Volume 3, Issue 1, January 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Scalability Techniques of MPEG-2 Standard for Video Compression

T.Vishnu Priya[1]
Assistant Professor
Department of ECE
SRK Institute of technology
JNTU-Kakinada
Vijayawada, India

M.Sravya[2]
IV/IV ECE
SRK Institute of technology
JNTU-Kakinada
Vijayawada, India

B.Bhavana[3]
IV/IV ECE
SRK Institute of technology
JNTU-Kakinada
Vijayawada, India

K.Pravallika[4]
IV/IV ECE
SRK Institute of technology
JNTU-Kakinada
Vijayawada, India

K.Gayathri[5]
IV/IV ECE
SRK Institute of technology
JNTU-Kakinada
Vijayawada, India

Abstract* — *Today, video coding is used in a wide range of applications ranging from multi-media messaging, video telephony, video conferencing over mobile TV, wireless and internet video streaming to Standard Definition TV Broadcasting (SDTV) and High Definition TV Broadcasting (HDTV). In order to provide efficient communication the audio and video signals are compressed before broadcasting. Differences exist among various compression standards and their implementation based on the primary requirements of the target application. Though MPEG-1, MPEG-2, MPEG-4 and MPEG-7 standards exists, only MPEG-2 provides efficient transmission of the compressed digital content across a network. Both SDTV and HDTV end users have different quality, resolution and must have an encoding-decoding mechanism, such that they can receive same broadcast and deliver the respective quality and resolution videos. These variable transmission conditions can be dealt using the scalability technique. The ability to encode and decode a video at varying quality and resolution is termed as scalability. In this project we mainly concentrate on scalability techniques of MPEG-2 standard, that have an ability to encode and decode a video at varying quality and resolution and we also implement the performance of each scalability technique for different video signals.

Keywords- Compression, Video coding, Motion estimation, MPEG-2, Scalability techniques.

I. INTRODUCTION

A video signal [1] is the term used to describe any sequence of time varying images. Movies (films) and television are both examples of video signals. Digital video has become very important form of information technology and is now used in many different areas, such as board casting, teleconferencing, mobile telephone, surveillance, and entertainment. People now expect to access a video through a wide range of different devices and over various networks. To provide these kinds of services we must know what a video compression is, for storage and transmission. Compression of video imagery has become a necessity and very important because of transmission and storage of uncompressed video would be extremely costly and impractical. For instance, a video sequence running for (90) minutes, at (25) frames per second, at standard resolution of (720 *576), and with (24) bit per pixel would require (1343692800000) bit or approximately 156.43 gigabytes". This is not a problem if we only wish to access and deal with video through high - end systems with a lot of storage and network band width. However, since it is desired that video will be accessible from a wide range of devices having different capabilities and connected to different networks, therefore some form compression for digital video is required.

II. VIDEO COMPRESSION

Video compression [2] [4] is the process of compressing (encoding) and decompressing (decoding) video. Digital video takes up a very large amount of storage space or bandwidth in its original, uncompressed form. Video compression makes it possible to send or store digital video in a smaller, compressed form. Source video is compressed or encoded before transmission or storage. Compressed video is decompressed or decoded before displaying it to the end user.

Compression refers to the process of reducing the number of bits required to represent the image and video. It comes in two forms namely lossless and lossy. The lossless compression is a process to reduce image or video data for storage and transmission while retaining the quality of original image (i.e. the decoded image quality is required to be identical to image quality prior encoding). In lossy compression, on the other hand, some information present in the original image or video is discarded so that the original raw representation of image or video can only be approximately reconstructed from the compressed representation with high compression ratio. In order for a compressed video bit stream to be decodable uniformly by various platforms and devices, the bit stream format must be predefined. These pre-defined formats are given by MPEG [2]. MPEG is an acronym for Moving Pictures Expert Group which was formed by ISO (International Standards Organization) to set standards for audio and video compression and transmission.

MPEG-1 [2] was finalized in 1991, and was originally optimized to work at video resolutions of 352x240 pixels at 30 frames/sec (NTSC based) or 352x288 pixels at 25 frames/sec (PAL based), commonly referred to as Source Input Format (SIF) video. MPEG-1 is defined for progressive frames only, and has no direct provision for interlaced video applications, such as in broadcast television applications.

MPEG-2 [2][3] is an extension of the MPEG-1 international standard for digital compression of audio and video signals. MPEG-1 was designed to code progressively scanned video at bit rates up to about 40 Mb/sec for applications such as CD-i (compact disc interactive). MPEG-2 is directed at broadcast formats at higher data rates; it provides extra algorithmic 'tools' for efficiently coding interlaced video, supports a wide range of bit rates and provides for multichannel (up to 5 audio channels) surround sound coding.

