Design Analysis for Ensuring Multimedia Qos over Scarce Resource Network Applications

Rohit Kumar[1] Kishan Singh Rawat[2]
[1]Assistant Engineer [2]Assistant Professor
[2]Department of Electronics & Communication Engineering
[1]Uttarakhand Council for Biotechnology, Biotech Bhawan, haldi, Udham Singh Nagar, Uttarakhand-263146 [2]Amrapali Institute, Lamachaur, Haldwani, Uttarakhand-263139

Abstract— The existing multimedia software in e-learning does not provide par excellence multimedia data service to the common user; hence e-learning services are still short of intelligence and sophisticated end user tools for visualization and retrieval. Delivery of video in the presence of bandwidth constraints is one of the most important video processing problems. The work presents an innovative and complete end-to-end solution for e-learning multimedia delivery package for the client-server model. An efficient approach has been proposed to achieve content based embedded scalable motion compensated video compression, for the e-learning videos, to save and/or efficiently utilize the network bandwidth. The serious bottleneck in multimedia e-learning is the non-availability of required bandwidth. While much research has been done on multimedia compression and network resource optimization, multimedia delivery is still a big concern especially for scarce resource networks. The virtues of network aware video coding and application of domain specific compression techniques have not yet been fully explored. This work deals with the issue of intelligent transmission of important segments of the video sequence over scarce resource networks. We incorporate the knowledge of the network conditions to determine how various parts of the video frames are encoded. An estimate of the available network bandwidth is obtained which is then distributed optimally between the different frame constituents based on their relative importance and motion by the bandwidth allocation module. We have devised a simple approach to classify the videos into visual objects and used Discrete Wavelet Transform based Color Embedded Zerotree Wavelet (CEZW) coding to obtain a scalable bitstream that provides dynamic response to changing network conditions. The above approach helps to reach an optimal trade off between perceptual quality of the video and the available network bandwidth. The results are compared with the state of the art compression and transmission schemes over the network to illustrate the superior performance of our approach.

Key words: QoS, video encoding, e-learning, CEZW coding, multimedia compression

I. INTRODUCTION

The demands for the multimedia services are rapidly increasing while the expectation of quality for these services is becoming higher and higher. To attain the highest possible quality, analog signals such as speech, audio, image, and video, are sampled and digitized for storage, transmission, and reconstructed at the receiving ends in order to be free from noise etc. However, these digitized data are usually voluminous. Even though the technology is continuously progressing and pushing up the bandwidth limit and reducing the transmission storage cost, still the channel bandwidths and storage capabilities are limited and relatively expensive in comparison with the volume of these raw digital signals. To make all the digital services feasible and cost effective, data signal compression is an essential step. The main objective of compression is to retain as little data as possible that is sufficient to reproduce the original images without causing unacceptable distortion of the image.

Compression relies on the fact that information, by its very nature, is not random but exhibits some order and patterning. If this order and patterning can be extracted, the essence of the information can often be represented using less data than would be needed for the original data. For video systems, compression reduces the volume of data by exploiting spatial and temporal redundancies of the picture and by eliminating the data that cannot be displayed suitably by the associated display or imaging device. Further, much higher compression ratio can be attained using content-aware video compression algorithm.

E-learning and streaming media is becoming increasingly prominent on the internet. Many institutes, such as MIT (USA) and IITs (India), have opened their web servers for free lecture-on-demand on several courses [1-4]. The Open Courseware Consortium[5 is working to provide free lectures on demand to users worldwide. MHRD, Govt. of India has started a project called NPTEL, National Program for Technology Enhanced Learning[4] to enhance the quality of engineering education in India by developing curriculum based video and web courses. Various developed countries have reserved a big proportion of education funds to support e-learning to enhance the educational exports [6]. [7-9] discuss the potential and increasing popularity of e-learning. The concept of remote laboratories also demands real-time multimedia content delivery [10-12]. Research is going on to provide real time streaming support for teleconferencing applications [13,14].

However, a serious bottleneck towards multimedia e-learning is the non-availability of required bandwidth to view the lecture videos at good resolution because of their large size. Other challenges in video streaming are the dynamic change of bandwidth (as in wireless and dial-up networks), packet loss, and the differences of video content and users' preferences [15]. Although a lot of work has been done on content-based classification [16-18], content-based streaming [19,20], bandwidth adaptation and on network issues [21,22], a complete framework that addresses all these issues to provide an end-to-end solution for e-learning videos does not exist. Meeting bandwidth requirements and maintaining acceptable image quality simultaneously is a challenge.

General video coding standards and formats like MPEG-1, MPEG-2, and H.261 etc. achieve a high rate of video compression but educational videos are not separately dealt by these standards. Continuous rate scalable applications can prove valuable in scenarios where the channel is unable to provide a constant bandwidth to the application. Such decoders are particularly attractive because of their flexibility in allowing only one image or sequence to be stored in the database, avoiding the overhead of maintaining several coded images or sequences at different data rates. A specific coding strategy known as color embedded rate scalable coding is well suited for continuous rate scalable applications [23,24]. This coding scheme generates a bitstream such that the reconstructed image quality progressively improves as more and more symbols from the encoded stream are used for decoding.

[25] presents a novel rate control scheme to allocate the bits between texture and non-texture parts of the image. In this paper we propose a new approach for network aware optimal resource allocation. This approach, applicable to all low motion videos is exploited and explained in this paper for educational videos.

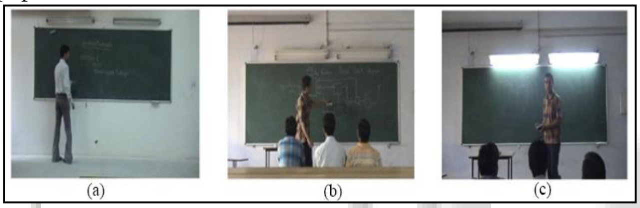

Fig. 1: Snapshots of classroom Lecture sessions

Some of the issues considered in this project are:

a) The solution must enable a streaming server to perform real-time processing
b) It should generate an optimal bitstream that is best perceptually based on the estimates from the network
c) It must enable low-complexity decoding with low-memory requirements
d) The scalable bitstream must be resilient to packet loss events.

II. Background and Related Work

Different approaches have been proposed to provide service guarantees to multimedia applications. These approaches can be broadly divided into network level protocols, reservation-based schemes, and adaptation-based schemes for QoS support. This section discusses the basics of video processing and computer networks and also presents a literature review of recent developments in these fields.

A. Video Compression Scheme

Transform coding has been a dominant method of video and still image compression. It takes advantage of energy compaction properties of various transforms (such as DCT, DFT, DWT, etc.) and properties of the Human Visual System to minimize the number of useful coefficients. DCT has been the popular choice for image and video processing schemes [26,27].

1) DCT: The Discrete Cosine Transform

Since its conception in 1974, the Discrete Cosine Transform (DCT) [26,28] is the dominant transform for video and still image compression applications. The DCT produces nearly optimal de-correlated frequency domain coefficients, only slightly surpassed by the theoretically optimal and computationally intensive Karhunen-Loeve Transform. Due to its most desirable properties, the DCT has been part of virtually every video and still image compression published standard, i.e. JPEG, MPEG-1, MPEG-2 etc.[28]. DCT is basically a separable transform, i.e. (n+1)-dimensional DCT coefficients can be attained by applying 1-dimensional DCT over n-dimensional DCT coefficients. Usually DCT is used as a block transform for compressing a 2Dimensional image. Most current image compression algorithms operate on 8 x 8 blocks pixels, because it provides a good trade-of between moderate computational requirements and efficient temporary redundancy reduction [22].

2) DWT: The Discrete Wavelet Transform

In image processing, the DWT(Discrete Wavelet Transform) is obtained for the entire image, and it results in a set of independent, spatially oriented frequency channels or subbands. The wavelet transform is typically implemented using separable and possibly different filters. It allows localization in both the space and frequency domains.

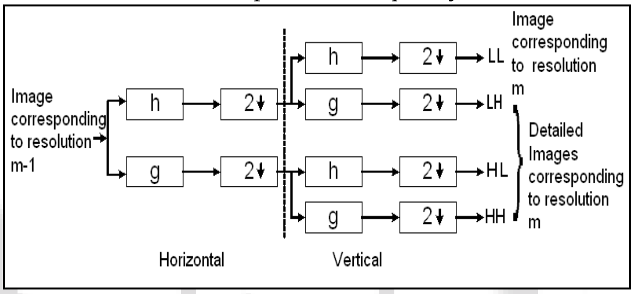

Fig. 2: One level of Wavelet Decomposition

Typically the full image is decomposed into a hierarchy of frequency sub-bands. The decomposition is achieved by filtering along one spatial dimension at a time to effectively obtain four frequency bands. The lowest subband (LL), represents the information at all coarser scales (as shown in Figure.) and it is decomposed and subsampled to form another set of four subbands.

