International Journal of Computational Intelligence and Informatics, Vol. 3: No. 4, January - March 2014

A Novel Video Coding Technique for Robust Video Transmission

K Muthulakshmi
Department of Electronics and Communication Engineering
S Veerasamy Chettiar College of Engineering and Technology
Puliyangudi, Tamil Nadu, India
muthu_bala12@yahoo.com

V Seenivasagam
Department of Computer Science Engineering
National Engineering College
Kovilpatti, Tamil Nadu, India
yespee1094@yahoo.com

M Ganeswari
S Veerasamy Chettiar College of Engineering and Technology
Puliyangudi, Tamil Nadu, India
ganeswari02@gmail.com

***Abstract*-Now-a-days efficient video compression techniques are essential in order to make digital video applications feasible. A new approach to the video coding techniques is introduced here. For the efficient video transmission the Set Partitioning In Hierarchical Trees (SPIHT) algorithm is used. The three dimensional (3-D) SPIHT coder has proved its efficiency and its real-time capability in compression of video. Since Three Dimensional Spatio-temporal orientation trees coupled with powerful SPIHT sorting and refinement, it provides comparable performance to H.263 standard videos. The 3D wavelet transform (WT) algorithm is based on the "Group Of Frames" (GOF) concept. The group of frames are decomposed both temporally and spatially. The decomposition process utilizes the wavelet filters. The transform coefficients are coded using "Three Dimensional Set Partitioning in Hierarchical Trees" (3-D SPIHT). In the reconstruction phase, the 3-D SPIHT decoding algorithm and the inverse wavelet transform are employed respectively.**

Keywords- Video Compression, Set Partitioning In Hierarchical Trees (SPIHT), 3-D Wavelet Transform, Group Of Frames (GOF)

I. INTRODUCTION

The high bit rates that result from the various types of digital video make the transmission through the intended channels very difficult. In addition, the processing power needed to manage raw volumes of data for storage and transmission makes the receiver hardware very expensive. To overcome the problem, engineers and researchers work not only on developing high speed processors and peripherals that are capable of processing huge volumes of data but also on establishing video compression standards. The latter facilitates manipulation and storage of full-motion video as a form of computer data, and its transmission over existing and future computer networks, or over worldwide broadcast channels. Compression is the conversion of data to a format that requires fewer bits, usually so performed that the data can be stored or transmitted more efficiently. The size of the data in compressed form (C) relative to the original size (O) is known as the compression ratio (CR=C/O). If the inverse of the process, decompression, produces an exact replica of the original data then the compression is lossless. On the other hand, lossy compression has a higher compression ratio but does not allow reproduction of an exact replica of the original image .The success of data compression depends largely on compressible than others the data itself. Some data types are inherently more. Any compression algorithm should satisfy the following requirements, if we have to standardize the compression-un compression procedure.

- It should have a high compression rate at the same time maintain a good enough quality so that uncompressed file is not so much different from the original file and contains almost all of the important information.
- The technique of compression should be simple in fact the least complex technique is preferred.
- The delay introduced in compression should be very small.
- Based on application requirements, there are two kinds of codecs - symmetric and asymmetric. Symmetric algorithms take the same amount of time to compress and uncompress files, while asymmetric codecs spend more time in one of the above processes. Asymmetric codecs are usually preferred, because the general user would like to compress the file once, and then uncompress it whenever needed, so that the decompression time to be small while the compression time can be allowed to be big because to compress the file once.