The MPEG-1 & -2 standards are similar in basic concepts. They both are based on motion compensated block-based transform coding techniques, while MPEG-4 deviates from these more traditional approaches in its usage of software image construct descriptors, for target bit-rates in the very low range, < 64Kb/sec. Because MPEG-1 & -2 are finalized standards and are both presently being utilized in a large number of applications, this paper concentrates on compression techniques relating only to MPEG-2 standard. Note that there is no reference to MPEG-3. This is because it was originally anticipated that this standard would refer to HDTV applications, but it was found that minor extensions to the MPEG-2 standard would suffice for this higher bit-rate, higher resolution application, so work on a separate MPEG-3 standard was abandoned. MPEG-2 was finalized in 1994, and addressed issues directly related to digital television broadcasting, such as the efficient coding of field-interlaced video and scalability. Also, the target bit-rate was raised to between 4 and 9 Mb/sec, resulting in potentially very high quality video. MPEG-2 consists of *profiles* and *levels*. The profile defines the bit stream scalability and the color space resolution, while the level defines the image resolution and the maximum bit-rate per profile.

MPEG-4 [2] was standardized in 1998 and is aimed at very low data rates as well as content-based interactivity on CD-ROM, DVD, and digital TV and universal access, which includes error-prone wireless networks MPEG-4 .MPEG-4 is the multimedia standard for the fixed and wireless web, enabling integration of multiple applications.

MPEG-7 [2] was standardized in 2004. MPEG-7, titled *Multimedia Content Description Interface*, provides a rich set of tools for the description of multimedia content. It provides a comprehensive set of audiovisual *description tools* to create descriptions, which will form the basis for applications that enable the needed effective and efficient access to multimedia content.

III. WHY MPEG-2?

- Allows storage and transmission of movies using currently available storage media and transmission bandwidth
- The MPEG-2 standard is capable of coding standard-definition television at bit rates from about 3-15 Mbit/s and high-definition television at 15-30 Mbit/s.
- Better coding quality at higher data rates in comparison with MPEG 1
- Provides for multichannel surround sound coding
- Often used to create movies to be distributed over the internet
- Offers resolutions of 720x480 and 1280x720 at 60 fps, full CD-quality audio
- Significant requirements for computer encoding
- Supports both interlaced and progressive video
- Supports up to 5 audio channels including surround sound
- May contain GUI, interaction, encryption, data transmission

MPEG 2 [2] [3] describes a combination of lossy video and audio data compression methods. MPEG 2 supposes compression of the raw frames into three kinds of frames: intra-coded frames (I-frame), predictive-coded frames (P-frames), and bidirectional-predictive-coded frames (B-frames). This format defines 4 profiles and 4 levels that correspond to the quality and resolution.

IV. VIDEO CODING

There are different coding techniques like simple differencing, optimal linear prediction and motion compensated prediction, of which, motion compensation prediction is mostly preferred since the object motion is estimated with the help of blocks which makes the task simpler.

Motion-Compensated Prediction

A reason for large variance in the differential frame is due to the motion of objects between temporally adjacent frames. If we can determine how much an object has moved from the previous frame to the current frame, then it is possible to align the object in the current frame with that in the previous frame and obtain the difference. This will result in zero error. An object's motion is estimated using variations in pixel values. In order to determine the motion of an object in a frame, one has to first define the object boundaries in the current and reference frames. To make the task simpler, we will only use rectangular blocks of a constant size. The first task is to estimate the translational motion of a rectangular block of pixels between the current and previous or *reference* frames known as the *motion estimation*. This will yield a *motion vector* with components in the horizontal and vertical directions. The magnitude of the components will be in number of pixels. Once the motion vector is determined, the rectangular block in the current frame can be aligned with that in the reference frame and the corresponding differential pixels can be found. The process of aligning and differencing objects between successive frames is called *motion-compensated (MC) prediction*. Since we are dealing

with images defined on rectangular grids, motion can only be estimated to an accuracy of a pixel. However, motion can be estimated to an accuracy of less than a pixel only approximately because it involves interpolating the values between pixels.

A. Motion Estimation

We want to estimate the translational motion of a block of pixels in the current frame with respect to the previous frame. This can be accomplished by *phase-correlation* method or pel-recursive method, by *block-matching* method in the spatial domain, or by the use of *optical flow equation* (OFE).