This process can be continued until the desired number of levels of decomposition is attained. Two analysis filters, namely g and h, carry out the decomposition, into independent frequency spectra of different resolutions, producing different levels of detail. Formulation of sub-bands does not cause any compression (same number of samples are required to represent the subbands as is for the original image), but arranges the data in more efficiently codable format. Figure 2 shows one level of wavelet decomposition.

Several efficient coding schemes have been used for coding DWT coefficients. Embedded Zerotree Wavelet (EZW) introduced by Shapiro[23] is one such coding scheme for gray scale images. It exploits the interdependence between the coefficients of the wavelet decomposition of an image, by grouping them into spatial orientation trees (SOT). It outputs an embedded bitstream. An embedded bit stream can be truncated at any point during decoding, and can be used to obtain a coarse version of the image. Decoding additional data from the compressed bit stream can then refine this version.

3) Color Embedded Zerotree Wavelet Scheme

For color images, the same coding scheme can be used on each color component. However, this approach fails to exploit the interdependence between color components. It has been noted that strong chrominance edges are accompanied by strong luminance edges [29,30]. However, the reverse is not true, that is, many luminance transitions are not accompanied by transitions in the chrominance components. This spatial correlation, in the form of a unique spatial orientation tree (SOT) in the YUV color space, is used in a technique for still image compression known as Color Embedded Zerotree Wavelet (CEZW) [31-33]. CEZW exploits the interdependence of the color components to achieve a higher degree of compression. The parent child dependency in CEZW is illustrated in figure 3.

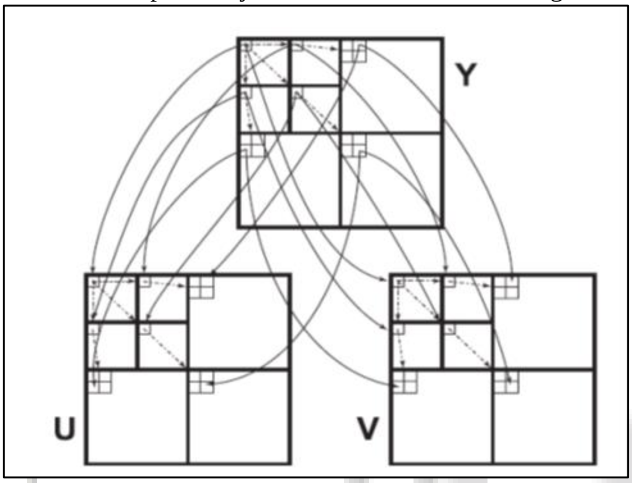

Fig. 3: Parent Child Dependency in CEZW scheme

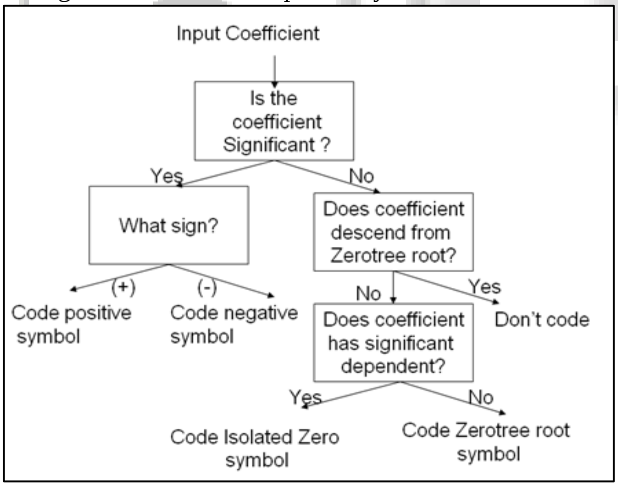

Fig. 4: Flow diagram of CEZW coding algorithm

4) Motion Estimation and Motion Compensation

In image and video coding schemes image is divided into small blocks for operation by prediction techniques. Motion estimation is used to determine the movement of a macroblock from the reference frame to the current frame. Motion is estimated by searching for the macroblock in the reference picture that provides the closest match, as shown in figure 5. The difference between the values of both the macroblocks is coded for reconstruction at the decoder. To reduce the distortion between the decoded and the original picture, the encoder uses a reconstructed reference frame to perform motion estimation. This reconstructed reference frame is same as used at the decoder side.

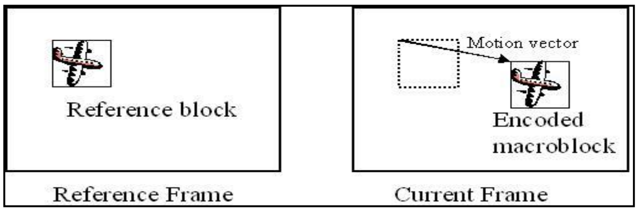

Fig. 3: Motion estimation of a block

Motion estimation computes one motion vector per macroblock. Usually, the search is conducted for the luminance component only.

A predictive frame is constructed from the motion vectors obtained for all macroblocks in the frame, by replicating the macroblocks from the reference frame at the new locations indicated by the motion vectors. The difference between the values of the predicted and the current frames, known as predictive error frame (PEF), is then encoded using the same procedure as for an intracoded frame. The frame obtained by adding the predictive frame to the PEF is known as the reconstructed frame. The energy of the PEF is low, thus many coefficients are zero, reducing the number of bits needed to encode the frame.

5) Frame Packaging

Video compression system identifies three types of frames, depending on the location of all reference block, I-frame (Intra-coded frame), P-frame (predictive frame), and the B-frame (bi-directionally predictive coded frame). The video frame is efficiently coded for the I-frame by motion estimating within the same frame (to reduce spatial redundancy); whereas, the location of target block for motion estimation/compensation decides the type of coded frame (P-frame or B-frame). P-frames are coded relative to a temporarily preceding I or P frame and B-frames are coded relative to the nearest previous and/or future I and P frame. Reasons for coding B-frames are: 1) better matching of blocks, as it can be matched in reverse as well as forward direction, hence lower bpp (bits per pixel) after compression and 2) B-frames can be filtered out from the video stream for lowering the frame rate to improve the load condition of under-performing network, without considerably affecting the video quality at client side.

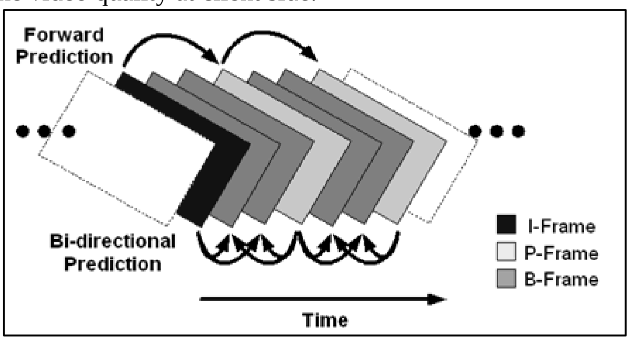

Fig. 6: Temporal relations between I, P and B

B. Multimedia Networking Issues

The availability of multimedia resources places new demands on the service that a network must provide. The most important of these are the bit error rate, the packet or cell loss, delay and delay variation. Network resources need to be committed to multimedia data streams to accommodate the peak bit rate, mean bit rate and burstiness of the data stream. Compression algorithms and techniques are critical to the viability of multimedia networking.

Uncompressed digital television requires about 140 Mbps. Since few users have this sort of network access compression is the only hope for the widespread deployment of digital video and multimedia.

While compression can ease the demands on networks and storage media there are several trade-offs. Since some compression techniques remove information considered to be less important a loss in resolution may result. The cost of complex hardware and software and compression and decompression delay are other factors important to users. Different uses require different compression methods. Video conferencing must be done in real time so fast encoding and decoding is needed. This is the aim of the H.261 standard. Video film distribution via cable networks, radio or CD is essentially a playback process, so encoding is not time critical, and decoding should be easy to implement to reduce consumer costs. The MPEG standards address these applications.

Multimedia applications, particularly those using video and images demand large bandwidths. However bandwidth for the foreseeable future will be limited. The limitations arise from the cost of installing optical fiber transmission, terminal equipment complexity and speed, tariffing regimes, switching speeds, and increasing numbers of users sharing equipment and networks. In Indian context, bandwidth is a very crucial factor. Poor network infrastructure in most areas implies the need of video transmission over poor network conditions with sufficient quality.

Multimedia also makes new demands on the workstations used to reproduce audio and video. Processor speeds, operating systems, displays, storage medium and network interfaces and application must all be capable of handling multimedia.

1) Existing Internet Protocols

The Internet is a more of a phenomenon than a network, but is important when discussing multimedia because a popular Internet Application, the World Wide Web is capable of accessing and displaying multimedia formats such as pictures, audio and video. The current Internet has thrived and grown due to the existence of TCP implementations for a wide variety of classes of host computers. These various TCP implementations achieve robust inter-operability by a "least common denominator" approach to features and options.