ISSN: 2349 - 6363

II. RELATED WORK

The Set Partioning in Hierarchical Trees (SPIHT) algorithm has made great strides in wavelet-based video coding [2], [3]. It has the potential of providing a scalable representation of a video coding in spatial resolution temporal resolution and quality with low computational complexity [4]. The Video transmission over the Internet by forward error correction (FEC) is proposed [7] to enable the quality streaming of real time video. However, FEC is ineffective in packet-based networks when a large number of packets may be lost in clusters and such loss exceeds the recovery capability of the FEC code. Retransmission [21] may be more suitable since it only introduces the overhead when needed. The existence of a backward channel from the decoder to the encoder also affects the deployment of some schemes. In applications such as broadcast, where there is no backward channel, none of the interactive error concealment techniques can be applied. However, the effectiveness of such techniques is limited by the available information. Also, this technique may be either too complicated for cost-effective implementation or introduce unacceptable processing delay for real-time applications. The forward-error-correcting (FEC) channel (RCPC) code combined with a single ARQ (automatic-repeat-request) is proposed [23] and proved to be an effective means for protecting the bitstream. There were two problems with this scheme: the noiseless reverse channel ARQ may not be feasible in practice; and, in the absence of channel coding and ARQ, the decoded sequence was hopelessly corrupted even for relatively clean channels. ARQ however may not be feasible in certain sceneries and has the unfortunate consequence of increasing traffic on already congested channels. An asymmetric tree structure is proposed [15] for utilization with an error resilient form 3-D SPIHT, called ERC-SPIHT with the fast error concealment scheme for embedded video bitstream using ERC-SPIHT. By using the simple averaging method in the root subband, the presence or absence of edges in the lost block of every image was detected. Then the interpolation scheme is used to recover the missing coefficients are estimated by the average value of their four adjacent neighbours which improves the rate distortion performance. But the video quality of this proposed method is very low. The Set Partitioning in Hierarchical Trees (SPIHT) [1] is proposed to combat packet loss the additional redundancy is introduced in each substream by adding a partial version of the other substreams. However, the problems with these solutions are that they only exploit the additional redundancy to estimate the missing coefficients but such additional redundancy may considerably decrease source coding efficiency and eventually yield a significantly reduced reconstruction quality.

Here the 3-D SPIHT algorithm for the efficient video transmission is used. The video coding system consists of two phases.

A. Compression

Wavelet Decomposition: The input video which is in the form of Group Of Frames (GOF) is decomposed into subbands by using wavelet filters.

3-D SPIHT Encoding: The wavelet coefficients which are obtained by the decomposition is coded using 3-D SPIHT algorithm.

B. Reconstruction

3-D SPIHT Decoding: When all the substreams are received at the decoder, it is necessary to decode all the coefficients to restore the original positions of the wavelet coefficeints.It is done by using 3-D SPIHT decoder.

Wavelet Reconstruction: In order to obtain the original frames the inverse wavelet transform is applied to the decoded coefficients.

III. THE NOVEL VIDEO CODING TECHNIQUES FOR ROBUST VIDEO TRANSMISSION

The Fig.1 shows the flow methodology for the efficient video transmission.

A. Wavelet Decomposition

Initially, a video which is represented as the group of frames or the images is transformed into the 3-D Wavelet transform. In general time ordered sequence images are known as video. An image is represented as a two-dimensional array of coefficients each coefficient representing the brightness level in that point. When looking from a higher perspective, it can't be differentiated between coefficients as more important ones, and lesser important ones. But thinking more intuitively, it can be done. Most natural images have smooth color variations, with the fine details being represented as sharp edges in between the smooth variations. Technically, the smooth variations in colour can be termed as low frequency variations and the sharp variations as high frequency variations. The low frequency components (smooth variations) constitute the base of an image, and the high frequency components (the edges which give the detail) add upon them to refine the image, thereby giving a detailed image. Hence, the smooth variations are demanding more importance than the details. Separating the smooth variations and details of the image can be done in many ways. One such way is the decomposition of the image using a Discrete Wavelet Transform (DWT).

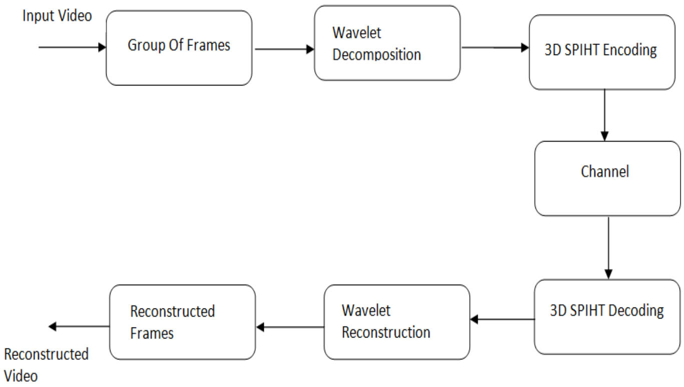


Figure 1. Flow Model for the efficient video transmission

A single stage of a 1D discrete wavelet transform (DWT) decomposes a 1D signal into a low pass signal and a high pass signal.

$$^{1D}W_n^{\ SR} \longrightarrow \left(\begin{array}{c|c} a_n & h_n \\ v_n & d_n \end{array}\right)$$

a_n, h_n, v_n, d_n - wavelet coefficients.

