Wavelet Based Video Encoder Using KCDS

Sudhakar Radhakrishnan, Guruprasad Subbarayan, and Karthick Vikram
Department of Electronics and Communication Engineering, PSG College of Technology, India

Abstract: *Video coding schemes for low bit-rate is of high importance and traditional coding schemes which use block transforms suffer from blocking artefacts. Here we propose a Video Codec based on Wavelet transform and hence performance of the coder is superior to other block transform based codecs. The Wavelet coefficients are coded using the computationally simpler no list SPIHT whose performance is similar to that of set partitioning in hierarchical trees. Motion estimation is done using the recently proposed kite-cross-diamond search algorithm which is the fastest among the block matching algorithms. The codec is ideally suited for sequences with smooth and gentle motion of the video conferencing kind. Simulation results are provided to evaluate the performance of the codec at various bit-rates. The codec is scalable in terms of bandwidth requirement which means only one compressed bit stream is produced for different bit-rates. The use of NLS makes the codec scalable since it has the embedded coding property. However for resolution scalability different compressed files are required.*

Keywords: *Codec, motion estimation, SPIHT, video encoder, wavelets.*

Received October 6, 2007; accepted January 6, 2008

1. Introduction

Video coding schemes are generally transformed based coding schemes to exploit the spatial redundancy that exists within a frame followed by a feedback loop to exploit the temporal redundancy that exist between frames. Here we propose a Video Codec based on Wavelet transform and hence performance of the coder is superior to other block transform based codecs.

Like typical codecs we use the intra-coded I-frames and predictive P-frames. The intra-coding of frames is done using transform coding techniques [6, 7]. Wavelet transformation is preferred due absence of blocking artefacts and superior performance compared to other block transform schemes. The Wavelet coefficients are coded using the computationally simpler No List SPIHT (NLS) [8] whose performance is similar to that of Set Partitioning In Hierarchical Trees (SPIHT) [6]. We use biorthogonal 9/7 wavelet due to its superior performance compared to other wavelets when NLS is used. Using the bi9/bi7 wavelet 55% of the coefficients becomes zero and hence the quantization phase becomes more efficient. No entropy coding stage is necessary since the output of SPIHT is of high entropy and arithmetic coding stage produces an improvement of only 4-5% for low bit-rate and 0-1% for high bit-rate videos. The main bottle neck in a video codec exists in three places namely the motion estimation phase, the transform coding phase and the entropy coder module. Here we use NLS which is a computationally simpler transform coding technique and entropy coding is not used due to the reasons stated earlier. Motion estimation [1] is done using the recently proposed Kite-Cross-Diamond Search (KCDS) [3] algorithm which is the fastest among the block matching algorithms. KCDS achieves about 39% improvement in speed compared to Cross-Diamond Search (CDS) [2] for sequences with low motion content and 8-27% for sequences with high motion content. Hence the codec is fast and its performance is superior to other block transform based codecs.

The codec is ideally suited for sequences with smooth and gentle motion of the video conferencing kind. Simulation results are provided to evaluate the performance of the codec at various bit-rates. Our proposal is to use SPIHT for coding the P-frames in the video encoder and to do motion estimation using the KCDS algorithm for those sequences with smooth motion content. Unlike MPEG blocking artefacts are absent in our codec because of the use of wavelets in the transformation phase.

2. Wavelet based Compression of I-Frames Using NLS

The I-frames are intra-coded which means that the frames are compressed independently of other frames. Compression is achieved by exploiting the spatial redundancy that exists within those frames. For compression of I-frames we use a transform coding technique based on wavelets. Wavelet decomposition is applied to the image and the resulting coefficients are quantized using the NLS algorithm. SPIHT basically exploits the statistical dependencies that exist between the wavelet coefficients and only those coefficients that have significant magnitude are coded and thus compression is achieved.

The compression of I-frames can be regarded as compression of individual images. The image is wavelet decomposed and the wavelet decomposed coefficients are quantized using the NLS algorithm. The performance of the image compression system shown in Figure1 was tested at various levels of wavelet decomposition with different wavelets. The encoding process is done for various compression ratios (CR) the results are used to ascertain the ideal conditions for the compression of I-frames or intra coded frames. NLS is a variant of the popular SPIHT algorithm and operates without the use of lists unlike SPIHT algorithm.

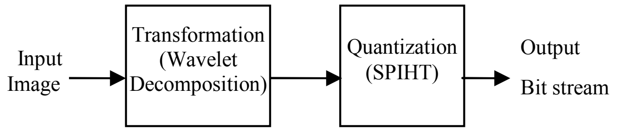


Figure 1. I-frame compression block diagram.

In SPIHT the use of lists causes a variable, data-dependant memory requirement and there is need for memory management as list nodes are created or moved from one list to another list. Since no lists are involved the NLS algorithm is computationally simpler compared to SPIHT. Further NLS yields similar results like SPIHT but at a 56 % increase in the memory requirement [8].

Instead of using lists NLS uses 4 bit markers for each pixel to keep track of the coding and partitioning information. This use of the state table markers is the major reason for the increase in memory requirement. Unlike SPIHT which operates on a two dimensional array of coefficients, NLS uses the linear indexing technique where the coding process is done on the coefficients stored in a linear array. Such use of linear array makes skipping of child trees easy and also the parent-child relationship in Hierarchical Trees [6, 7] is simplified. Hence the algorithm becomes less complex compared to SPIHT. The three passes used in NLS are

- