-perceivable objects such as portion of video showing a white paper.

2. Its accuracy depends on the parameter defined.

Feature Based Tagging

In this from of tagging, features are defined based on the content present in the video, and each frame in the video is tested as whether it contains this feature or not. For example, in CDEEP videos, a white paper element usually has a white background and has CDEEP logo on the top. An attempt to tag video based on visual features in the video and representative keyframes can be found in [5] and [15].

Advantages

1. It is content-based and hence domain knowledge can be used to identify user-perceived features.

Disadvantages

1. Misclassifications can occur if the assumed feature is not present. For example, if a white paper element does not have a CDEEP logo, then the process can fail.

3.0.6 Overview of Tagging Process

The method used in this adaptation scheme to tag videos for the different element time boundaries, is a feature based scheme. In this method, images are extracted from the video every one second, and then each image is classified as either a white paper element, or an instructor element or a presentation element. To classify any image, a feature needs to be defined for each type of element, that uniquely identifies an element. For example, in CDEEP videos, every white paper used has a CDEEP logo on top right corner of the sheet. So, any image that contains this logo on the top-right corner, can be classified as white paper element.

The downside of this method is that there can be misclassifications. So, there must be a provision for manually correcting these errors, before the actual tag file is generated. So, the overall procedure for tagging is shown in Figure 3.2.

Figure 3.2: Tagging Process

3.0.7 Image Processing Library and Techniques

OpenCV 1.0 is an C++ library that has pre-built function for some common image processing functions. The image processing functions used in the classification scheme of the tagging method, are explained below:

Thresholding

Any image pixel has three components R,G,B whose values range from 0-255. By converting the image to gray scale(i.e. black and white), we have only one component viz. gray level with value from 0-255 (called intensity values).

In images, the background and foreground are usually of different colors and hence different intensities. For example, in a white paper image the background is white (intensity-255), while the writing and logo are in black (intensity-0). Hence, by fixing a threshold like 150, we can distinguish the pixels of the image into foreground and background, by making the foreground pixels to have value 0 and background pixels to have value 255. Then only the foreground or background pixels can be chosen for processing. For example, to check whether the logo exists or not, only the foreground pixels (intensity-0) pixels can be processed.

Thresholding functions are available as part of OpenCV 1.0 library and hence can be used directly on the image.

Template Matching

Template Matching is the process of finding whether a template image is present in the source image or not. For example, to identify whether an image consists of the CDEEP logo or not, the template image is the CDEEP logo while the source image is the image being checked for.

Template Matching is generally applied, after converting the gray scale and applying thresholding, so that the compared pixels have either the same value(0 or 255) or different values.

Template Matching function is also available in OpenCV 1.0 library.

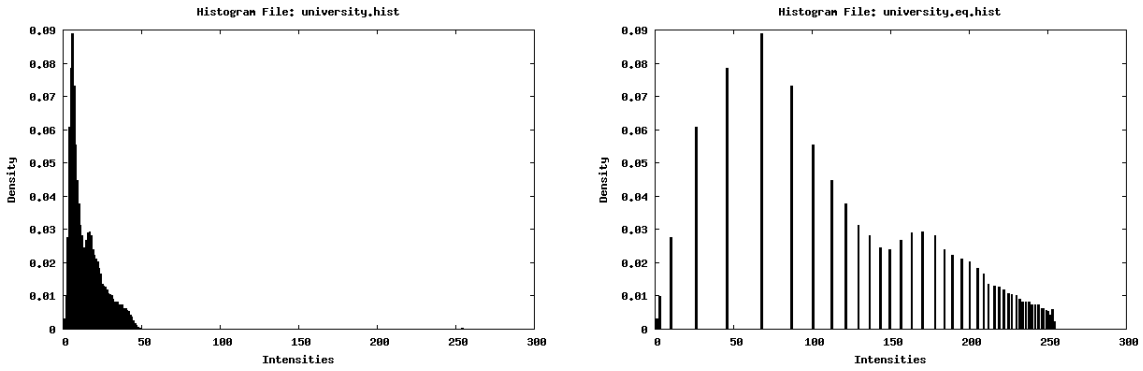
Histogram Equalization

When the lighting is not proper, there is no proper difference in intensities of foreground and background in images. This is similar to a low contrast images where no object is

clearly visible. To distinguish foreground and background clearly, and to bring objects to make objects of interest that are in the foreground clearly visible, we used this method of histogram equalization.

Generally, foreground and background are not clearly distinguishable, if the intensity of pixels are not equally distributed i.e. certain intensity levels are more predominant while virtually very less pixels exist in other intensity levels. This method, makes all intensity levels equally probable.

(a) University (b) Histogram Equalized



(c) Original Histogram (d) Equalized Histogram

Figure 3.3: Histogram Equalization

3.0.8 Classifying Images as White Paper

White paper elements are those portions of the video that show a whitepaper on which the instructor is explaining. To classify an image as a part of white paper element, the CDEEP logo on the top-right corner of the paper is used as a feature. All images that contain the CDEEP logo text “CDEEP” on the top-right corner, would be classified as whitepaper.

(a) CDEEP logo (b) Feature

Figure 3.4: CDEEP logo and the part used as feature

Template Matching after Thresholding(TMT)

To do the above, template matching is used. A sample image of the word "CDEEP"in the logo(template image) is taken from an existing white paper image. It then compared with all portions of the current image(template matching), to identify the portion of the image that is likely to contain the image.

Template matching returns a value that denotes how much the portion of the current image identified as containing the template, differs from the template. A threshold for this template was manually identified by trial-and-error, such that most of the classifications for videos from different college departments are classified correctly. Before template matching, thresholding was used to get a binary image, that captures the main foreground features.

Figure 3.5: Matching the CDEEP Logo Feature During White Paper Element Image Classification

Template Matching after Histogram Equalization(TMHE)

Histogram Equalization is an image processing technique that is used to improve the contrast of images, so that some features that are hidden because of improper lighting

are brought to the fore. After conducting sample trials, this technique was found to classify certain images that were not classified correctly by the previous method.

So, in addition to the above method, this method was also used to improve classification accuracy. In this method, the same template matching was used, but after histogram equalization of the image, rather than thresholding of the image.

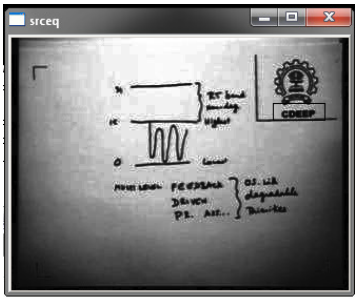


Figure 3.6: Matching the CDEEP Logo Feature

Overall Procedure

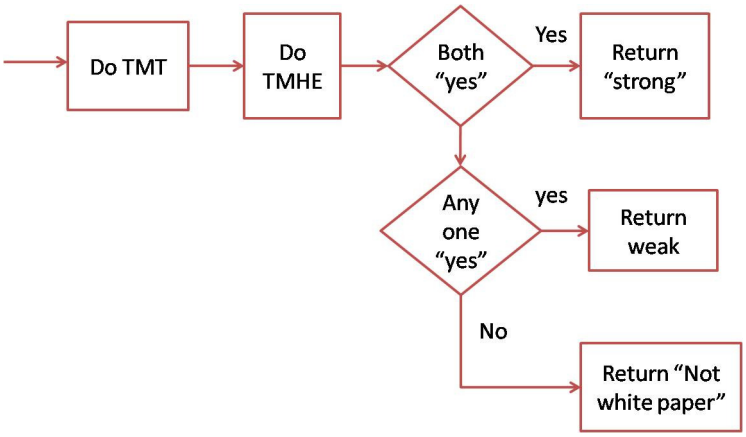


Figure 3.7: Overall Procedure to classify Image as White Paper

In addition, the first white paper element image classified is "marked", and every other image that is checked whether its whitepaper or not, is also checked for similarity with this marked whitepaper image. Only if the similarity is there, then it is accepted as white paper. This is used to further reduce the inaccuracies in the classification.