The system of connected networks which comprise the Internet has also been used to carry live audio and video. Extensions to the TCP/IP protocols currently used have been proposed as Real Time Protocols (RTP). Broadcast of audio and video has taken place on the Multicast Backbone (MBONE), by allocating higher priority to audio and video information from within routers.

The MBONE is being developed as a technology for low cost multimedia. Multicasting within MBONE enables multiple destinations to share the same information without replication. Internet routers and workstation software require some modifications to support multicasting. A virtual network has been implemented over the IP network to bypass routers which do not support multicasting, and to enable some bandwidth to be reserved for multicasting. However audio and video on the MBONE must still compete with other traffic on parts of the network. This limits the quality of both the voice and video obtainable.

However current transport protocols exhibit some severe problems for high performance, especially for using hardware support. Existing protocols require a processing overhead which takes longer than the transmission time on high speed networks. For example, TCP places the checksum in the packet header, forcing the packet to be formed and read fully before transmission begins. ISO TP4 is even worse, locating the checksum in a variable portion of the header at an indeterminate offset, making hardware implementation extremely difficult.

Special purpose transport protocols have been developed. Examples include special purpose transport protocols such as UDP (user datagram protocol), RDP (reliable datagram protocol), NVP (network voice protocol), PVP (packet video protocol) and XTP (Xpress Transfer Protocol), XTP fixes header and trailer sizes to simplify processing and places error correction in the trailer so that the code can be calculated while information bits are being transmitted.

2) Content-Based streaming & Issues

Liu et.al. [32] provides a real-time content analysis method to detect and extract content regions from instructional videos and then adjusts the Quality of Service (QoS) of video streams dynamically based on video content. Similarly, [20] uses a content based retransmission scheme. However, these methods identify the video frames based on their content, they will not prove useful in cases where a static camera is used for recording the lecture session, as is usually the case. In this case, all the frames will be either content frames or non-content frames and thus, discriminative frame rate or frame quality control cannot be applied effectively. In videos with static camera, there is a need to segment an individual frame into objects so that special coding techniques can be applied to each of them.

An approach to retrieve content pixels from the board regions by statistical modelling and classification is presented in [33]. However this method is computationally expensive and does not segment the video into a distinct number of visual objects among which the network bandwidth can be distributed. Similarly, [34] presents a scheme for knowledge-assisted image analysis using fuzzy logic and region growing algorithms. We have used a simple approach to achieve real time yet robust e-learning video classification. Moreover, the regions having instructional content obtained by the other approaches have arbitrary dimensions which cannot be directly used with the CEZW compression scheme or ISU/ITU standards like MPEG-1, H.261 etc[28].

The following factors should be considered in selecting a layered coding technique for packet video transmission.

- Overall Video Quality: In all cases, layered coding has some coding penalty which, under a constant bit-rate constraint, reduces the video quality compared to non-layered coding. The penalty is the smallest for data partitioning and the largest for spatial scalability. If both layers will be received most of the time, this is an important factor.
- Base Layer Video Quality: If, due to severe losses, only the base layer will be received for a significant portion of time, of if and application must use only the base

layer for a special purpose, e.g., reduced resolution video for picture-in-picture, this factor becomes important. The video quality of the base layer is determined not only by the rate allocated the base layer, but also by the layered coding technique used. For example, for a given bit rate, the decoded video quality of the base-layer only is best for spatial scalability and worst for data partitioning.

– Rate Partitioning Range: If dynamic rate partitioning is needed for traffic shaping, having a usable base-layer video for a large range of base bit-rates is important. Data partitioning can create a usable base layer picture only for a small range of base-layer bit-rates, while spatial scalability can produce reasonable base-layer pictures for a large range of base-layer bit rates.
– Video Quality with Cell Losses: If cell losses are moderate, the resulting video quality depends on both the overall lossless quality and the base-layer quality. In general, this is a nonlinear relationship, which is difficult to quantify analytically.
– Complexity of Implementation: The increase in implementation complexity due to layered coding ranges from addition of a simple multiplexer and demultiplexer in the case of data partitioning to the use of two coupled codecs for spatial scalability.

3) *Bandwidth Estimation*

Information about the bandwidth availability in the network helps to adaptively adjust the transmitted data according to its importance from the point of view of the user. Much research has been done on bandwidth estimation and a number of tools have been proposed which can be used with our approach. These set of tools can be distinguished according to the two main approaches underlying the estimation techniques:

a) The probe gap model (PGM): It exploits the information in the time gap between the arrivals of two successive probes at the receiver. Works of Strauss[35] and Ribeiro[36] are example of tools that use the gap model.
b) The probe rate model (PRM): This model is based on the concept of self-induced congestion. Tools such as Pathload[37] and Pathchirp[38] use the probe rate model.

Majumdar's work on packet pair [39] is the earliest attempt to estimate the available bandwidth using measurements conducted at the end hosts. Packet pair assumes Fair Queuing in the routers and as a result cannot estimate the available bandwidth in the current Internet.

Cprobe is a pioneering tool for estimating the available bandwidth using end-to-end measurements. Cprobe doesn't assume fair queueing. Instead of using a pair of packets, Cprobe sends a short train of ICMP packets and computes the available bandwidth as the probe traffic divided by the interval between the arrival of the last ICMP ECHO and the first ICMP ECHO in the train. A similar approach is used by Hu et.al. [40]. Jain et al [41] show that these techniques measure a metric called the Asymptotic Dispersion Rate (ADR), which is related to the available bandwidth but not the same.

III. SYSTEM OVERVIEW

The designed classroom video streaming architecture gives an interactive interface for user and also a practical adaptive backbone framework for on demand video (ODV) streaming and controls adjustments. Our primarily focus is on ODV streaming over a low and varying bandwidth network so that the best content reception is available to the student. The following two subsections provide an overview of the outer shell and knowledge of different blocks of the system. Subsection 3.1 discusses the backbone framework while 3.2 briefly discuss the client side set up. The ODV streaming application is presently experimented for classroom lectures setup.

A. *Video Compression and Transmission Framework*

The overall system overview is presented in figure 9. The edge processing based content retrieval scheme is effective as it works in real time and is robust. The CEZW coding of images has been implemented after conversion of frames to YUV space and down-sampling by 4:2:0. Our system provides end to end delivery of Multimedia with best perceptual quality. The interface design has been chosen very simple. Moreover, the user has the choice to just have the blackboard zoomed view in case of scarce network conditions. The unwanted motions of the classroom (like the movements of students which have no academic value) can be suppressed in case of low bandwidth to enable high quality reception of information content by student at remote site.

We have classified the classroom lectures into three main visual objects (VO): black board, teacher and the remaining background. The different uncompressed VOs are then packaged by a modified scheme. We classify the even and the odd frames separately. The I, B and P frames are packaged and processed. Motion compensation and estimation is then followed by DWT based CEZW+ encoding and adaptive arithmetic encoding. The CEZW) coding uses EZW over three color components while scaling the U and V.

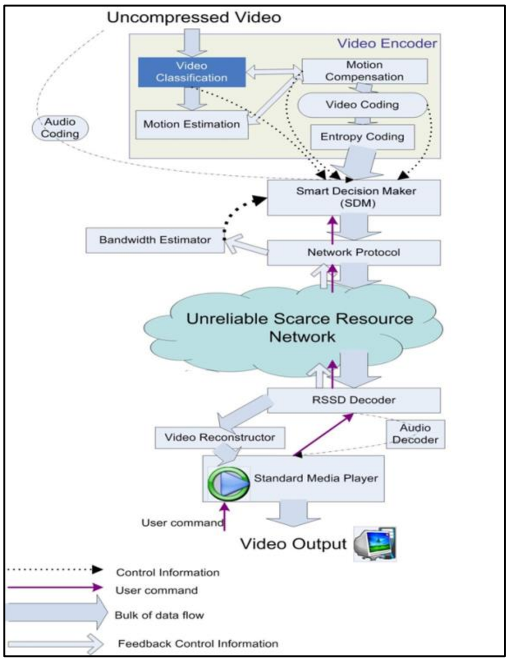

Fig. 7: System overview for Network aware video streaming framework

components by a scale factor (usually 5) to achieve further compression. The base and enhancement layers are discriminated for each VO based on its significance. Motion compensation scheme is resilient to changing quality of VOs since it takes the base layer reconstructed frame of parent as its parent frame. The bit stream for each individual VO is decided by the Smart Decision Maker (SDM) block incorporated into the system. SDM receives an estimate of the available network bandwidth and then it optimally distributes bandwidth to different VOs based on relative significance and current motion in each VO.

1) Classification Module

A typical classroom lecture has Teacher Writing On Board (TWOB) sessions. These sessions have a teacher (complete view) writing on Black Board and explaining the class points to the students. Other types of frames constitute a very low portion of video, have low motion and can be directly coded. For the sake of brevity we limit our discussion to TWOB sessions only.

TWOB frames possess some typical characteristics which help in determining them from other frames: The complete board is visible, motion is slow, some students may be visible in the view and a significant part of the tutor's body may be visible in the frame (see Fig.1). The height of teacher in these frames generally doesn't vary much. Thus they can be easily identified. They are classified into three visual objects: background, Black Board and the tutor. In this work, we have modeled the background for static camera.

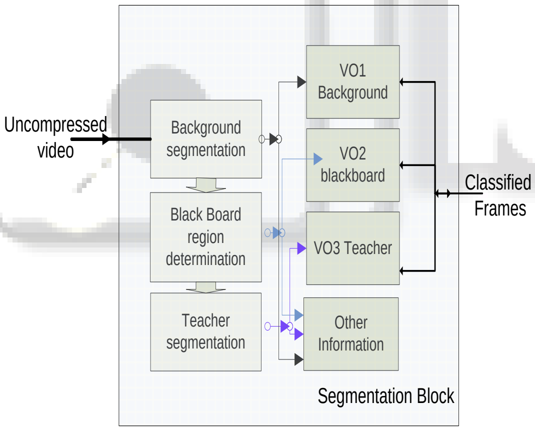

Fig. 9: Classification scheme

Figure 9 shows the output VOs for a single frame. Thus the motion and non motion regions have been classified. The teacher region is the moving region and of most attention to the student besides blackboard. Figure 9 summarizes the classification process.

2) CEZW+ coding details and Frame Packaging

[29-31,46-48] discuss the implementation of CEZW scheme. The compressed bit stream consists of the initial threshold T, followed by the resulting symbols from the dominant and subordinate passes, which are entropy coded using an arithmetic coder. We have used a modified scheme. The Y and UV components of the VOs are coded at different bit rates. If BPP_{total} is the overall bit rate available for coding the VO, and if we code the U and V components at the same bit rate, then we can write

$$BPP_Y = BPP_{total}\left[1+\frac{2}{s}\right]^{-1}$$, and

$$BPP_U = BPP_V = \frac{BPP_{total}}{s}\left[1+\frac{2}{s}\right]^{-1}$$

Where BPP_Y, BPP_U, BPP_V are the bits per pixel available to the Y, U and V components respectively. s is the scaling factor and is the ratio of the bits per pixel available to the Y component to the bits per pixel available to the U and V components.

We use the peak signal-to-noise ratio (PSNR), based on the mean-squared error (MSE), as our "quality" measure. The PSNR of a YUV image is obtained by using the following equation:

$$PSNR = 10\log_{10}\left[\frac{255^2}{\frac{MSE_Y + MSE_U + MSE_V}{3}}\right]$$

Here, MSE_Y, MSE_U, and MSE_V are the mean square errors of the Y, U, and V component of the reconstructed frame with respect to the original frame. Figure 10 explains the working of the compression block. The input VO frame is read as R,G and B components and converted to Y,U,V color space and downsampled.

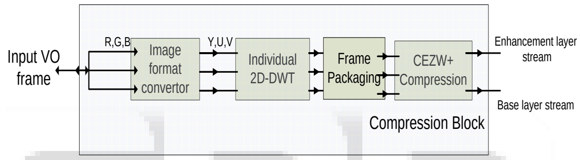

Fig. 10: Compression and Frame Packaging block schematic

The frame packaging scheme used is similar to MPEG-1 standard[28]. Then we apply our CEZW+ scheme to obtain the base and enhancement layer data stream. The enhancement layer stream is scalable and used to adjust data flow to changing network conditions.

3) SDM: Smart Decision Maker Module

Dynamic bandwidth allocation module called as SDM (Smart Decision Maker) tracks the unpredictable bandwidth variations and then adaptively manages the bandwidth allocation to the different VOs according to relative importance and their present perceptible quality. Thus the network traffic is decided based on

a) Desired PSNR threshold for the VO
b) Energy of current enhancement layer stream for the VO
c) Available network bandwidth

SDM uses UPD socket for video data transfer and TCP connection to pass information like user preferences, packet acknowledgement etc.UDP is needed for effective transfer of video data over an unreliable network. But UDP doesn't make the provision for packet acknowledgement but that is important for Bandwidth estimation and other parameters. Hence the TCP based connection sends back the acknowledgement information. The NBE (Network Bandwidth Estimation) tool used was developed using network profiling which helps in the accurate estimation of the available bandwidth.

B. Client Interface

The client interface is designed simple and the decoder operation is reverse to the encoder structure. It reconstructs the various VOs and then adds them to reconstruct the frame. The client interface is computationally less intensive and simple.

Thus, the processing unit at client performs the following operations:

1) It responds to the server to respond to quality to data delivery.
2) The Video is transmitted in our scheme as a combination of several visual blocks (VOs). The integration and synchronization of different VBs, background and audio stream is another task for CDPU. It also performs smoothening of reconstructed frames.

The various inputs are plotted against each other at user's discretion to observe the performance of system. For example it may plot the step response of the given system using the inputs and outputs. The plotting is done at client side to reduce network traffic and also to reduce the overhead for server which may be required to run several laboratory experiments for different users.

IV. Implementation

The implementation and testing was carried out in MATLAB. Then we tried to implement the code in Microsoft Visual C++ 2005 to achieve real time processing.

A. *Steps That Led To the Final Design*

The following steps were involved:

Step 1: Development of the segmentation module.

Step 2: Development of the CEZW coder and decoder.

Step 3: Use of DirectX and OpenCV to obtain individual image frames of a video sequence. See figure 11.

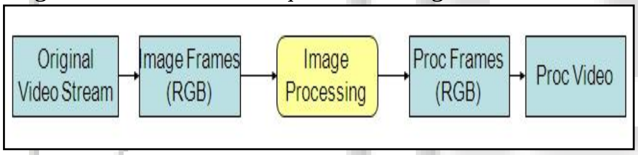

Fig. 11: Obtaining individual image frames to process

Step 4: Integration of the CEZW coder and decoder with the image processing part of step 3. CEZW coder was developed as in step 2. See figure 12.

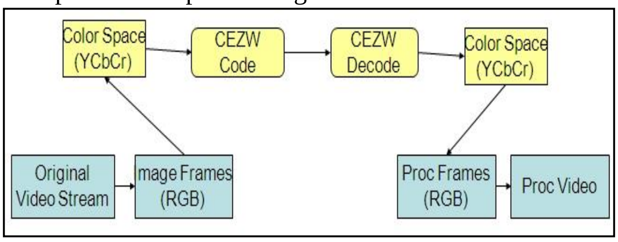

Fig. 12: Integration of CEZW codec

Step 5: Development and Integration of the network streaming module as shown in figure 13. Currently, we store the CEZW coder output on the server and stream it at different bitrates as per the bandwidth availability at the receiver.

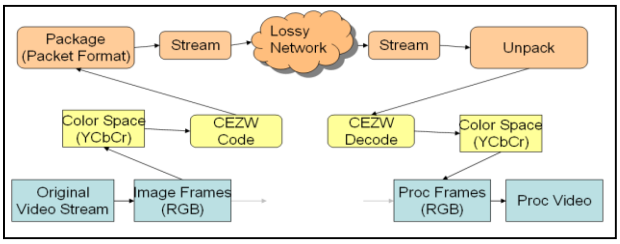

Fig. 13: Development and Integration of Network Streaming Module

B. *Code Development Model*

The code development was done using Modular approach. The project was divided into different modules and each module was independently developed. Figure 14 shows the different modules.

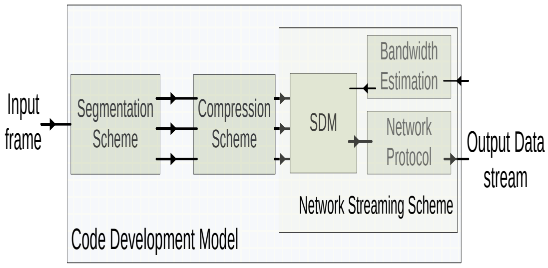

Fig. 14: Overview of Code Development Model

1) *Segmentation Module*

The segmentation module was developed and tested in Matlab. It segments each video frame into three Visual objects. The GUI developed for implementation is shown in fig.18. Several techniques like using edge detection based approach, Kalman filtering approach for motion prediction, contour mapping based approach for object approach for object segmentation etc were applied and the optimal results obtained.

