An Approach towards Video Compression using Transform Domain Technique

S. Anand
PG Scholar
S.Veerasamy Chettiar College
of Engineering and Technology
Puliangudi, Tirunelveli, Tamilnadu

M. Anjugam M.Tech.
AP/ECE
S.Veerasamy Chettiar College
of Engineering and Technology
Puliangudi, Tirunelveli, Tamilnadu

K. Muthulakshmi M.E.
HOD/ECE
S.Veerasamy Chettiar College
of Engineering and Technology
Puliangudi, Tirunelveli, Tamilnadu

Abstract— Recently, there has been considerable interest in video streaming over the internet's. Multiple Descriptions Coding (MDC) is the excellent solution for real time applications. Video compression enables a number of applications by reducing the required bit rate needed to represent a video sequence, however the compressed video is much more susceptible to errors such as bit errors or packet loss. Resilience to packet loss is a critical requirement in predictive video coding for transmission over packet-switched networks, since the prediction loop propagates error and causes substantial degradation in video quality. Multiple description video coding (MDC) is one of the approaches for reducing the detrimental effects caused by transmission over error-prone networks. . In this paper, a MDC model based on hierarchical B pictures is proposed to optimize the trade-off between coding efficiency and error resilience. The model produces two descriptors by applying transform, quantization followed by lossless compression technique. Our approach employed duplication, spatial splitting, and temporal splitting for the frames at different hierarchical levels. Duplication (high redundancy) is for key frames: spatial splitting (medium redundancy) for reference B frames, and temporal splitting (low redundancy) for non reference B frames. The analysis has been carried out in terms of PSNR, MSE.

Index Terms— Duplication, hierarchical B pictures, multiple description coding (MDC), spatial splitting, temporal splitting.

I. INTRODUCTION

Video as a sequence of pictures (or frames). It is the relationship of high correlation between successive frames. Only small portion of each frame is involved with any motion that is taking place. Analog video is a video signal transferred by an analog signal. An analog color video signal contains luminance, brightness (Y) and chrominance (C) of an television image. When combined into one channel, it is called composite video as is the case, among others with NTSC, PAL and SECAM. Analog video may be carried in separate channels, as in two channel S-Video (YC) and multi-channel component video formats. Analog video is used in both consumer and professional television production applications. However, digital video signal formats with higher quality have been adopted, including serial (SDI), Digital Visual Interface (DVI), High-Definition Multimedia Interface (HDMI) and Display Port Interface. Digital video is a type of digital recording system that works by using a digital rather than an analog video signal. Digital video is audio/visual data in a binary format. Information is represented as a sequence of zeroes and ones, rather than as a continuous signal as analog video. Digital video are becoming more popular and accessible through the various media technology advances which enable users to capture, manipulate and store video data in efficient and inexpensive ways. Reason for taking digital video is given as: Direct random access which is good for nonlinear video editing, No problem for repeated recording, No need for blanking and synchronize pulse. Video compression plays an important role in modern multimedia applications Video compression uses modern coding techniques to reduce redundancy in video data. Most video compression algorithms and codec's combine spatial image compression and temporal motion compensation. Video compression is a practical implementation of source coding in information theory. In practice, most video codec's also use audio compression techniques in parallel to compress the separate, but combined data streams as one package. The idea behind compression is to save time and the number of bits sent between images by taking the difference between them instead of sending each frame again. With video streaming and storage becoming so popular this is a very useful tool to have. Some video compression schemes typically operate on square-shaped groups of neighboring pixels, often called macro blocks. These pixel groups or blocks of pixels are compared from one frame to the next, and the video compression codec sends only the differences within those blocks. In areas of video with more motion, the compression must encode more data to keep up with the larger number of pixels that are changing. Commonly during explosions, flames, flocks of animals, and in some panning shots, the high-frequency detail leads to quality decreases or to increases in the variable bit rate. Digital video compression is thus necessary even with exponentially increasing bandwidth and storage capacities. Fortunately, digital video has significant redundancies and eliminating or reducing those redundancies results in compression. Video

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
TITCON-2015 Conference Proceedings

compression can be lossy or loss less. Loss less video compression reproduces identical video after decompression. video contains much spatial and temporal redundancy. In a single frame, nearby pixels are often correlated with each other. This is called spatial redundancy, or the intraframe correlation. Another one is temporal redundancy, which means adjacent frames are highly correlated, or called the interframe correlation. Therefore, our goal is to efficiently reduce spatial and temporal redundancy to achieve video compression.