3.0.9 Classifying Images as Instructor

Instructor elements are portions of the video showing an instructor talking to students explaining something. Since, an image that is part of the instructor element, must show the face of the instructor, face detection is used to detect faces in the images. If a face is found, then its part of an instructor element.

Figure 3.8: Sample Face Detection while Classifying Instructor Element Image

There are two types of face-detection classifiers available in OpenCV - frontal face detection and profile-face detection(side-view). Both are used. The algorithm is shown below:

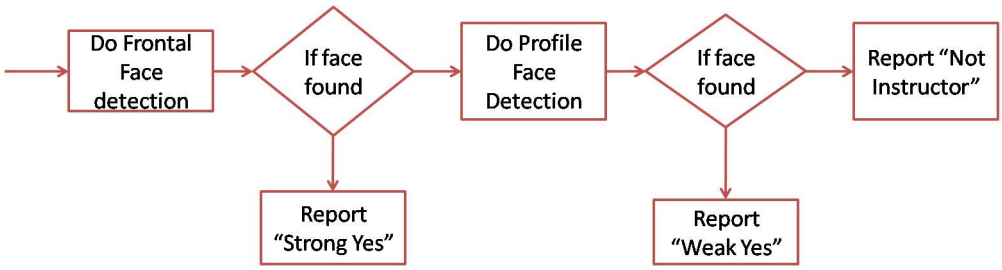


Figure 3.9: Overall Procedure to classify image as Instructor

Problems in Face Detection

Face detection algorithms do not work for all postures of the face. Only if the face is straight or totally in side-view, the algorithms detect faces. Cross-views of the face generally are problematic and may or may not be classified. Hence, this misclassifications possible in face-detection must be compensated by some other means.

The misclassifications are compensated by comparing the image with an earlier image that has been classified as instructor element, for similarity. The similarity is measured by direct pixel to pixel color comparison. Though this may not be the best way for identifying similarity, a proper threshold can be identified.

3.0.10 Classifying Images as Presentation

Presentation element is that portion of the video, that shows presentation slides. No unique features exist to tell whether an image shows a slide or not.

But, usually the headings of all the slides are similar in color as the template of the slide does not change. This fact can be utilized to tell whether an image is part of presentation element or not, by comparing its heading with an earlier presentation element. The comparison process is a simple color comparison(instead of RGB, HSV space of images is used).

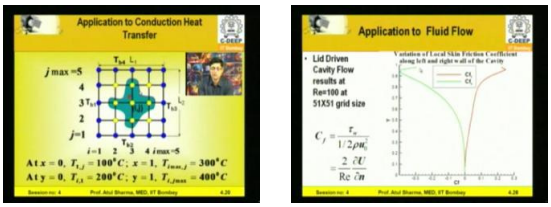


Figure 3.10: Presentation Slide Images with Similar Headings

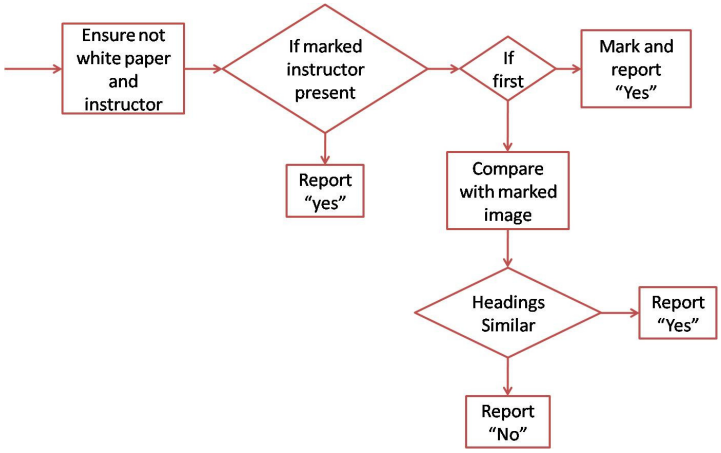


Figure 3.11: Overall Procedure for Classification of Image as Presentation

Marking Instructor before Marking Presentation

Since instructor element classification is error-prone, as explained earlier, no image can be classifed as presentation before we get a sure classification for instructor. So its essential that the initial part of the video shows instructor properly for some time, so that a proper identification can be made.