The Wavelet coefficients can be computed as follow.

$$a_n : a_{x,y,n} = \sum\sum h_i h_j{}^{1D} W_n^{SR}(2x-i)^{1D}W_n^{SR}(2y-j) \\ = hWh^T \tag{1}$$

$$h_n : h_{x,y,n} = \sum\sum g_i h_j{}^{1D} W_n^{SR}(2x-i)^{1D}W_n^{SR}(2y-j) \\ = gWh^T \tag{2}$$

$$v_n : v_{x,y,n} = \sum\sum h_i g_j{}^{1D} W_n^{SR}(2x-i)^{1D}W_n^{SR}(2y-j) \\ = hWg^T \tag{3}$$

$$d_n : d_{x,y,n} = \sum\sum g_i g_j{}^{1D} W_n^{SR}(2x-i)^{1D}W_n^{SR}(2y-j) \\ = gWg^T \tag{4}$$

$^{1D}W_n^{SR}$ denotes the wavelet coefficients resulting from the 1D Wavelet decomposition. h, g denote the low-pass and high-pass filters. Multidimensional wavelet decompositions are typically constructed by such 1D wavelet decompositions applied independently along each dimension of the image dataset, producing a number of subbands. The decomposition procedure can be repeated recursively on one or more of the subbands to yield multiple levels of decomposition of lower and lower resolution. The most commonly used multidimensional DWT structure consists of a recursive decomposition of the lowest resolution subband. The procedure is, a low pass filter and a high pass filter are chosen, and they exactly halve the frequency range between themselves. This filter pair is called the Analysis Filter pair.

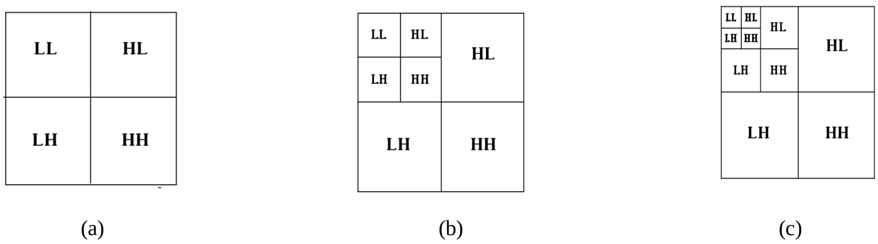


Figure 2. Wavelet Decomposition (a) One level (b) Two level (c) Three level

First, the low pass filter is applied for each row of data, thereby getting the low frequency components of the row. Now, the high pass filter is applied for the same row of data, and similarly the high pass components are separated, and placed by the side of the low pass components. This procedure is done for all rows. Next, the filtering is done for each column of the intermediate data. The resulting two-dimensional array of coefficients contains four bands of data, each labeled as LL (low-low), HL (high-low), LH (low-high) and HH (high-high). The LL band can be decomposed once again in the same manner, thereby producing even more sub bands which is shown in the figure 2. This can be done up to any level, thereby resulting in a pyramidal decomposition. The number of decomposition levels depends on the original image size. It may last until one coefficient remains to represent the entire image.

Note that a wide range of wavelet transform filters could be chosen at this step. For example, previously the Haar wavelet transform and the 9/7 biorthogonal wavelet filter (also called CDF 9/7) were used. One reason to choose the Haar wavelet is that computation is simpler and the results are easier to analyze mathematically. Another nice feature of the Haar wavelet is that it is a 2-tap filter. In addition to the Haar transform, the 9/7 biorthogonal wavelet filter also implemented. Compared to the 2-tap Haar filter, CDF 9/7 is a longer filter with up to nine taps and typically offers better performance on image and video coding. However CDF 9/7 biorthogonal wavelet filter's coefficients are all irrationals which increases the computational complexity of the DWT/IDWT .So this case is analysed with different wavelet filters. So here the Symlet wavelet transform, Spline Wavelet transform and Coiflet Wavelet transform are also experienced.

B. 3-D SPIHT Encoding

The SPIHT algorithm utilizes three basic concepts: (1) Searching for sets in spatial orientation trees in a Wavelet transform. (2) Partioning the wavelet coefficients in these trees into sets defined by the level highest significant bit in a bit-plane representation of their magnitudes. (3) Coding and transmitting bits associated with the highest remaining bit planes first. The actual algorithm used by SPIHT is based on the realization that there is really no need to sort all the coefficients. The main task of the sorting pass in each iterations to select those coefficients that satisfy the below condition $2^n \leq | c_{i,j} | < 2^{n+1}$.