Block Matching Method

Block matching [5] is a spatial domain process. In block matching technique, a block of pixels in the current frame is matched to a same size block of pixels in the reference frame using a suitable matching criterion. Because the matching process is computationally intensive and because the motion is not expected to be significant between adjacent frames, the matching process is limited to a *search window* of a much smaller size than the image frame. The figure below illustrates the idea of block matching. The block size is assumed to be $M1 \times N1$ and the search window of size $(M1+2\,M2) \times (N1 + 2N2)$. In the MPEG-2 standard, the macro block size is 16 × 16 pixels and the search window is of size 32 × 32 pixels. There are several matching criteria to choose from for motion estimation.

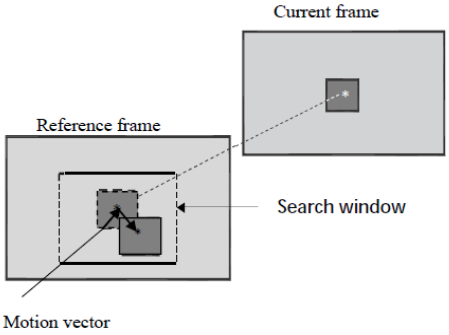

Fig 1: A diagram illustrating the process of matching a rectangular blockof pixels in the current frame to that in the reference frame within a search window that is larger than the rectangular block but smaller than the image.

1). Minimum MSE Matching Criterion

In minimum MSE case, the best matching block is the one for which the MSE between the block in the current frame and the block within the search window W in the reference frame is a minimum. The MSE between the blocks with candidate displacements d_x and d_y is defined as

$$MSE(d_{x,}d_y) = \frac{1}{M_1 N_1} \sum_{(m,n)\in W} (b[m,n,k] - b[m-d_y, n-d_x, k-1])^2 \qquad \text{-----1}$$

The estimate of the motion vector then corresponds to that vector for which the MSE is a minimum, that is,

$$\begin{bmatrix} \hat{d}_x \\ \hat{d}_y \end{bmatrix} = \arg\min MSE(d_x, d_y) \qquad \text{-----2}$$

2).Mean Absolute Difference Matching Criterion

The number of arithmetic operations in equation 1 can be reduced if we use the Mean absolute difference (MAD)[6] as the best block-matching criterion MAD between the current and reference blocks is defined as

$$MAD(d_{x,}d_y) = \frac{1}{M_1 N_1} \sum_{(m,n)\in W} (b[m,n,k] - b[m-d_y, n-d_x, k-1]) \qquad \text{-----3}$$

The estimate of the motion vector is then given by

$$\begin{bmatrix} \hat{d}_x \\ \hat{d}_y \end{bmatrix} = \arg\min MAD(d_x, d_y) \quad \text{-----4}$$

MAD is also known as the *city block distance* and is the most popular matching criterion especially suitable for VLSI implementation and is the recommended criterion for Moving Picture Experts Group (MPEG) standard. In equation 3 if we omit the normalizing constant 1/(*M*1*N*1), the resulting matching criterion is called the *sum of absolute difference* (SAD). Thus, SAD is defined as

$$SAD(d_{X,}d_y) = \sum_{(m,n)\in W} \left| (b[m,n,k] - b[m-d_y, n-d_x, k-1]) \right| \quad \text{-----5}$$

3). Matching Pixel Count Matching Criterion

In matching pixel count (MPC), each pixel inside the rectangular block within the search window is classified as matching or mismatching pixel according to

$$C(m,n;d_x,d_y) = \begin{cases} 1 & \text{if } \left| b[m,n,k] - b[m-d_x, n-d_y, k-1] \right| \le T \\ 0 & \text{otherwise} \end{cases}$$

where *T* is a predetermined threshold. Then the MPC is the count of matching pixels and is given by

$$MPC(d_x,d_y) = \sum_{(m,n)\in W} C(m,n;d_x,d_y) \quad \text{-----6}$$

From equation **6** the estimate of the motion vector is obtained as d_x , d_y for which the MPC is a maximum, that is,

$$\begin{bmatrix} \hat{d}_x \\ \hat{d}_y \end{bmatrix} = \arg\max MPC(d_x, d_y) \quad \text{-----7}$$

V. SEARCH TECHNIQUES

Once a matching criterion is chosen, the task of estimating the motion vector involves searching for the block within the search window [6] that satisfies the chosen criterion. There are various search techniques like exhaustive search or full search technique, three-step search, pyramidal or hierarchical motion estimation, logarithmic search technique, orthogonal search algorithm, cross-search … etc out of which full search technique is best preferred since it predicts the best matching block for all possible displacements.

A. Exhaustive or Full Search:

To find the best matching moving block using any of the criteria described above, one has to compute the appropriate matching metric within the search window for all possible integer-pel displacements. This results in the exhaustive search, also known as the *full search*. Full search is the optimal search method but is computationally intensive. For instance, if *M*1 = *N*1 = *M*2 = *N*2 = 8, then the number of arithmetic operations required per block can be obtained as follows:

Number of searches/block = 2*M*2 × 2*N*2 = 256
Number of OPS/search = *M*1 × *N*1 = 64

VI. MC PREDICTIVE CODING

After having learnt the method of pixel prediction with motion compensation, we now describe a technique to compress video images using MC prediction.

Fig 2(a) depicts video encoding using MC prediction. The current input image is divided into rectangular blocks in a raster scanning order and input to the encoder. The motion, if any, between the current and the reference blocks is estimated and the resulting motion vector is used in the predictor to compensate for the motion and find the predicted block. This predicted pixel block is subtracted from the current input block, the differential block is forward discrete cosine transform (DCT) transformed, and the DCT coefficients are quantized. The quantized DCT coefficients of the MC prediction error are de-quantized, inverse DCT transformed, and added to the predicted pixels to recreate the input pixels. Thus, the video coder operates in a closed loop with the decoder in the encoder loop. At the decoder, the quantized pixel blocks are de-quantized, inverse DCT transformed, and added to the MC pixel block to reconstruct the image.

At the decoder (Fig 2(b)), the received quantized error block is de-quantized, inverse DCT transformed, and added to the previously reconstructed block. Thus, the video sequence is compressed and decompressed.

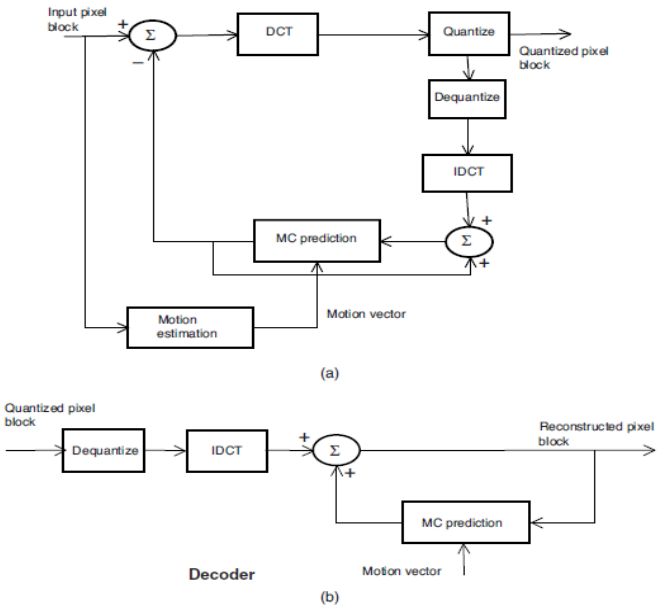

Fig 2: Block diagram depicting video coding using MC predictive coding: (a) encoder and (b) decoder.

VII. SCALABILITY

A single program may be broadcast in different formats, such as SDTV and HDTV. These two end users have different quality and resolution. There must be an encoding/decoding mechanism, whereby both end users can receive the same broadcast and deliver the respective quality/resolution video. The idea is to transport layers of compressed data consisting of a *base* layer followed by one or more *enhancement* layers. Decoding the base layer will deliver lower quality/resolution video, and decoding additional layers will produce the higher quality/resolution video. Because the base layer carries a larger part of the compressed data and the enhancement layers carry only incremental data, the encoding becomes much more efficient. Otherwise we must employ separate coder for each of the intended application. The ability to encode/decode a video at varying quality/resolution is termed *scalability* [7]. MPEG-2 allows three types of scalability [8], namely, SNR, spatial, and temporal. We will briefly describe each of the scalable schemes as follows.

A. SNR SCALABILITY

Fig 3 shows qualitatively how the compressed bit stream is scalable to achieve pictures of different quality. Here, quality is associated with the signal-to-noise ratio (SNR). By decoding only the base layer, a lower quality image is obtained. By decoding the base layer and an additional enhancement layer, we get the same spatial resolution picture but with a better quality. By decoding all the data, we obtain the highest quality picture. In the context of MPEG-2, a typical SNR scalable encoder using DCT is shown in Figure 4(a). The base layer consists of the bit stream generated by quantizing the DCT coefficients of the prediction errors with quantization $Q1$. The difference between the actual and quantized/dequantized DCT coefficients at the encoder is quantized by quantization $Q2$ and transmitted as the enhancement layer. The SNR scalable decoder shown in Figure 4(b) decodes the base layer with inverse quantization $Q1^{-1}$ and delivers a base quality video. If both layers are decoded in the manner shown in Figure 4(b), one obtains the higher quality video.