3.0.11 Overall Algorithm for Classification

The overall procedure for classification is shown in Figure 3.12. There are three essential phases in this algorithm:

1. Image Classifying into an element.
2. Storing the first classified image for every element.
3. Comparing subsquent images with this image for similarity, in addition to the standard procedure detailed in the above sections.

So, in addition to the standard features used for classifying each image as one of the elements, comparison with earlier similar images has been used, to further improve accuracy of classification.

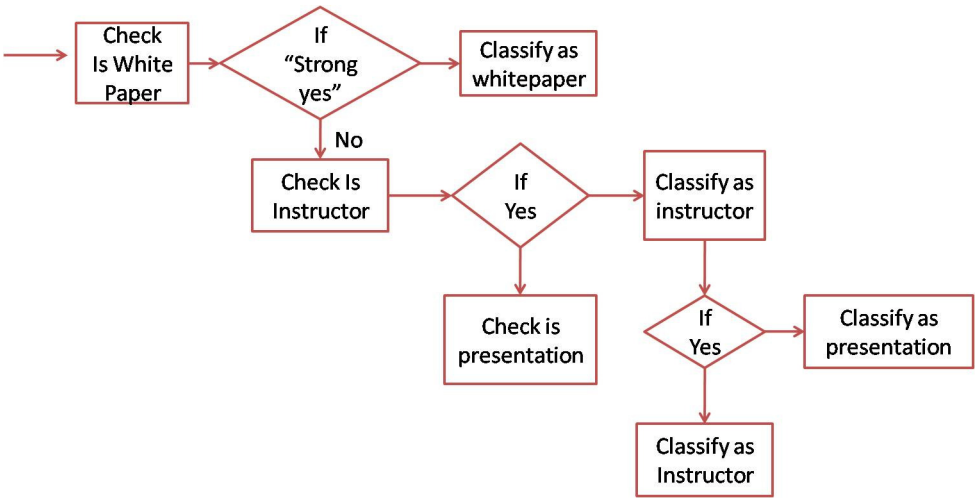


Figure 3.12: Overall Procedure for Classifying an Image

3.1 Classification Accuracy

Accuracy percentages were calculated after adaptation of videos taken from different college departments - Computer Science, Mechanical, Electrical to name a few. The results are tabulated below:

VideoName	Chief Features	Accuracy		
		White Paper	Presentation	Instructor
cs634	Has all elements Slides color close to white paper	100%	100%	99.96%
cs684	Has all elements Less slides More whitepaper	91.24%	99.44%	100%
me704	Has only slides and instructor elements	100%	100%	100%

Table 3.1: Classification Accuracy Results

Video Characterisitics for Favourable Classification

The classification algorithms work well generally if the video has the following properties:

1. Instructor is shown for atleast 10-20 seconds at the beginning.
    - Since presentation elements do not have a feature, images can be classified as presentation elements accurately only if atleast one instructor or white paper image has been classified accurately. Since all videos do not have white paper images, atleast one instructor image must be classified accurately.
2. Instructor explains the slides in full screen mode.
    - Presenation image headings are compared with earlier presentation images headings for similarity.
3. All white papers have CDEEP logo on them (which generally is the case).
    - This is the feature by which images are classified as white paper elements.

Chapter 4

System Implementation

The aim of this system, was to enable viewing of a video that was adapted by using the Study-Element based adaptation scheme, on a mobile device. This means that the user should be able to see the images of each element, at an interval based on the desired user-experience as selected by him/her, along with continuos audio of the lecture. There were many issues to be considered while taking decisions about the design of the system as explained below:

4.1 Design Considerations

4.1.1 Port Blocking in GSM networks

Mobile networks have a proxy server to which all the mobile phones send their packet data. This proxy server does not allow direct TCP socket connections to any port except for HTTP ports 80 and 8080. This was tested by deploying a simple java server listening on port 15000, on a home PC connected to a home broadband connection. A connection to this server was attempted from a mobile java application, throught GPRS and it was unsuccessful.