This task is divided into two tasks. For a given value of n, if a coefficient satisfies the above condition then it is significant; otherwise it is called insignificant, but their number increases iteration to iteration. Since n keeps getting decremented. In the SPIHT coding technique both the encoder and decoder follows the same flow of execution. It is important to have the encoder and decoder test sets for significance in the same way. So the coding algorithm uses three lists.

- List of Significant Pixels (LSP),
- List of Insignificant Sets (LIS),
- List of Insignificant Pixels (LIP).

The LIP contains coordinates of coefficients that were insignificant in the previous sorting pass. In the current pass they are tested, and those that test significant are moved to the LSP.

In a similar way, sets in the LIS are tested in sequential order, and when a set is found to be significant, it is removed from the LIS and is partioned. The new subsets with more than one coefficient are placed back in the LIS, to be tested later, and the subsets with one element are tested and appended to the LIP or the LSP, depending on the result of the test. The refinement pass transmits the n^{th} most significant bit of the entries in the LSP. The steps involved in the 3-D SPIHT Encoding algorithm is given below.

Step 1Initialization:

Set the threshold. Set LIP to all root nodes coefficients. Set LIS to all trees. Set LSP to an empty set.

Step 2 Sorting pass:

1. Check the significance of all coefficients in LIP:
    1. If significant, output a sign bit, and move the coefficient to the LSP.
    2. If not significant, output 0.
2. Check the significance of all trees in the LIS according to the type of tree:
    1. For a tree of type D:
    1. If it is significant, output 1, and code its children:
        1. If a child is significant, outputs 1, then a sign bit, add it to the LSP.
        2. If a child is insignificant, output 0, and add the child to the end of LIP.
        3. If children have descendants, move the tree to the end of LIS as type L, otherwise remove it from LIS.
    2. For a tree of type L:

1. If it is significant, output 1, add each of the children to the end of LIS as an entry of type D and remove the parent tree from the LIS.
2. If it is insignificant, output 0.

Step 3 Refinement pass:

For each entry in the LSP, output the n^{th} most significant bit.

Step 4 Loop:

Decremented the threshold and go to step 2 if needed.

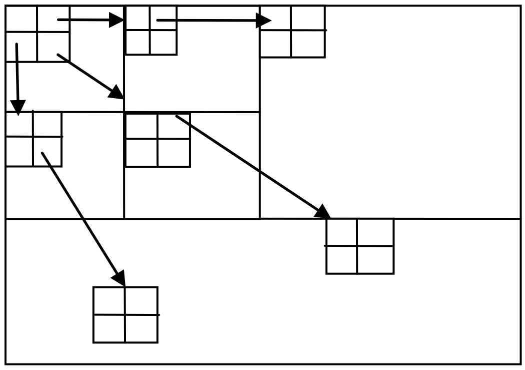


Figure 3. Dependencies in Spatial Orientation trees

C. 3-D SPIHT Decoding

The decoder recovers the ordering because its three lists are updated in the same way as those of encoder. When the decoder inputs data, its three lists are identical to those of the encoder at the moment it output that data. The decoder however, has to display the image and update the display in each iteration. In each iteration, when the coordinates (i,j) of a coefficient c(i,j) are moved to the LSP as an entry, it is known to both encoder and decoder. As a result the best value that the decoder can give the coefficient that is being reconstructed. During the refinement pass, when the decoder inputs the actual value of the nth bit, it improves the value 1.5 * 2^n by adding 2^n-1 to it (if the input bit is 1); otherwise subtracting 2^n-1(if the input bit is 0). This way the decoder improves the appearance of the image during both the sorting and refinement passes.

D. Wavelet Reconstruction

The reconstruction of video signal from its coefficients is called inverse 3-D WT. In other words, the synthesis of video signal implements the (t + 2-D) scheme in reverse order. That is, the spatial (2-D) composition is followed by the temporal composition (t). The former and the latter are exactly inverses of the 2-D and 1-D WTs, respectively. The composite starts with one level spatial composition of subbands at level-3. Then, resultant low and high subbands are temporally composed. This entire process is described as one-level spatial-temporal synthesis, and creates the LLL subband at level-2. It can be regarded as merging of sub-cubical at level-3 to a larger volumetric data that fits exactly to empty location of the cubic at level-2. The synthesis is an iterative procedure. The eight subbands, including LLL at level-2 are spatial-temporal composed, yielding the LLL subband at level-1. Finally, the composition is performed once more for 8 subbands to obtain the original video.