II. RECENT WORKS

For real-time applications, since retransmission is often not acceptable, error resilience (ER) and error concealment (EC) techniques are required for displaying a pleasant video signal despite the errors and for reducing distortion introduced by error propagation. Several ER methods have been developed, such as forward error correction [1], intra/intercoding mode selection [2], layered coding [3], and multiple description coding (MDC) [4]. This paper is concerned with MDC. MDC is a technique that encodes a single video stream into two or more equally important substreams, called *descriptions*, each of which can be decoded independently. Unlike the traditional single description coding (SDC), where the entire video stream (single description) is sent in one channel, in MDC, these multiple descriptions are sent to the destination through different channels, resulting in much less probability of losing the entire video stream (all the descriptions), where the packet

losses of all the channels are assumed to be independently and identically distributed. The first MD video coder, called multiple description scalar quantizer [5], was realized in 1993 by Vaishampayan who proposed an index assignment table that maps a quantized coefficient into two indices each could be coded with fewer bits. Due to effectiveness in providing error resilience, a variety of research on different MDC approaches had been proposed afterward. These approaches can be intuitively classified through the stage where it split the signal, such as frequency domain [5], [6], spatial domain [7], [8], and temporal domain [9], [10]. In our previous work [11], a hybrid MDC method has been proposed, which applies MDC first in spatial domain to split motion compensated residual data, and then in frequency domain to split quantized coefficients. The results in [11] show that, by properly utilizing more than one splitting technique, the hybrid MDC method can improve error-resilient performance. Although a variety of MDC approaches have been proposed, most of them were built upon conventional H.264/AVC coding structure and did not utilize hierarchical B-picture prediction. In a hierarchical B-picture prediction framework, the B frames at the coarser temporal levels can be used as a reference for the B frames at the finer temporal levels, and therefore the coding efficiency can be further improved. Compared with classical H.264/AVC prediction structure IBBP, the improvement can be more than 1 dB as described in [12]. Even though hierarchical- B picture coding has been

widely used in scalable extension of H.264/AVC [13] to provide temporal scalability, it is rarely adopted in multiple description coding. In [14], an MDC based on hierarchical B pictures was proposed, where two descriptions are generated by duplicating the original sequence and then coded by hierarchical B structures with staggered key frames in the two descriptions. By using different QPs at different levels, their approach enables each frame to have two different quality fidelities in different descriptions. When two descriptors are received, their approach simply selects the frame with high-fidelity, or uses a linear combination of the high-fidelity and low-fidelity frames to generate a better reconstruction. When only one descriptor is received, the lost frame is recovered by copying from the corresponding frame in the other descriptor. It can be seen that although their MDC approach employs hierarchical B-pictures to improve coding efficiency, it still suffers from high bit-rate redundancy by duplicating the original sequence to two descriptions. This paper presents a MDC based on hierarchical B pictures. Our approach employed duplication, spatial splitting, and temporal splitting for the frames at different hierarchical levels to provide unequal redundancy to frames with different fidelity requirements.

III. MULTIPLE DESCRIPTION CODING BASED HIERARCHICAL B PICTURES

Recently there has been considerable interest in video streaming over the internet's. Multiple Descriptions Coding (MDC) is the excellent solution for real time applications. During data transmission, packets may be dropped or damaged, due to channel errors, congestion, and buffer limitation. Moreover, the data may arrive too late to be used in real-time applications. In the case of transmission of compressed video sequences, this loss may be devastating and result in a completely damaged stream at the decoder side. For real-time applications, since retransmission is often not acceptable, error resilience (ER) and error concealment (EC) techniques are required for displaying a pleasant video signal despite the errors and for reducing distortion introduced by error propagation. MDC is a technique that encodes a single video stream into two or more equally important sub streams, called descriptions, each of which can be decoded independently. Unlike the traditional single description coding (SDC), where the entire video stream (single description) is sent in one channel, in MDC, these multiple descriptions are sent to the destination through different channels, resulting in much less probability of losing the entire video stream (all the descriptions), where the packet losses of all the channels are assumed to be independently and identically distributed. The first MD video coder, called multiple description scalar quantizer proposed an index assignment table that maps a quantized coefficient into two indices each could be coded with fewer bits. Due to effectiveness in providing error resilience, a variety of research on different MDC approaches had been proposed afterward. These approaches can be intuitively classified through the stage where it split the signal, such as frequency domain, spatial domain and temporal domain. Although a variety of MDC approaches have been proposed, most of them were built upon

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
TITCON-2015 Conference Proceedings

conventional H.264/AVC coding structure and did not utilize hierarchical B picture prediction. In a hierarchical B-picture prediction framework, the B frames at the coarser temporal levels can be used as a reference for the B frames at the finer temporal levels, and therefore the coding efficiency can be further improved.