Hence, a web server along with Java servlets has been used as a server to which all the client mobile applications connect to. Since, the web server requests and responses go over http, there will be no port issues involved.

4.1.2 NAT in GSM networks

As already mentioned above, mobile networks have a proxy server to which all the mobile devices are connected. These mobile devices are connected in a private LAN to the proxy server. The proxy server acts as a NAT server.

The significance of this is that mobile device's IP address is not directly available. This can have an impact on streaming audio to the client. Audio cannot be streamed by the server to the mobile by sending the stream to the mobile device's IP address. Instead the streaming must be RTSP based, with a dedicated RTSP server deployed to which the mobile sends its requests for audio on demand.

But due to certain problems mentioned below, audio streaming has not been used, but progressive download has been implemented.

4.1.3 RTSP and Audio Streaming to Mobile Phones

In this system, pure audio streaming has not been implemented, but progressive download has been implemented, the reason being explained below.

RTSP is a streaming protocol, that is used to stream audio and video along with controls like pause, play, stop and so on. Attempt was made to stream audio by deploying Helix Media Basic stream server which hosted all the audio files, and then connecting to it from a mobile java application and requesting for a specific audio. But this attempt was unsuccessful and the audio could not be streamed.

The precise reason could not be identified. But the probable reason could be, that the mobile device's support was only for video formats but there was no support for audio formats like mp4a(mpeg4 audio). Hence, video could be played from sites like YouTube but not audio.

4.2 Block Diagram

The system mainly consists of three entities:

1. **Web Server** - This server hosts all the images and audio for every video file. It also has a Java Servlet that handles all the user queries.

2. **Mobile Client** - This is a J2ME application that runs on the mobile, which the user uses to see videos.

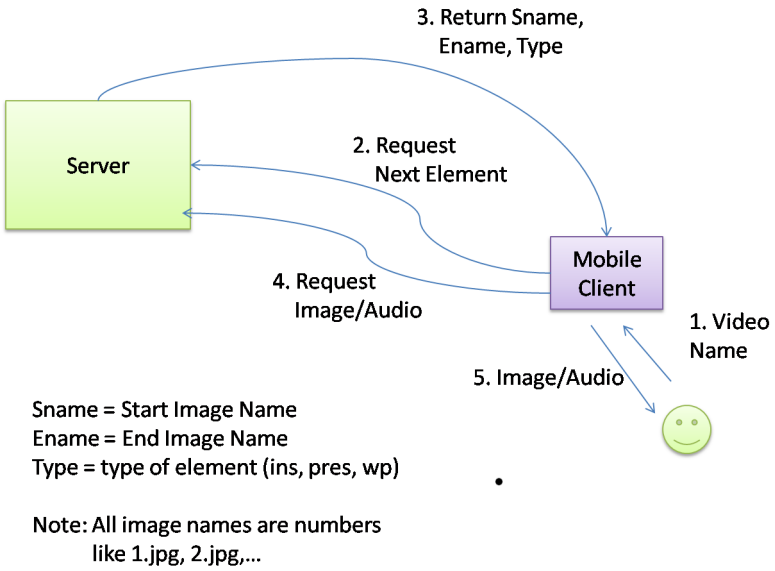


Figure 4.1: Request Response Pattern on Video Request

3. **User** - The user who sees the video.

As shown in Figure 4.1, when the user request a video name, the client application requests the first element details from the server which includes the starting image name(all image names are numbers like 1.jpg 2.jpg), ending image name and the type of the element that these images are part of. Then, the period for that element type is fetched from the server by supplying the server with the user experience requested by the user for that type of element.