2) *Compression Module*

The compression module was first implemented in Matlab. Block and image processing was done using the inbuilt Matlab image processing toolbox functions. The Matlab prototype is not real time as it takes around 1 second for decoding an individual frame. Also, it does not deal with the audio stream. The utility of this prototype is reduced further if we consider the fact that all the users of this software will not have Matlab on their machines. Also, Matlab has large memory requirements which restrict its usage for

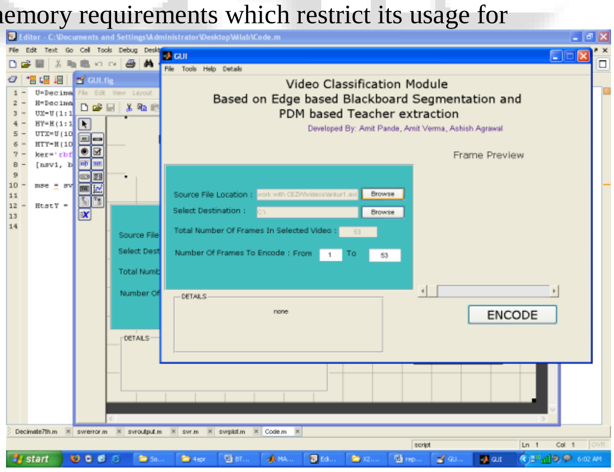

Fig. 15: Matlab Implementation of Classification Module

research work only. Thus, another optimized, realtime implementation of the software on Visual C++ was required. The project was then implemented in Visual C++ for real time functionality and further widespread usage.

The following packages and libraries were used in the implementation:

– Microsoft DirectShow: DirectShow is a multimedia framework and API produced by Microsoft for software developers to perform various operations with media files. DirectShow provides a common interface for media across many of Microsoft's programming languages, and is an extensible, filter-based framework that can render or record media files on demand at the behest of the user or developer. Earlier DirectShow Extras were a part of DirectX SDK, but now they are available as a separate package.

– Intel Image Processing Library (IPL): It is a set of highly optimized C functions that implement image processing functions on Intel architecture processors. It focuses on taking advantage of the parallelism of the new SIMD (single-instruction, multiple-data) instructions of the latest generations of Intel processors. These instructions greatly improve the performance of computation-intensive image processing functions.
– Open Computer Vision (OpenCV) libraries: It is a collection of C functions and a few C++ classes that implement many popular Image Processing and Computer Vision algorithms. It is a wrapper to Intel Image Processing Library. It simplifies the usage of Intel Image Processing Library and is also useful in removing memory issues.
– DWT and EZW libraries

Figure 16 shows the DirectShow filter graph. DirectShow has been used firstly for obtaining separate audio and video streams from the audio. Then we decode the video by adding a DirectShow video decoder filter to the filter graph. The decoded video has to be handled frame by frame for compression and streaming operation. DirectX is implemented using the Microsoft COM technology. This means that when we initialize something, we have to do it by using a given COM interface. In order to initialize the COM layer, we call: CoInitialize(NULL);And similarly, when we are done with COM, we need to uninitialize it: CoUninitialize();

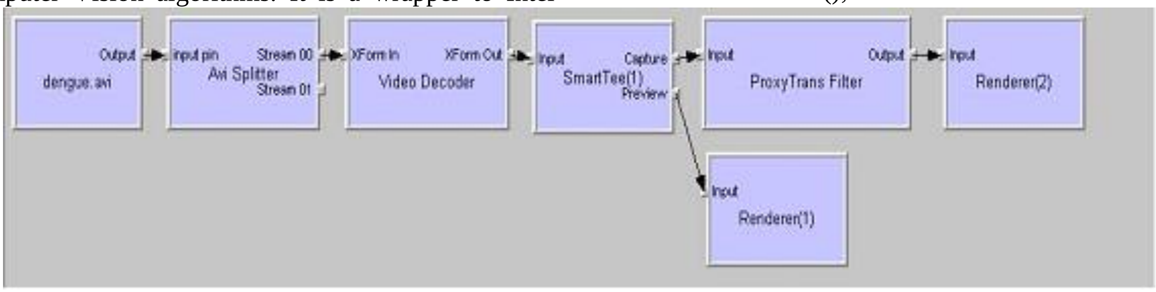

Fig. 16: Direct Show filter graph

A COM interface is an abstract class containing pure virtual functions (forming together the interface). Using a COM interface is the only way to communicate with a COM object. They are obtained by calling the appropriate API function. These functions return a value of type HRESULT representing an error code. The simplest way to verify whether a COM call failed or succeeded is to check the return value using the FAILED macro. Finally the resources allocated are released by calling the SAFE_RELEASE macro. This is done to prevent resource leakage.

To use a component of DirectX, we first call its top-level interface. These are identified by a CLSID identifier and each interface is identified by an IID. The input video file is selected by calling the *addSourceFilter* method. It takes the name of the video as an argument and thus the video to be processed is selected. We use GetPin method for obtaining the input and output pins of the filter. This method takes the name of the filter, pin direction and number of pins as arguments and gives us pin(s) in return.
These pins are then connected to each other to complete the filter graph. We use the *addFilter* method for creating the various filters. This method takes the filter CLSID_ID and name as an input. Then the filter created is added to the graph by connecting it to the output pin of some other filter.

Firstly the source filter is added using the *addSourceFilter* method. Then GetPin method is used to obtain its output pin. Its output pin is connected to the input pin of the *AviMux* filter. *AviMux* filter splits the video into audio and video. Thus *AviMux* Filter gives us two output pins. The first pin contains the video output. The video stream is passed as an input to the Video Decoder filter which decodes the video. The decoded video is passed as an input to the Smart Tee Filter. Smart Tee filter has two output pins. The capture pin controls the sequence flow; the preview pin will receive frames only if extra computational resources are available. When processing a sequence, we could also use two Smart Tee filters, one to display the original sequence, the other to display the processed one. The decoded video is previewed through the preview pin of Smart Tee filter for testing purpose. While the capture output pin of Smart Tee filter is connected to the *proxytrans* filter provided by OpenCV.

Fig. 17: MS VS2005 Implementation of CEZW coder with some distortions at low bpp decoding

ProxyTrans filter is designed specially to simplify the work with DirectShow. It enables using DirectShow videos in applications without creating any additional transform filter. In fact, the objective of this filter is to give access to the programmer to each frame of the sequence that can thus be processed. This is realized through a callback function that is automatically called for each frame of the sequence. This callback function passes in argument a pointer to the current image, the user is then free to analyze and modify this image. The frame packaging used here is similar to MPEG standard. For demonstration purpose we are neglecting the B frame and selected every eighth frame as I frame. The output from the proxytrans filter will be saved on the server and will be streamed to the client via the streaming application. Decoding is done on the client side.

Figure 17 illustrates the input and output frame of our CEZW. The GUI interface provides flexibility and ease to the user in knowing the system performance and in adjusting the parameters. A close up of GUI is given in figure 18 giving the details of various parameters.

Fig. 18: E-learning CODEC Graphic User Interface

3) *Network Streaming Scheme*

The proposed network streaming strategy was tested on Network Simulator-2 with 8 wireless nodes. One of the nodes acted as a multimedia server for the remaining seven nodes. We found that the characteristics of packet loss in wireless networks were different from what was assumed for the proposed streaming strategy. As a result, we concluded that the streaming strategy will incur additional overhead with no significant gains.

Most of the bandwidth estimation tools available as well those proposed in research papers are based on the principle of self-induced congestion. Since bandwidth estimation has to be done every 5 seconds, it is very inefficient to use them as they may lead to network congestion, thus deteriorating performance. Therefore, we simply used the feedback based packet transmission technique which has inherent bandwidth estimation capabilities. The data is sent to the receiver at the rate at which feedback is received.

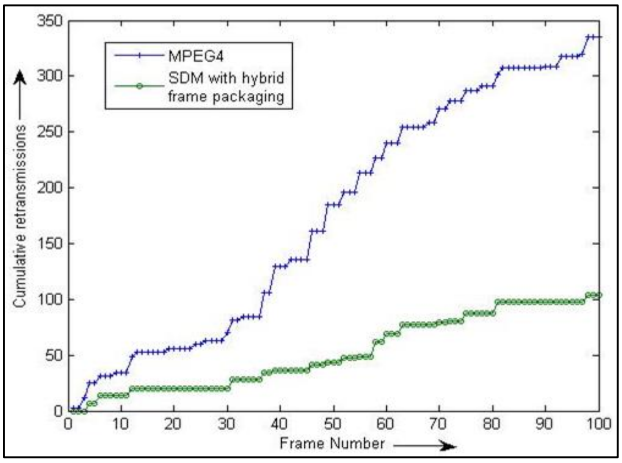

Fig. 19: Comparison of retransmissions in MPEG4 with SDM with hybrid frame packaging

V. RESULTS

Motion estimation and compensation blocks, CEZW coder and arithmetic coding modules were first simulated on Matlab 7. Network Bandwidth was estimated and the response of SDM was tested in Matlab. We then switched over to implement compression and network streaming part over C++.

We use the peak signal-to-noise ratio (PSNR), based on the mean-squared error (MSE), as our "quality" measure. The PSNR of a downsampled YUV image is obtained by using the following equation:

Here, , , and are the mean square errors of the Y, U, and V component of the reconstructed frame with respect to the original frame. The PSNR values can be mapped to ITU-R Quality and Impairment Scale and MOS. MOS is the human impression of the video quality, which is given on a scale from 5 to 1 [50].