A. HIERARCHICAL B PICTURES

In the hierarchical B pictures-based video coding, some pictures are regularly (or irregularly) selected from the original sequence as key pictures. A typical regular hierarchical prediction structure is depicted in Figure, where the key pictures can be I pictures. A key picture and all pictures that are temporally located between the current key picture and the previous key picture construct a group of pictures (GOP). The remaining B frames are hierarchically predicted using two reference frames from the nearest neighboring frames of the previous temporal level. In a hierarchical B-picture prediction framework, the frames at lower hierarchical levels can be used as a reference for the frames at higher hierarchical levels. Due to this dependence, the decoding quality of a frame strongly depends on the quality. The frame lost at the lower level will result in more corrupted frames. The proposed MDC model is illustrated in Fig. 1, where a nondyadic hierarchical B-picture structure with four levels is used. We refer to the I frames at the lowest hierarchical level as key frames; the B frames at intermediate levels as reference B (RB) frames because they are used as reference; and the B frames at the highest level as non reference B (NRB) frames because they are not used as reference. The I-frame is used to predict the first P-frame and these two frames are also used to predict the first and the second B-frame. The second P-frame is predicted using the first P-frame and they join to predict the third and fourth B-frames. This structure suggests a problem because the fourth frame (a P-frame) is needed in order to predict the second and the third (B-frames). So we need to transmit the P-frame before the B-frames and it will delay the transmission (it will be necessary to keep the P-frame). As Fig. 1 shows, we apply duplication (denoted by D) on key frames for providing the highest error resilience; spatial-splitting (S) on RB frames for modest error resilience; and temporal-splitting (T) on NRB frames for the lowest error resilience. The resulting two descriptions are illustrated in Fig. 3.2, where the rectangles with a missing corner represent incomplete frames (due to spatial splitting). It can be seen that, due to different MDC methods applied, the frames at different hierarchical levels have unequal redundancy to provide robustness again errors. Assuming that description D0 is lost, the lost key frames (1 to 10) can be easily reconstructed at decoder by using the same frames in description D1. The partially lost level-1 and level-2 frames (4 and 7) can be estimated by using the information of their counterparts in description D1, while the lost level-3 frames (2,3,5,6,8 and 9) which are not in D1, can only be estimated by using other frames.
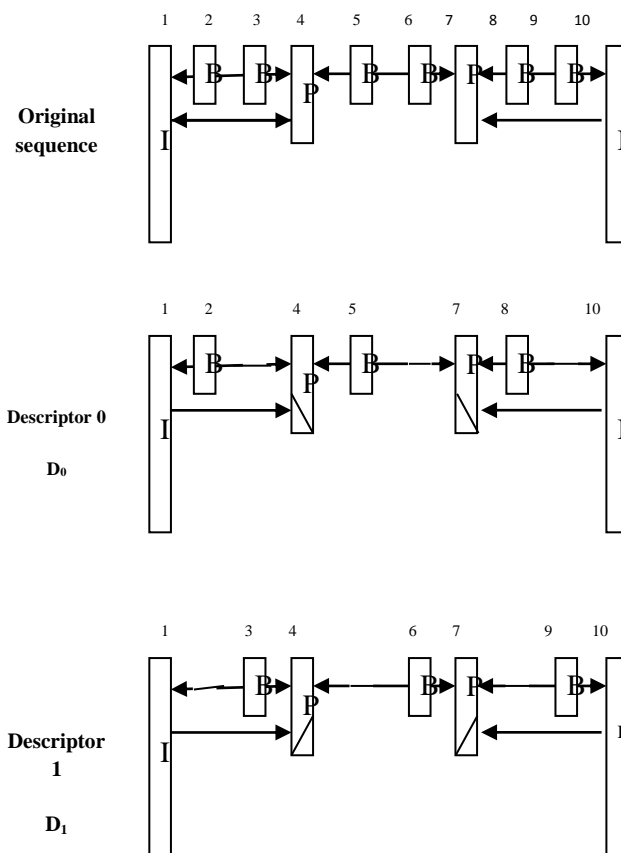