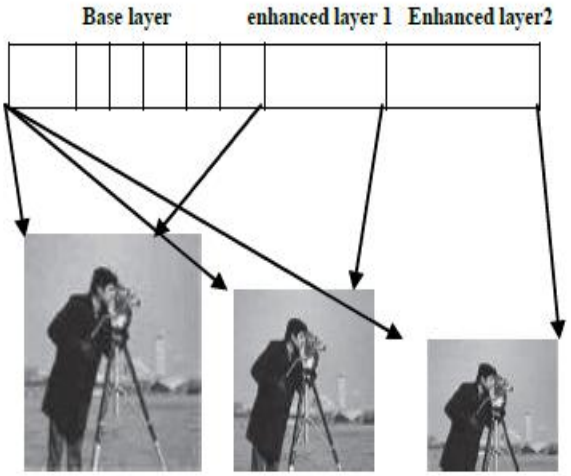

Fig 3: A qualitative way of showing SNR scalability from an MPEG-2 bit stream.

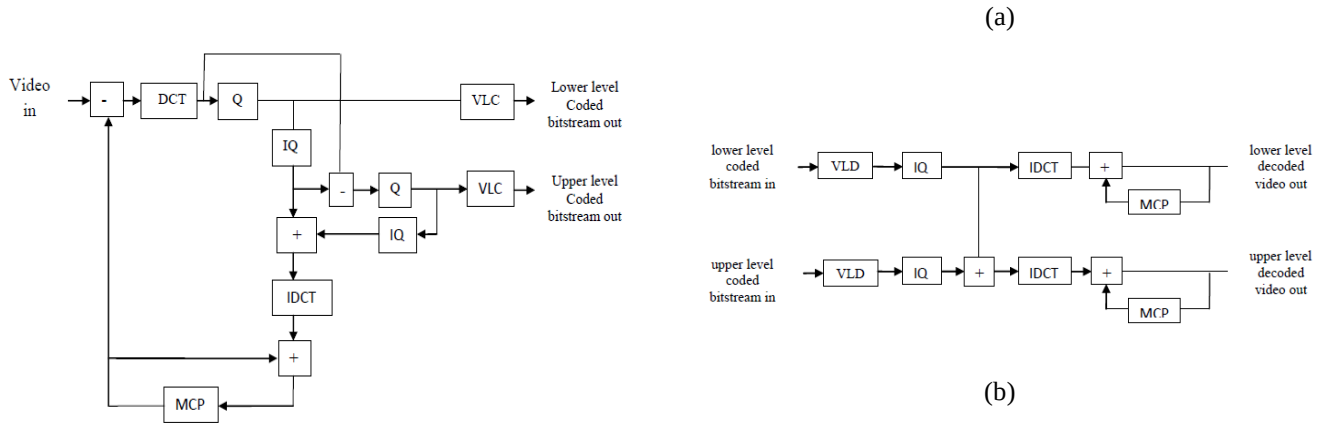

Fig 4: A block diagram showing SNR scalable DCT-based MPEG video coding: (a) encoder and (b) decoder

B. SPATIAL SCALABILITY

The idea behind spatial scalability is similar to that of SNR counterpart, namely, that the base layer carries a compressed lower spatial resolution video and the enhancement layers carry incremental data .When only the base layer bit stream is decoded, one obtains a base resolution video. When the base layer and enhancement layers are together decoded, the highest resolution video is obtained. However, the quality of both the base and enhanced resolution videos may be about the same because the intent is to distribute the same video material in a single bit stream which carries information about different resolutions of the videos. The spatial scalability is illustrated in block diagram form in Figure 5. Figure 5(a) is the spatially scalable encoder. The input video is in full resolution. It is low pass filtered and down sampled to the base resolution and then encoded using MPEG2 scheme. The difference between the full resolution input video and the base layer decoded (locally) video is coded using the DCT transform. The two bit streams are transmitted or stored. On the decoder side (Figure 5(b)), only the base layer is decoded to obtain the base resolution video. To obtain the full resolution video, both base and enhanced layers are decoded, the decoded base resolution pictures are up sampled and filtered, and the two are added. The quantizer scale may be the same in both layers and as a result the decompressed videos may have essentially the same quality but at different resolutions. It must be pointed out that spatial scalability can also be accomplished in the wavelet domain.