Results are presented on educational IITR video sequences. The first video sequence is of relatively poor quality. However, our algorithm works satisfactorily. The classification of video into various VOs is illustrated in fig. 10. Fig. 20 shows the reconstruction of frame no. 24 at different bpp values using SDM, compression and segmentation modules. Some distortions are evident at lower bitrates. The GOP size was taken as 10 and B frames were not included for simulation purpose. Fig.20 shows the evident errors in reconstruction at low bpp without segmentation. This illustrates the need of segmentation module. As illustrated by fig. 9 its use enhances the quality of reconstructed video by optimally allocating the bandwidth. While Fig. 20(c) shows error in reconstruction of frame 24 at 0.8 bpp, using segmentation module, this error was reduced even for a bpp of 0.5.

Fig. 21 shows the change in perceptual quality of the reconstructed frame no. 24 by different commonly used codecs at the same value of PSNR, which is taken to be 35dB for each of these codecs. The frame numbered 24 corresponds to a frame with significant teacher motion. Table. 1 shows the bit rates used by different standard codecs.

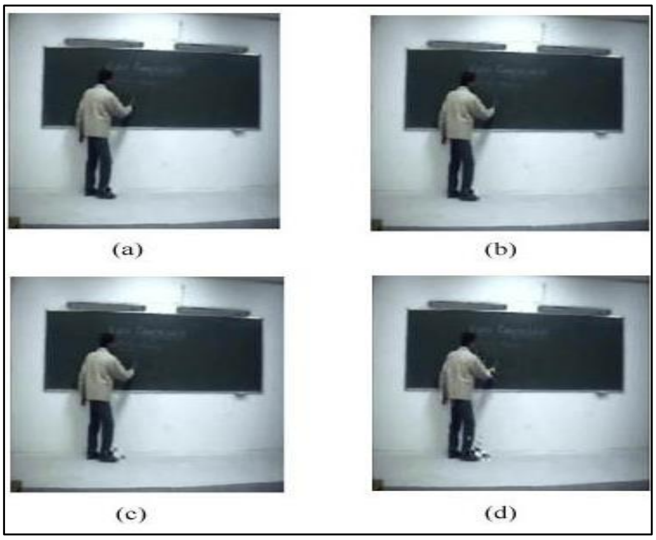

Fig. 20: Reconstructed frames for a GOP (a) frame 21 (I frame), (b) frame 22, (c) frame 24, (d) frame 28 at 0.8 bpp without segmentation

The above figure shows that even at a very low bpp of 0.8, the reconstructed frames can be seen clearly. Thus it shows that quality of picture has been maintained even in low bandwidth constraint. The results obtained are significantly improved by the use of segmentation module.

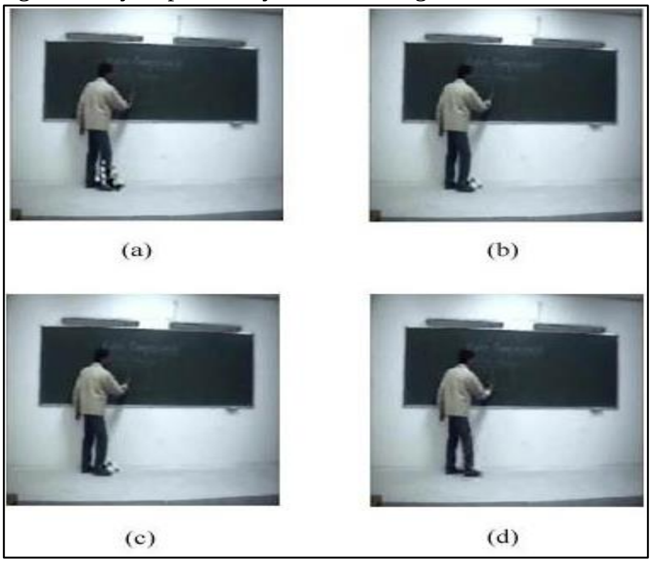

Fig. 21: Reconstruction of frame 24 at (a) 0.1bpp (b) 0.2 bpp (c) 0.3 bpp (d) 0.5 bpp

The above figure shows the reconstruction of a frame at different bits per pixels. There are some distortions in Fig 24 (a) because of a very low bpp of 0.1. As the bpp for the frame is increased, the picture quality is improved. Thus by utilizing the varying bandwidth effectively we are able to provide the best possible picture quality to the client. The changes in network bandwidth are tracked and the best possible reconstruction is provided to the client. We can further see that even in low bpp the reconstruction is satisfactory as compared to other standards like real media[2] or MPEG [28] etc where appearance of blocks hamper the perceptual quality of image.

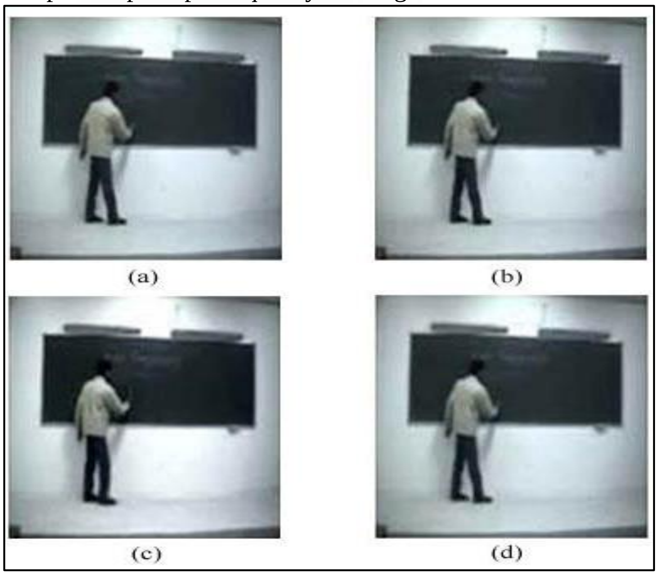

Fig. 22: Reconstruction of frame 24 by (a) Microsoft MPEG4 V2 (b) Windows Media (c) MPEG2 (d) Our algorithm.

Our algorithm shows significant improvement over others. Thus, the bandwidth allocator optimally allocates greater part of the bitstream to the teacher region and obtains good reconstruction of the frame even at low bpp.

Codec	Bits per pixel
Microsoft MPEG4 V2	0.58
Windows Media	0.51
MPEG2	0.88
Our algorithm	0.16

Table 1: Performance comparison of different codecs for IITR video sequence (PSNR=35)

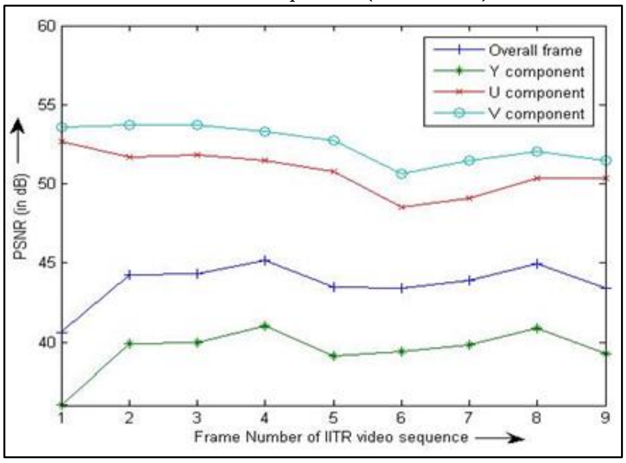

Fig. 23: PSNR measures of different image components for one GOP

Fig. 23 shows the PSNR measures of Y,U, and V components of frames for a sample GOP taken as 10. The results show that PSNR values of U and V components are quite high and justify our use of scaling factor in CEZW+ discussed in section 3.1.2. This helps us to achieve further compression over CEZW. The result is obtained over IITR video sequence.

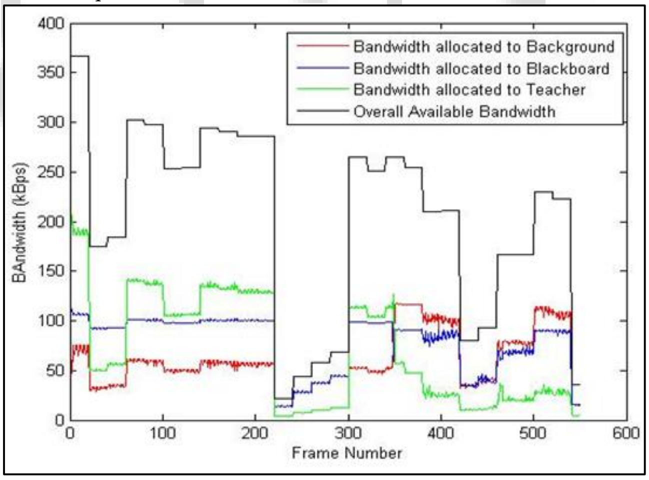

Fig. 24: Bandwidth distribution amongst various VOs

Figure 24 shows the bandwidth allocation amongst various VOs by the bandwidth allocation scheme. It has been obtained for IITR video sequence. It can be seen that the bandwidth allocation adjusts itself to bandwidth variations. We can see that background was allocated less bits in the beginning but later when the motion of teacher decreased and blackboard changes also reduced, its allocated bandwidth increased according to the equations summarized in section 3.1.3. The network bandwidth was simulated and bandwidth distribution was made according to the motion and significance of VO in the current frame.