Fig 1 proposed MDC based on Hierarchical B-pictures

IV. PROPOSED METHOD

The overall block diagram consists of two parts such as encoder and decoder. In encoder first the video is separated into three types of frames by means of using intra prediction or motion compensation technique. Then the frames are transformed and quantized followed by Huffman encoding. The encoded output has two descriptors named as D_0 and D_1. In the decoder side the process is inverse of encoding technique. The video be reconstructed when the decoding process is completed. The encoder architecture of the proposed MDC model is depicted in Fig 2.
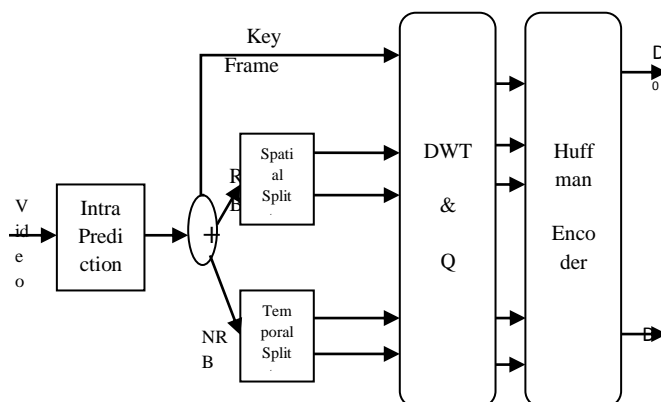

Fig 2 Encoder Architecture

The encoder architecture consist of three paths for three different kinds of frames defined as key frame, RB frame, and NRB frame. Key frames will go through discrete cosine transform, quantization, and Huffman coding stages before it is duplicated to two descriptions. NRB frames will go to a temporal splitter which assigns the input frames, in turn, to the two output paths such that successive NRB frames will go to different descriptors. RB frames will enter a spatial splitter which splits each input frame into two parts which are then separately discrete cosine transformed, quantized, and Huffman encoded before going to their respective descriptors.

A. INTRAPREDICTION

There are three types of frames be considered in this project named as I frames, P frames and B frames. An I frame is an 'Intra-coded picture', in effect a fully specified picture, like a conventional static image file. P frames and B frames hold only part of the image information, so they need less space to store than an I frame and thus improve video compression rates. A P frame ('Predicted picture') holds only the changes in the image from the previous frame. For example, in a scene where a car moves across a stationary background, only the car's movements need to be encoded. The encoder does not need to store the unchanging background pixels in the P frame, thus saving space. P frames are also known Reference B frame. A B frame ('Bi-predictive picture') saves even more space by using differences between the current frame and both the preceding and following frames to specify its content.

B. SPATIAL SPLITTER

The spatial splitter performs splitting on an 8×8 block basis in the residual domain. For each 8×8 residual block, it is first poly phase permuted inside the block and then is split into two. The permuting mechanism is that, for every 2×2 pixels inside the 8×8 residual block, the top-left pixel is re-arranged to the top-left 4×4 block, the top-right pixel to the top-right 4×4 block, the bottom-left pixel to the bottom-left 4×4 block, and the bottom-right pixel to the bottom right 4×4 block. After poly phase permutation, the 8×8 block is split into two 8×8 blocks, each carries two 4×4 blocks chosen in diagonal and the remaining two 4 ×4 blocks are given all-zero residuals. Note that there are four 8 ×8 residual blocks in each macro block, all of them are permuted and split in the same way. In the decoder side the spatial merger is used to merge two complementary RB frames into a full RB frame.

C. TEMPORAL SPLITTER

It can be performed in Non reference B frames such as B-frames. It is a technique which splits the NRB frames into two temporally splitted into two NRB frames by means of the frames splits to two descriptors consiqutely. This means the first NRB frame moves to descriptor 1 and the next NRB frame moves to descriptor 2. The temporal

merger is used to reconstruct the order of NRB frame for output sequence.

D. DWT

Wavelet transform has recently become a very popular when it comes to analysis, de-noising and compression of signals and images. The Discrete Wavelet Transform (DWT) is sufficient for most practical applications and for the reconstruction of the signal. The DWT provides enough information and offers a significant reduction in the computation time. In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time). The main idea is the same as it is in the CWT. A time-scale representation of a digital signal is obtained using digital filtering techniques. The CWT is a correlation between a wavelet at different scales and the signal with the scale (or the frequency) being used as a measure of similarity. The continuous wavelet transform was computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating over all times. In the discrete case, filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies. The resolution of the signal, which is a measure of the amount of detail information in the signal, is changed by the filtering operations, and the scale is changed by up sampling and down sampling (sub sampling) operations. Sub sampling a signal corresponds to reducing the sampling rate, or removing some of the samples of the signal. For example, sub sampling by two refers to dropping every other sample of the signal. Sub sampling by a factor *n* reduces the number of samples in the signal *n* times. Up sampling a signal corresponds to increasing the sampling rate of a signal by adding new samples to the signal. For example, up sampling by two refers to adding a new sample, usually a zero or an interpolated value, between every two samples of the signal. Up sampling a signal by a factor of n increases the number of samples in the signal by a factor of n.