IV. SIMULATION RESULTS

The performance of the proposed system is investigated in this section.we compare the performance of the wavelet filters by finding their PSNR (Peak Signal to Noise Ratio). The Peak Signal to Noise Ratios (PSNRs) are used as a quality measure of the reconstructed frames.

The experiment is conducted by using the YUV video sequence. The wavelet coefficients are independently encoded at a coding rate of *c* bits/pixel (bpp).The higher the PSNR, the better the quality of the compressed or reconstructed image. The Peak Signal to Noise Ratio (PSNR)

$$\text{PSNR (dB)} = 10 * \log \left(\frac{255^2}{MSE}\right) \quad (5)$$

is used as the distortion metric where MSE is the mean squared error between the original and the reconstructed video sequence.

Where MSE - Mean Squared Error, x-the number of rows, y-the number of columns and A_{ij}, B_{ij} are the original image and the reconstructed image. And another one important factor is compression ratio. Compression Ratio is used to know how much the image gets compressed. It is defined as the ratio between the uncompressed image size and the compressed image size. Here we estimate the compression ration along with the PSNR estimation. The compression is estimated by using the following equation.

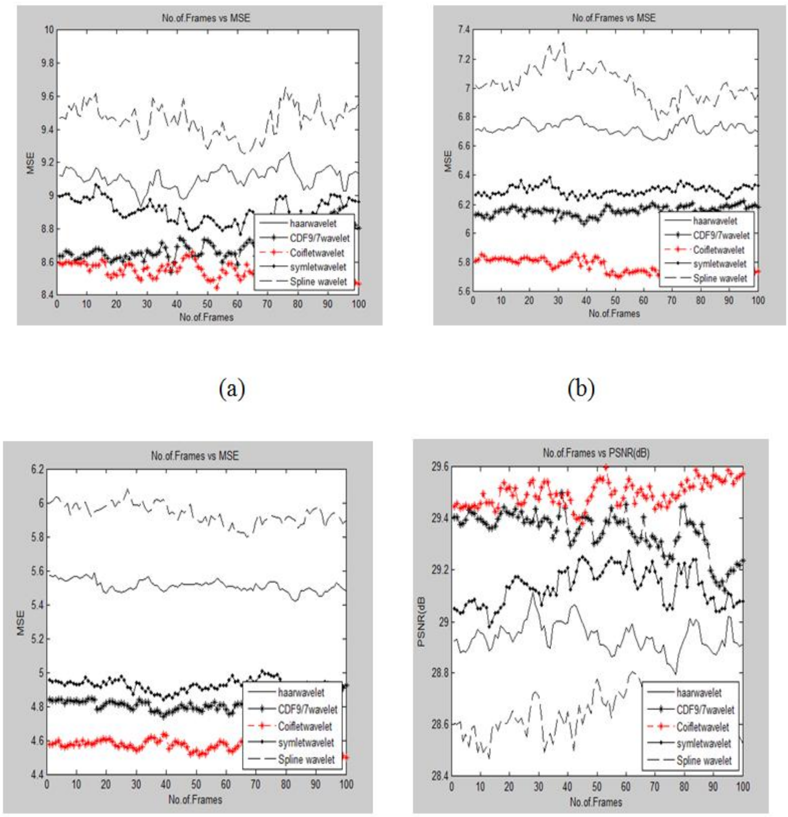
$$\text{Compression Ratio} = \frac{\text{Uncompressed Image Size}}{\text{Compressed Image Size}} \tag{6}$$

(a) (b) (c)

(d) (e) (f)