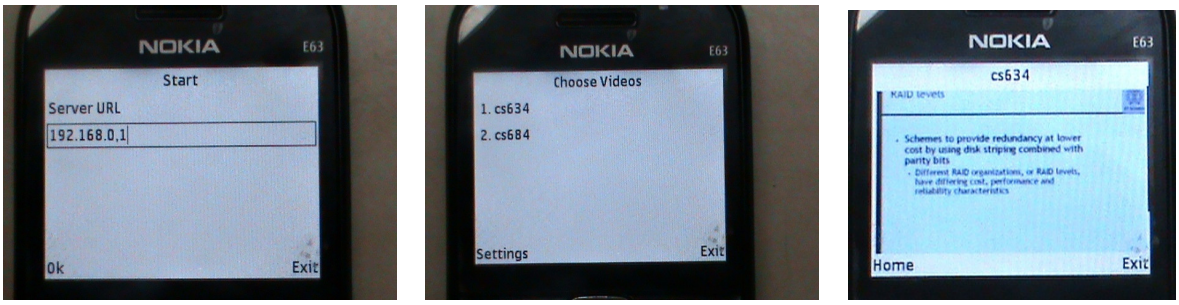
Once the period is obtained for that element, images are requested from the server, one-by-one according to the period. Once all the images in the element are shown, then the details of the next element in the video is requested and the process continues.

In the background, audio is continuosly played, by a form of progressive download. The server hosts the audio as small chunks of 20 seconds each. The client app downloads the 2nd chunk while playing the first chunk and so on.

4.2.1 Watching Video

As shown in Figure 4.2 , the user initially enters the http URL of the server. He is then shown a list of videos available on the server, on which he chooses one of the videos. Once the user selects a video, he/she would see a slideshow of images along with background audio. The images would change at an interval as per the user experience

desired by the user for each element. The default user-experience values are set to suit GPRS bandwidths.



(a) Start Page (b) Video List (c) Slideshow of Images

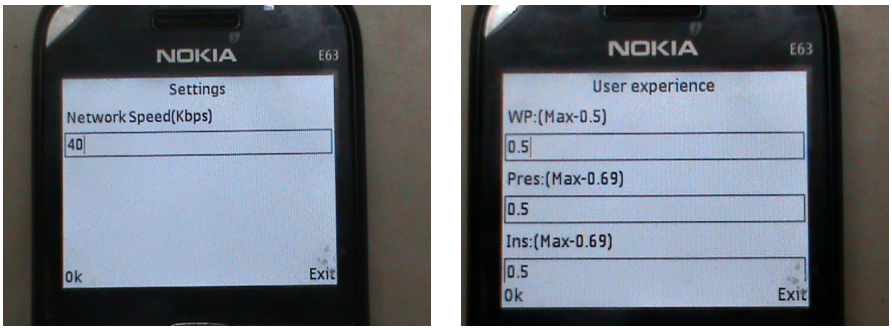
Figure 4.2: Seeing a Video

4.2.2 Changing Settings

The settings in the mobile application, consist of:

1. Network Bandwidth (Kbps)
2. User Experience of each element (0-1)

The default values for these settings already exist in the applicaition, and are tuned for GPRS access. If however, the user wishes to do so, he can change the settings on the video list screen.



(a) Network Bandwidth Screen (b) User Experience Screen

Figure 4.3: Settings of Network Bandwidth and User Experience

Figure 4.3 shows the screens the user sees, when he presses the settings button. In the first screen, the user has to enter the network bandwidth available to him. The

application then contacts the server to fetch the maximum permissible user experience value less than or equal to 1, at this network bandwidth.

The user then has to enter the desired value of user experience for each element, and the mobile application then contacts the server with these values, to fetch the interval at which the images have to be shown.

Chapter 5

Conclusion and Future Work

In conclusion, it can be said that lecture videos can be adapted so that they can be viewed on networks with low-network bandwidths and cost-constraints. Further, identifying study-elements within the lecture video, enables defining user-experience of the adapted video, thereby making the adaptation process customizable also.