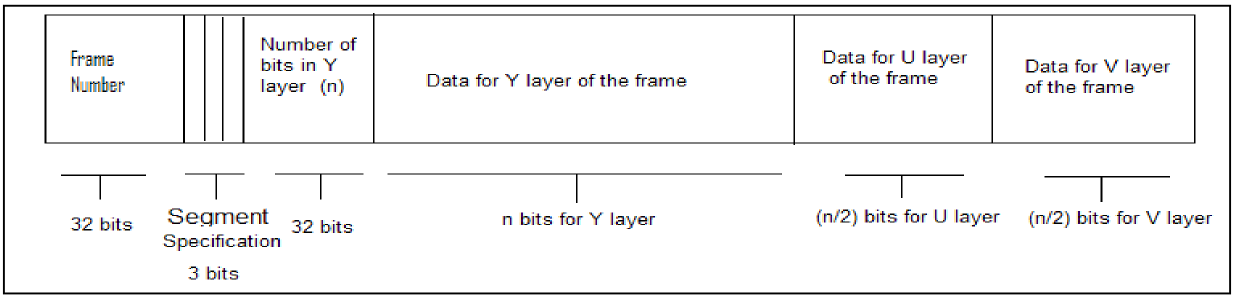

Fig. 25: Packet format

Fig 25 shows the packet format for streaming the compressed frames. The first four bytes have been reserved for specifying the frame number. Next three bits specify the type of segment (teacher, blackboard, background). For example a field value of 100 indicates that the packet contains the teacher region while the other two regions are absent. Only one of the bit values can be 1 at a time. Next, the data of the Y component is present followed by the data of the U and V components. The network bandwidth was estimated at intervals of 1 second and the video has frame rate of 20fps. At high available bandwidths, if the blackboard and teacher motion is small, the scheme allocates higher bit rate to background which is otherwise curbed. Next we present results over second video sequence. This video was recorded from a live Operating Systems lecture (EC324) taught by Dr. Ankush Mittal and thus represents a real classroom. The video was obtained at high quality and resolution of 720*480 pixels at 30fps. This video had proper illumination conditions and excellent results were obtained. Figure 26 explains the working of segmentation module over the video sequence to yield different VOs. We first extract the background Fig 26 (c) by using two spatially distant frames Fig 26 (a) and (b). Then the extracted background and the spatially distant frames are used to extract the teacher and blackboard for these individual frames as specified in the segmentation algorithm.

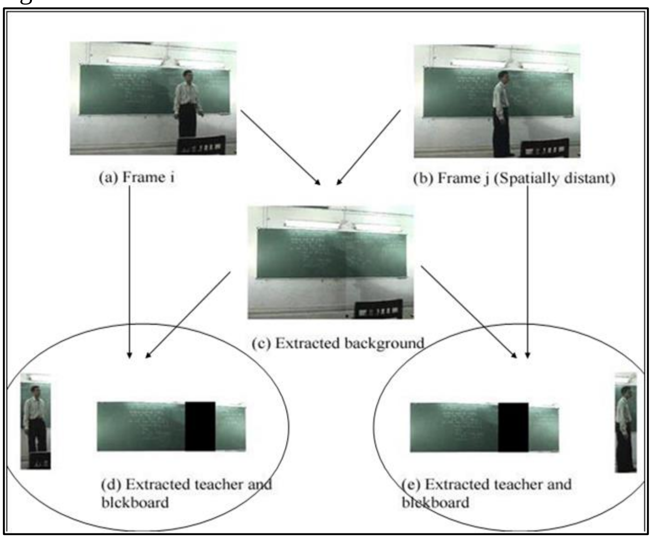

Fig. 26: Segmentation results on OS Lecture Video

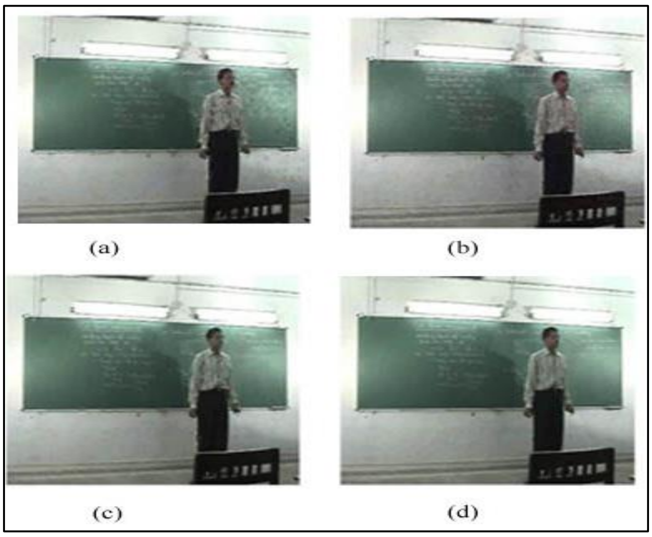

Fig. 27: Reconstruction of frame 34 at 0.05, 0.1, 0.2, 0.5 bpp respectively

The above figure shows the reconstruction of frames at very low bpp varying from 0.05 to 0.5. The picture at 0.5 bpp (Fig 27 d) is better than that at 0.05 bpp (Fig 28 a). The picture quality has improved by increasing the bits per pixel for frames.

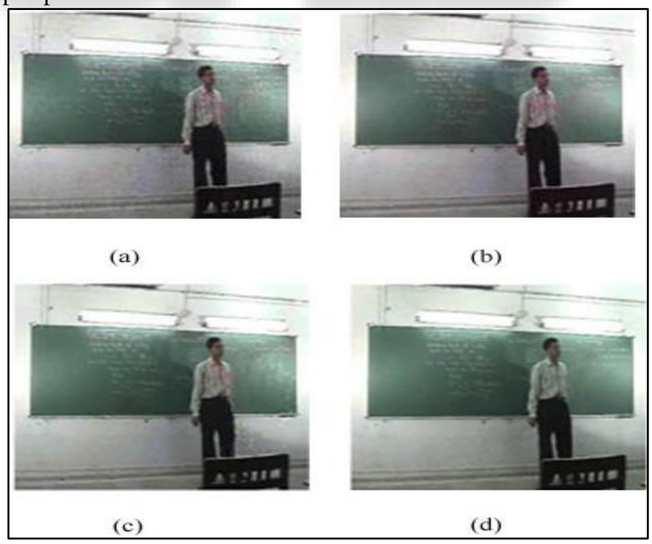

Fig. 28: Reconstruction of frame 106 at 0.05, 0.1, 0.2, 0.5 bpp respectively

The above figure shows that even for a higher frame number 106, the results are same as that for frame 34. Thus it shows that increasing the frame number will not hamper the picture quality.

VI. CONCLUSIONS

The project provides a generic algorithm that significantly reduces the lecture video size, while maintaining the resolution of important video segments. It also provides a scheme for optimal network resource allocation. The CEZW and the SDM modules provide a highly scalable bit stream that is optimally shared between the frame constituents based upon feedback from the network. Further, the coding scheme is highly asymmetric which enables simple reconstruction at user's end. Simple rectangular VOs instead of MPEG-4 Visual objects of arbitrary shapes enable quick decompression. We have also proposed a flexible algorithm for rate-scalable e-learning video compression. For achieving the desired compression ratio we presented the concept of content-based scalable wavelet based video compression scheme. Although the proposed framework has been discussed for e-learning videos, it is equally applicable to other low motion videos with similar visual objects. The work has a remarkable impact on future of e-learning in India and abroad. The proposed framework can be utilized for compression of all slow motion videos and enables video transmission over scarce resource networks like dial up connection etc.

ACKNOWLEDGMENTS

This research was supported/partially supported by Uttarakhand Council for Biotechnology, Haldi and Amrapali Institute, Lamachaur, Haldwani. We thank our guides from G. B. Pant University, Pantnagar who provided insight and expertise that greatly assisted the research.