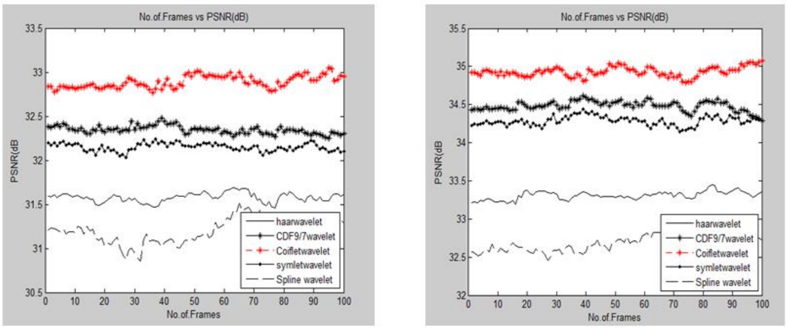
Figure 4. (a) 100th frame of akiyo video sequence (b),(c),(d),(e),(f) Reconstructed frames using Coiflet Wavelet transform, CDF 9/7 Wavelet transform, Haar Wavelet transform, Symlet Wavelet transform, Spline Wavelet transform respectively.

In the above Fig. 4 the 100th frame akiyo video sequence is taken and the reconstructed frames of that 100th frames are shown which has been analyzed by using the multi resolution transforms.



(a) (b)

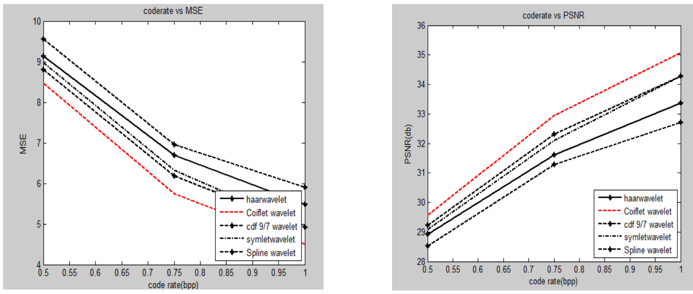
(c) (d)



(e) (f)

Figure 5. (a),(b),(c) Number of Frames vs MSE at bit rates 0.5,0.75,1 respectively,(d),(e),(f) Number of Frames vs PSNR at bit rates 0.5,0.75,1 respectively

Fig.5 shows the MSE versus number of frames and PSNR versus number of frames which compares the performance of the five different wavelet transforms for the three different code rates of the akiyo video sequence. In the Fig 5(a) and Fig 5(d), the MSE versus number of frames and PSNR versus number of frames are calculated at the code rate 0.5 and they are analysed for the 5 different wavelet transforms for the akiyo video sequence. Similarly figure 5(b) and figure 5(e) are plotted at the code rate 0.75 and figure 5(c) and figure 5(f) are plotted at the code rate 1.



(a) (b)

Figure 6. PSNR (dB) and MSE of Akiyo video sequence for different code rates with multiresolution wavelet transform (a) MSE (b) PSNR (dB).

Fig. 6 shows the average MSE and PSNR for Haar, CDF 9/7, Coiflet, Symlet, Spline wavelet transforms with code rates of 0.5, 0.75 and 1of akiyo video sequence.

From Fig. 5 and Fig. 6, among those five different wavelet transforms (Haar, CDF 9/7, Coiflet, Symlet, Spline) the Coiflet Wavelet transform gives the better result.

V. CONCLUSION

The efficient approach for video transmission has been presented here. However video compression is a very challenging problem for the existing algorithm. Previously the Haar wavelet transform and the CDF 9/7 wavelet transform were used. Compared to the Haar wavelet transform, CDF 9/7 offers better performance on image and video coding. However CDF 9/7 wavelet transform has its coefficients as irrational which increases the computational complexity as well as decreases the image quality. In order to improve the video quality, the system is analysed with different wavelet transform and the good video quality is obtained by using the 3-D SPIHT algorithm which has proved so successful in still video coding. Three dimensional spatio-temporal orientation trees coupled with powerful SPIHT video coder so efficient that it provides comparable performance to H.263 standard videos. And hence it is proved that Coiflet wavelet transform gives the better results. The compression ratio is estimated by taking the ratio between the original video frames and the encoded frames. Here the compression ratio is obtained appreciably as 4:1. The MSE (Mean Square Error) is obtained for the different code rate. And the MSE (Mean Square Value) and PSNR (Peak Signal to Noise Ratio) comparison plots are plotted. And the experimental results shows multiplication-free DWT/IDWT using the Coiflet Wavelet Transform are promising in the realization of real-time image and video.

REFERENCES

[1] A. C. Miguel, A. E. Mohr, and E. A. Riskin, “SPIHT for generalized multiple description coding”, in Proc. IEEE Int. Conf. Image Process., vol. 3. Oct. 1999, pp. 842–846.