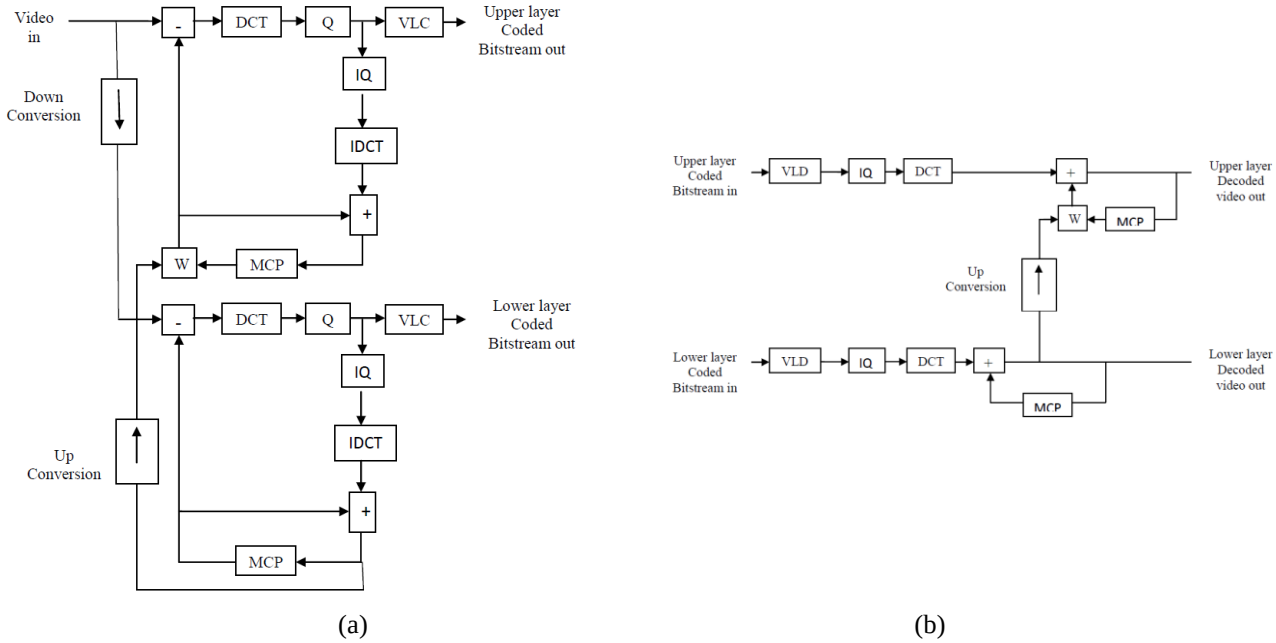

Fig 5: A block diagram illustrating a spatially scalable MPEG video coding system: (a) encoder and (b) decoder.

C. TEMPORAL SCALABILITY

As the name implies, temporal scalability provides different frame rate videos. The base layer yields the basic lower frame rate video. Higher rate video is obtained by decoding additional pictures in the data stream using temporal prediction with reference to the base layer pictures.

D. HYBRID SCALABILITY

MPEG-2 allows combination of individual scalabilities such as spatial, SNR or temporal scalability to form hybrid scalability for certain applications. If two scalabilities are combined, then three layers are generated and they are called the base layer, enhancement layer 1 and enhancement layer 2. Here enhancement layer 1 is a lower layer relative to enhancement layer 2, and hence decoding of enhancement layer 2 requires the availability of enhancement layer 1.

VIII. ALGORITHMS FOR SCALABILITY TECHNIQUES

A. *SNR SCALABILITY*

The following is the algorithm for SNR scalability technique

- SNR scalable video coding using motion compensated prediction
- Block motion is estimated to an integer pel accuracy using full search and SAD matching metric.
- Block size is 8 x 8 pixels.
- Search window size is 16 x 16 pixels.
- The differential block is 2D DCT transformed, quantized using uniform quantizer with constant quantization step of 16.
- The reconstructed block becomes the reference block for the next input frame.
- Base layer carries the above quantized data.
- The enhanced layer is created by coding the difference between the unquantized DCT coefficients and the base layer quantized DCT coefficients and then quantizing this differential DCT with a different quantization step.
- Decoding the base layer yields a lower quality video and adding the enhanced layer to the base layer results in a higher quality video.

The following are the SNR and PSNR characteristics for SNR scalability of different .avi files in MATLAB.

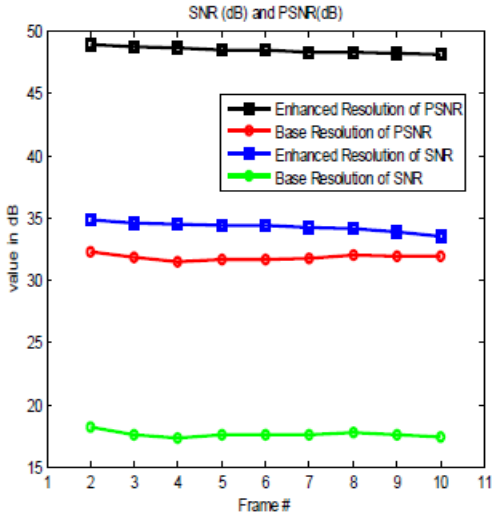

Fig 6: SNR and PSNR values for the base and enhanced quality pictures for SNR scalability of rhinos.avi file in MATLAB.