E. HUFFMAN ENCODER

Huffman encoding is a form of entropy encoding and it is based on Shannon's Information theory. The fundamental idea behind Huffman encoding is that symbols, which occur more frequently, should be represented by fewer bits, while those occurring less frequently should be represented by more number of bits. This scheme is similar to the one utilized in Morse code. Shannon has proved that the entropy of the total message gives the most efficient code, with minimum average code length, for sending a message. Given n symbols S_1 to S_{n-1} with probabilities of

occurrence P1 to Pn-1 in a certain message, the entropy of the message will be given by

$$\text{Entropy} = \sum p_1 log_2 \left(\frac{1}{p_1}\right)$$

Huffman encoding attempts to minimize the average number of bits per symbol and try to get a value close to entropy. Huffman encoding isn't implemented in this manner in the JPEG image compression standard. The standard code tables for Huffman coding are defined in the standard for the DC values and the non-DC values as well. These values are looked up both for encoding and decoding. Huffman code is a prefix code and hence it can be uniquely decoded.

The decoder architecture of the proposed MDC model is depicted in Fig. 3, where the two descriptors D_0 and D_1, are first Huffman decoded, dequantized, and inversely discrete cosine transformed separately, then spatial merger and temporal merger are applied to RB and NRB frames respectively. The spatial merger is used to merge two complementary RB frames into a full RB frame. The temporal merger is used to reconstruct the order of NRB frame for output sequence. If the decoder does not receive the two descriptors intact, then estimation process such as spatial or temporal estimation will be adopted to reconstruct the lost data.
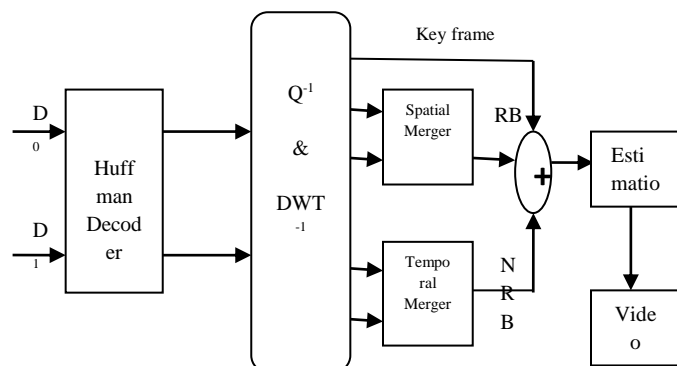

Fig 3 Decoder Architecture

V. EXPERIMENTAL RESULTS

The results for video compression for various videos is done. The video is separated into frames. Frame is separated into I,P,B frames. They are encoded and decoded based on motion vectors. The I frame is the key frame which is first frame and last frame in the GOP. After frame separation the frames are given to the DCT and Quantization. With the help of I1 frame the P4 frame is predicted and the motion vector for the P frame is estimated. In the decode side the motion vector of P4 is used for the P4 frame prediction. The number of motion vectors differ from video to video. Depends upon the block size considered for the frame the number of vectors is changed. With the help of I1 frame and P4 frame the B2 frame is predicted. In the decode side the B2 frame is predicted by using the motion vector of B2 in addition with the I1 and P4 frames. The B2 which is also predicted with

the help of I1 frame and P4 frame. In the decode side the B3 frame is predicted by using the motion vector of B3 in addition with the I1 and P4 frames. In the same way encoding of 10 frames are done. They are called as Group of Pictures. Each set of 10 frames form a GOP. The GOP consists of 2 I frames, 2 P frames and 6 B frames. The comparison of PSNR values using DCT and DWT is given in table 1.