REFERENCES

[1] J. Gerhard and P. Mayr, "Competing in the e-learning environment-strategies for universities," Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS), January 2002 pp. 3270 – 3279.
[2] http://ocw.mit.edu/index.html, last accessed: 14 March 2007.
[3] http://calpact.berkeley.edu/resources/elearning.html, last accessed: 14 March 2007.
[4] http://nptel.iitm.ac.in, last accessed: 14 March 2007.
[5] http://www.ocwconsortium.org/,last accessed: 14 May 2007.
[6] J. Yong, "Workflow-based e-learning Platform," Proceedings of the Ninth International Conference on Computer Supported Cooperative Work in Design, Volume 2, May 2005, Vol. 2, pp. 1002 – 1007.
[7] D. Strazzo and T. L. Wentling, "A study of e-learning practices in selected fortune 100 companies," Technical Report No. 2, National Center for Supercomputing Applications, UIUC, USA , March, 2001.
[8] T. Liu and J. R. Kender, "Lecture videos for E-learning: current research and challenges," IEEE Multimedia Software Engineering. Proceedings of IEEE Sixth International Symposium on 13-15 Dec. 2004, pp. 574 – 578.
[9] C. S. Lee and T. H. Tan, "Humanizing E-Learning," International Conference on Cyberworlds, 2003, pp. 418 – 422.
[10] Z. Nedic, J. Machotka, A. Nafalski, "Remote laboratories versus virtual and real laboratories", 33rd Annual Frontiers in Education, 5-8 Nov. 2003, Vol.1, Iss. pp T3E-1- T3E-6.
[11] H.H Hann, M.W. Spong, "Remote laboratories for control education", Proceedings of the 39th IEEE Conference on Decision and Control, 12-15 Dec. 2000, Vol.1 pp895 – 900.
[12] I. Gustavsson, " Remote laboratory experiments in electrical engineering education", Proceedings of the Fourth IEEE International Caracas Conference on Devices, Circuits and Systems, 17-19 April 2002, pp:I025-1 - I025-5.
[13] http://research.microsoft.com/~cohen/WangLowFrameRate.pdf last accessed 20 January 2007.
[14] G.Cote, B. Erol, M. Gallant, F. Kossentini. "263+: Video coding at low bit rates", IEEE Trans. Circuit and Systems for Video Technology, vol. 8, no. 7, 1998.
[15] T. Liu, and J. R. Kender, "Lecture videos for e-learning: current research and challenges", IEEE International Workshop on Multimedia Content-Based Analysis and Retrieval, 2004.
[16] B. G. Haskell, A. Dumitras, " A background modeling method by texture replacement and mapping with application to content-based movie coding", ICIP (1) 2002 pp 65-68
[17] A. Dumitras and B. G. Haskell," A Texture Replacement Method at the Encoder for Bit-Rate Reduction of Compressed Video", IEEE Trans. Circuits Syst. Video Techn., 2003.
[18] S. Murphy, S. Goor and L. Murphy, "Performance comparison of multiplexing techniques for MPEG-4 object based content", IPCC 2005. 24thIEEE International 7-9 April 2005 pp 505 – 510.
[19] B. Cavusoglu, D. Schonfeld, R. Ansari, D.K. Bal, "Real-time low-complexity adaptive approach for enhanced QoS and error resilience in MPEG-2 video transport over RTP networks", IEEE Transactions on Circuits and Systems for Video Technology, Volume 15, Issue 12, Dec. 2005, pp. 1604 – 1614.
[20] J.Dong, Y.F.Zheng, "Content-based retransmission for 3-D Wavelet video streaming over lossy networks", IEEE Transactions on Circuits and Systems for Video Technology, Volume 16, Issue 9, September. 2006, pp.1125 – 1133..
[21] N. Feamster and H. Balakrishnan, "Packet Loss Recovery for Streaming Video",12th International Packet Video Workshop, Pittsburgh, PA, April 2002.
[22] J. Lee, and H. Radha, "Interleaved Source Coding (ISC) for Predictive Video Coded Frames over the Internet," IEEE Proc. ICC, May, 2005, pp 1224-1228.
[23] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3445-3462, December 1993.
[24] E. J. Delp, P. Salama, E. Asbun, N. Saenz, and K. Shen, "Rate scalable image and video compression techniques," 42nd Midwest Symposium on Circuits and Systems, Volume 2, 8-11 Aug. 1999, pp. 635 – 638.
[25] Y. Liu, Z. G. Li, Y. C. Soh, "A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard", IEEE Transactions on Circuits

and Systems for Video Technology, Volume 17, Issue 1, Jan. 2007, pp. 68 – 78.
[26] A. K. Jain, "Fundamentals of digital Image Processing" Prentice Hall Inc.
[27] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digital Image processing using MATLAB Pearson edition, Prentice-Hall, 2004, ISBN 0-13-008519-7.
[28] I. Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for next generation multimedia", Wiley Publishing. 2003.
[29] A. N. Netravali and C. B. Rubinstein, "Luminance adaptive coding of chrominance signals", IEEE Transactions on Communications, vol. COM-27, no. 4, pp. 703{710, April 1979.
[30] J. O. Limb and C. B. Rubinstein, "Plateau coding of the chrominance component of color picture signals", IEEE Transactions on Communications, vol. COM-22, no. 3, pp. 12{820, June 1974.
[31] K. Shen and E. J. Delp, "Color image compression using an embedded rate scalable approach", Proceedings of the IEEE Intemational Conference on Image Processing, Santa Barbara, CA, October 26-29, 1997, vol. 3, pp. 34-37.
[32] K. Shen, "A Study of Real-Time and Rate ScalableVideo and Image Compression, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue Univeristy, December 1997.
[33] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," IEEE Transactions on Circuits and Systems for Video Technology, Volume 9, Issue 1, Feb. 1999 Page(s):109 – 122.
[34] T. Liu and C. Choudary, "Content-Adaptive Wireless Streaming of Instructional Videos", Multimedia Tools and Applications, 23(2), pp. 157-171, 2006.
[35] C. Choudary and T. Liu, "Extracting content from instructional videos by statistical modelling and classification", Pattern Analysis & Applications, 2006.
[36] T. Athanasiadism, P. Mylonas, Y. Avrithis, S. Kollias, "Semantic Image Segmentation and Object Labeling", IEEE Transactions on Circuits and Systems for Video Technology, Volume 17, Issue 3, March 2007 pp:298 – 312
[37] J Strauss, D Katabi, F Kaashoek, & B Prabhakar, "Spruce: A light weight End-to-End Tool for Measuring Available Bandwidth", MIT Laboratory for Computer Science, March 2003, 55-56.
[38] V. J. Ribeiro, M. Coates, R. H. Riedi, S. Sarvotham & R. G. Baraniuk, "Multifractal cross traffic estimation", Proceedings of ITC specialist seminar on IP traffic Measurement, September 2000.
[39] M. Jain & C. Dovrolis, "Pathload: A Measurement Tool for End-to-End Available Bandwidth", Proceedings of Passive and Active Measurements, Fort Collins, CO, March 2002.
[40] V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil & L. Cottrell, "PathChirp: Efficient available bandwidth estimation for network paths", Proceedings of Passive and Active Measurement Workshop, 2003.
[41] A. Majumdar, D. Sachs, I. Kozintsev, K. Ramchandran & M. Yeung, "Multicast and Unicast Real-Time Video Streaming", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No.6, June 2002, pp. 524-534.
[42] N. Hu & P. Steenkiste, "Evaluation and Characterization of Available Bandwidth Techniques", IEEE JSAC Special Issue in Internet and WWW Measurement, Mapping, and Modeling, 2003, pp. 879-894.
[43] M. Jain & C. Dovrolis, "Pathload: A Measurement Tool for End-to-End Available Bandwidth", Proceedings of Passive and Active Measurements, Fort Collins, CO, March 2002.
[44] Y. Ming, J.Jiang, and J.Ming, "Background modeling and subtraction using a local linear dependence based Cauchy statistical model," Proceeding of Digital Image Computing: Techniques and Applications, 2003, pp. 469-478.
[45] M. Onishi, M. Izumi and K. Fukunaga, "Blackboard Segmentation Using Video Image of Lecture and Its Applications," Proceedings of 15th International Conference on Volume 4, 3-7 Sept. 2000, pp. 615 - 618.
[46] Brejl M. Sonka M., "Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples." Medical Imaging, IEEE Trans. 19(10), 973–985 (2000).
[47] Carmichael O., Hebert M., "Object recognition by a cascade of edge probes." BMVC, pp. 103–112 (2002).
[48] M. Saenz, P. Salama, K. Shen and E. J. Delp, "An Evaluation of Color Embedded Wavelet Image Compression Techniques," Proceedings of the SPIE/IS&T Conference on Visual Communications and Image Processing (VCIP), January 23-29, 1999, San Jose, California, pp. 282-293.
[49] E. Asbun, P. Salama, K. Shen, and E. J. Delp, "Very Low Bit Rate Wavelet-Based Scalable Video Compression", ICIP (3) 1998 pp 948-952.
[50] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Transactions on Pattern Analysis and Machine intelligence, vol. 11, no. 7, pp. 674-693, July 1989.
[51] ITU, "Methodology for the Subjective Assessment of the Quality of Television Pictures", ITU-R Recommendation BT.500-10, 2000.