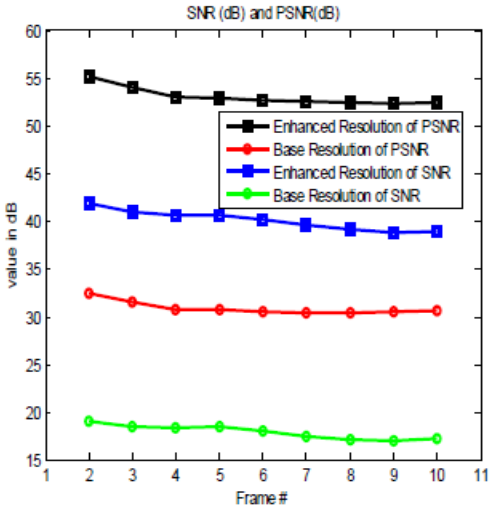

Fig 7: SNR and PSNR values for the base and enhanced quality pictures for SNR scalability of earth.avi file in MATLAB.

B. *SPATIAL SCALABILITY*

The following is the algorithm for spatial scalability technique

- Spatially scalable video coding using motion compensated prediction
- Block motion is estimated to an integer pel accuracy using full search and SAD matching metric.
- Block size is 8 x 8 pixels.
- Search window size is 16 x 16 pixels.
- The differential block is 2D DCT transformed, quantized using uniform quantizer with constant quantization step of 16.
- The reconstructed block becomes the reference block for the next input frame.
- Base layer carries the above quantized data.
- The enhanced layer is created by up sampling the base layer reference image and subtracting it from the current full resolution picture, taking the block DCT of the difference image, quantizing and VLC coding.
- Decoding the base layer yields a lower spatial resolution video and adding the enhanced layer to the up sampled base layer results in a higher resolution video.

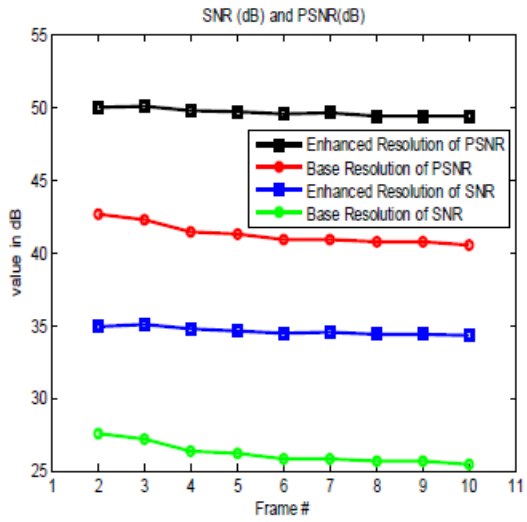

Fig 8: SNR and PSNR values for the base and enhanced quality pictures for SNR scalability of vipmen.avi file in MATLAB

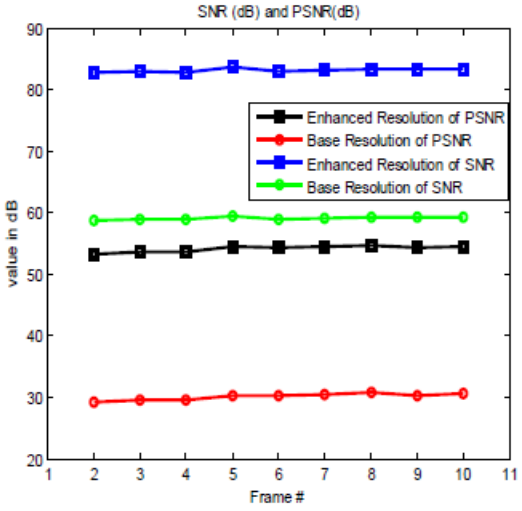

Fig 9: SNR and PSNR values for the base and enhanced quality pictures for SNR scalability of dfs.avi file in MATLAB

The following are the SNR and PSNR characteristics for spatial scalability of different .avi files in MATLAB.