[2] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees", IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 3, pp.243–250, Jun. 1996.

[3] B.J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees", in Proc. IEEE Data Compression Conf., Mar. 1997, pp.251–260.

[4] B.J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)", IEEE Trans. Circuits Syst. Video Technol., vol.10, no. 8, pp. 1374– 1387, Dec. 2000.

[5] C. D. Creusere, "A new method of robust image compression based on the embedded zero tree wavelet algorithm", IEEE Trans. Image Process., vol. 6, no. 10, pp. 1436–1442, Oct.1997.

[6] Dong Wei, Jun Tian, Raymond O. Wells, Jr., and C. Sidney Burrus , "A New Class of Biorthogonal Wavelet Systems for Image Transform Coding", IEEE transactions on image processing, vol. 7, no. 7, July 1998.

[7] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: Challenges and approaches", Proc. IEEE, vol. 88, no. 12,pp. 1855–1877, Dec. 2000.

[8] I. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video", IEEE Trans. Image Process., vol. 12, no. 10, pp.1211–1225, Oct. 2003.

[9] J. E. Fowler and B. Pesquet-Popescu, "An overview on wavelets in source coding, communications, and networks", EURASIP J. Image Video Process., vol. 2007, no. 60539, p. 27, 2007.

[10] Jie Zhu and Richard M.Dansereau, "Error-Resilient and Error Concealment 3-D SPIHT for Multiple Description Video Coding With Added Redundancy", IEEE Transactions on Circuits and Systems for Video technology, vol. 22,no. 6, June 2012.

[11] J. K. Rogers and P. C. Cosman, "Robust wavelet zero tree image compression with fix-length packetization", in Proc.IEEE Data Compression Conf., Mar. 1988, pp. 418-427.

[12] J. Kim, R. M. Mersereau, and Y. Altunbasak, "A multiple- substream unequal error-protection and error-concealment algorithm for SPIHT coded video bitstreams", IEEE Trans. Image Process., vol. 13, no. 12, pp. 1547–1553, Dec. 2004.

[13] J. Zhu, R. M. Dansereau, and A. Cuhadar, "Error-resilient and error concealment 3-D SPIHT video coding with added redundancy", in Proc.Int. Conf. Image Signal Process., LNCS 6134. Jun. 2010, pp. 351–358.

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transformation", IEEE Trans. Image Process., vol. 1, no.2, pp. 205–220, Apr. 1992.

[15] S. Cho and W. A. Pearlman, "A full-featured, error-resilient, scalable wavelet video codec based on the set partitioning In hierarchical trees (SPIHT) algorithm", IEEE Trans. Circuits Sys.Video Tech., vol. 12,no. 3, pp. 157-171, Mar. 2002.

[16] S. Cho and W. A. Pearlman, "Error resilience and recovery in streaming of embedded video", J. Signal Process., vol.82, no. 11, pp. 1545–1558,Nov. 2002.

[17] S. Cho and W. A. Pearlman, "Error resilient video coding with improved 3-D SPIHT and error concealment", in Proc.SPIE/ID&T Int. Conf.Electron. Imaging, vol. 5022. Jan. 2003, pp. 125–136.

[18] V. K. Goyal, "Multiple description coding: Compression meets the network", IEEE Signal Process. Mag., vol.18, no5, pp. 74–93, Sep.2001.

[19] W. A. Pearlman, B.J. Kim, and Z. Xiong, "Embedded video subband coding with 3D SPIHT", in Wavelet Image and Video Compression, P.Topiwala, Ed. Norwell, MA: Kluwer, 1998.

[20] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization", in Proc.SPIE VCIP Conf., Jan.1999, pp. 998–1008.

[21] Y. Wang and Q.-F. Zhu, "Error control and error concealment for video communications: A review", Proc. IEEE,vol. 86, no. 5, pp. 974–997,May 1998.

[22] Y. Guo, Y. Chen, Y.K. Wang, H. Li, M. M. Hannuksela, and M.Gabbouj, "Error resilient coding and error concealment in scalable video coding", IEEE Trans. Circuits Syst. Video Tech., vol. 19, no. 6, pp.781–795, Jun.2009.

[23] Z.Xiong, B.J Kim, and W. A. Pearlman "Progressive video coding for noisy channels", In Proc. IEEE International Conference on Image Processing (ICIP '98), vol. 1, pp. 334- 337, October, 1998.