As future work, there are few ideas that can be explored as given below:

1. **Collaborative Tagging** - As a supplement to the tagging methodology used currently, the viewers of the actual video files can be given an option to tag regions present in the video as one of the three elements. This serves two purposes:

   (a) The accuracy of the tagging methodology can be improved

   (b) The element instances within the video that are most frequently viewed by the users, can be identified and then can be stored as a seperate new element in itself.

2. **Presentation slides instead of images** - Currently, the section of the video where presentation slides are shown (i.e. presentation study elements) is of poor quality. We feel that the quality can be improved if the actual presentation file like a Power Point file can be sent to the client device and the corresponding slide be shown there whenever a presentation element occurs in the video. This would involve identifying the mapping between the slides shown in the video and the slides in the ppt file.

3. **Content Region Identification** - Quality of written text can be improved if the exact bounding rectangle of the content can be found in the video frames and the

images can be cropped to that extent. By doing this, we could actually encode the smaller region at high resolution thereby improving the quality.

4. **Conversion to text** - Since text is very easily supported by mobile devices, the user-experience and quality of viewing would be very good, if written material from the video can be converted to text, and be shown. Ofcourse, there would be certain complications like converting of equations.

Bibliography

[1] H.264 white paper. http://ati.amd.com/products/pdf/h264_whitepaper.pdf.

[2] MPEG background. http://bmrc.berkeley.edu/frame/research/mpeg/mpeg_overview.html.

[3] Secure, Portable, and Customizable Video Lectures for E-learning on the Move. http://www.informatica.si/PDF/33-1/16_Furini%20-%20Secure,%20Portable,%20and%20Customizable%20Video%20Lec.pdf.

[4] Video compression standards. http://www.cctvone.com/pdf/FAQ/Video%20Compression%20Standards%20Journal.pdf.

[5] Daniel Keysers Thomas M. Breuel Adrian Ulges, Christian Schulze. Content-based video tagging for online video portals. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.1168.

[6] Qiyam Tung Quanfu Fan Juhani Torkkola Ranjini Swaminathan Kobus Barnard Arnon Amir Alon Efrat Chris Gniady Andrew Winslow. Stuyding on the move - Enriched Presentation Video for Mobile Devices. http://kobus.ca/research/publications/09/movid-09.pdf.

[7] Scott Bateman, Christopher Brooks, and Gordon Mccalla. Applying collaborative tagging to e-learning. 2007.

[8] 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.

[9] 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.

[10] Stephan Kopf, Fleming Lampi, Thomas King, and Wolfgang Effelsberg. Automatic scaling and cropping of videos for devices with limited screen resolution. In Klara Nahrstedt, Matthew Turk, Yong Rui, Wolfgang Klas, and Ketan Mayer-Patel, editors, *ACM Multimedia*, pages 957–958. ACM, 2006.

[11] Marcelo G. Manzato and Rudinei Goularte. Live video adaptation: a context-aware approach. In *WebMedia '05: Proceedings of the 11th Brazilian Symposium on Multimedia and the web*, pages 1–8, New York, NY, USA, 2005. ACM.

[12] Adriana Reveiu, Ion Smeureanu, and Marian Dardala. Content Adaptation in Mobile Multimedia System for M-Learning. In *ICMB '08: Proceedings of the 2008 7th International Conference on Mobile Business*, pages 305–313, Washington, DC, USA, 2008. IEEE Computer Society.

[13] Wallapak Tavanapong and Srikanth Krishnamohan. A Characteristics-Based Bandwidth Reduction Technique for Pre-recorded Videos. In *IEEE International Conference on Multimedia and Expo (III)*, pages 1751–1754, 2000.

[14] A. Vetro, C. Christopoulos, and Huifang Sun. Video transcoding architectures and techniques: an overview. *IEEE Signal Processing Magazine*, 20(2):18–29, March 2003.

[15] Y. T. Zhuang, Y. Rui, T. S. Huang, and S.Mehrotra. Adaptive key frame extraction using unsupervised clustering. In *ICIP*, pages I: 866–870, 1998.