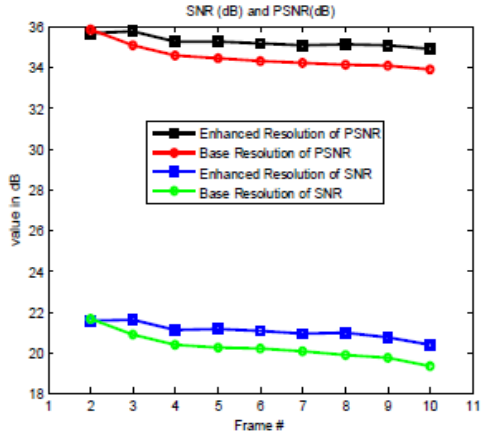

Fig 10: SNR and PSNR values for the base and enhanced quality pictures for spatial scalability of rhinos.avi file in MATLAB.

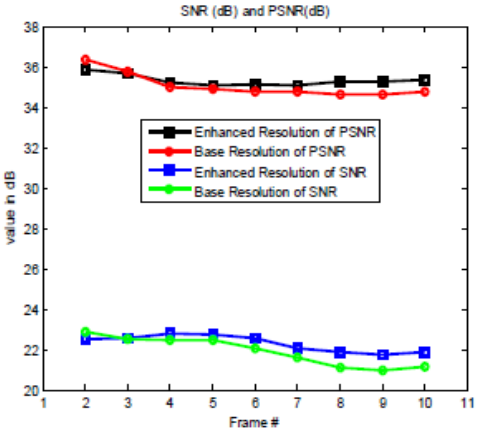

Fig 11: SNR and PSNR values for the base and enhanced quality pictures for spatial scalability of earth.avi file in MATLAB.

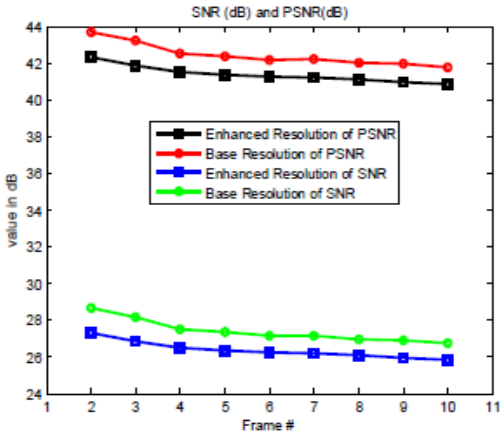

Fig 12: SNR and PSNR values for the base and enhanced quality pictures for spatial scalability of vipmen.avi file in MATLAB.

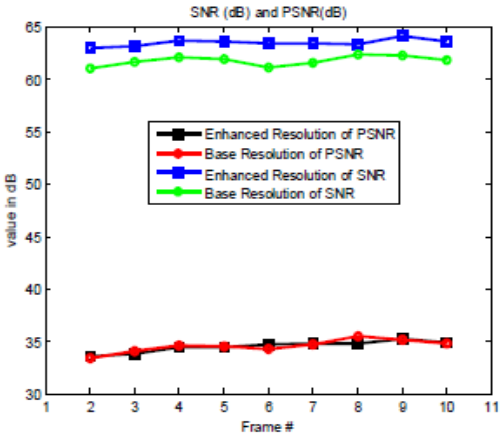

Fig 13: SNR and PSNR values for the base and enhanced quality pictures for spatial scalability of dfs.avi file in MATLAB.

IX. CONCLUSIONS

This paper depicts the extent to which a video can be encoded and decoded at varying quality and resolution by introducing an innovative concept of scalability in order to provide efficient communication. The performance of each scalability technique for different video signals is traced, which compares the characteristics of SNR and PSNR.

X. FUTURE WORK

MPEG-2 allows combination of individual scalabilities such as spatial, SNR or temporal scalability to form hybrid scalability for certain applications. There is a chance of efficient communication of video signal through this hybrid scalability technique.

REFERENCES

[1]. *Introduction to video compression*, Berkeley design technology, Inc.;
[2]. *Video Compression Djordje Mitrovic* – University of Edinburgh
[3].*MPEG-2 VIDEO COMPRESSION* by P.N. Tudor
[4].*ARCHITECTURE OF VIDEO PROCESSING*, Integrated System Laboratory C3I, Swiss Federal Institute of Technology, EPFL, can be referred at http://lsmwww.epfl.ch/Education/former/2002-2003/VLSIDesign/ch13/ArchiMultimed.htm
[5]. *Block-Matching Motion Compensation*; http://www.dcs.warwick.ac.uk/research/mcg/bmmc/index.html
[6]. H. Gharavi and M. Mills, "*Block-matching motion estimation algorithms*: new results," *IEEE Trans. Circ. Syst.*, 37, 649–651, 1990.
[7]. *Scalability Standard Codecs* (c) Image Compression to Advanced Video Coding by Ghandari.M.
[8]. *Overview of Scalable Video Coding extension of H.264/AVC* standard *IEEE Trans. Circ. Syst., Vol 17 Sept 2007.*

.