Video	Block	PSNR	
		DCT	DWT
Suzie.avi	8 x8	39.4138	42.2542
	16 x 16	38.8178	40.1363
Vipmen.avi	8 x8	41.6078	43.7854
	16 x 16	41.3621	43.5690
Rhinos.avi	8 x8	29.8566	33.7683
	16 x 16	28.6289	32.8359

Table1 comparison of DCT and DWT
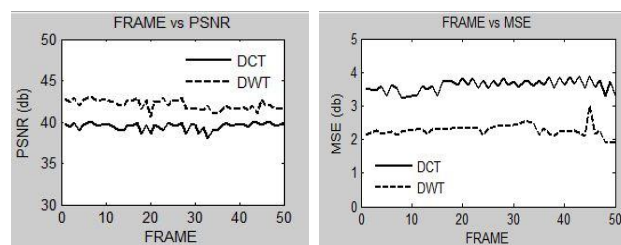
(a)   (b)

Fig 3 (a)Frames Vs PSNR (b)Frames VS MSE Comparison graph for DCT and DWT for a video sequence.

It is observed that PSNR value increases when discrete Wavelet transform is used instead of discrete cosine transform.

VI. CONCLUSION

Video Compression using MDC model based on hierarchical B pictures is proposed. The model produces two descriptors by applying MDC techniques such as duplication, spatial splitting and temporal splitting on frames at different hierarchical levels. The frames are transformed by DWT before applying Huffman coding that generate two descriptors. By taking account for importance of the frames in the hierarchical structure, the model is able to optimize the tradeoff between coding efficiency and error resilience. The results shows that DWT outperforms DCT with efficient compression ratio and increase in PSNR.

REFERENCES

[1] A. Nafaa, T. Taleb, and L. Murphy, "Forward error correction strategies for media streaming over wireless networks," *IEEE Commun. Mag.*, vol. 46, no. 1, pp. 72–79, Jan. 2008.

[2] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[3] C.-M. Fu, W.-L. Hwang, and C.-L. Huang, "Efficient post-compression error-resilient 3D-scalable video transmission for

packet erasure channels," in *Proc. IEEE ICASSP*, Mar. 2005, pp. 305–308.

[4] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proc. IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[5] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[6] O. Campana and R. Contiero, "An H.264/AVC video coder based on multiple description scalar quantizer," in *Proc. IEEE ACSSC*, Oct.–Nov. 2006, pp. 1049–1053.

[7] R. Bemardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, "Polyphase spatial subsampling multiple description coding of video streams with H.264," in *Proc. IEEE ICIP*, Oct. 2004, pp. 3213–3216.

[8] J. Jia and H. K. Kim, "Polyphase downsampling based multiple description coding applied to H.264 video coding," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E89-A, no. 6, pp. 1601–1606, Jun. 2006.

[9] J. G. Apostolopoulos, "Error-resilient video compression through the use of multiple states," in *Proc. IEEE ICIP*, vol. 3. Sep. 2000, pp. 352–355.

[10] S. Gao and H. Gharavi, "Multiple description video coding over multiple path routing networks," in *Proc. ICDT*, Aug. 2006, pp. 42–47.

[11] C. W. Hsiao and W. J. Tsai, "Hybrid multiple description coding based on H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 76–87, Jan. 2010.

[12] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MTCF," in *Proc. IEEE ICME*, Jul. 2006, pp. 1929–1932.

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[14] C. Zhu and M. Liu, "Multiple description video coding based on hierarchical B pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 511–521, Apr. 2009.

S.Anand received the Bachelor of Engineering (B.E) degree in Electronics and Communication Engineering from the Dr.Sivanthi Aditanar College Of Engineering, Tiruchendur, Tamilnadu, India in 2012. He is currently pursuing the Master of Engineering (M.E) degree in the Department of Applied Electronics in S.Veerasamy Chetttiar College of Engineering and Technology, Puliangudi, Tamilnadu, India. His research interests include image processing and digital electronics.

M.Anjugam received the Bachelor of Engineering (B.E) degree in Electronics and Communication Engineering from the PSN College of Engg & tech, tirunelveli, Tamilnadu India, in 2009 and Master of Engineering (M.E) degree in Embedded Systems Technology from the Veltech Dr.Rr & Dr.Sr Technical University, chennai Tamilnadu, India in 2012. Her research interests include embedded systems.

K.Muthulakshmi received the Bachelor of Engineering (B.E) degree in Electronics and Communication Engineering from the R.V.S. College of Engineering, Dindigul, Tamilnadu, India in 1999 and Master of Engineering (M.E) degree in Digital Communication and Network Engineering from the Arulmigu Kalasalingam College of Engineering, Srivilliputtur, Tamilnadu, India in 2007. She is currently pursuing the Ph.D degree in the Department of Information and Communication Engineering. Her research interests include image processing, image compression, video compression and multimedia communication.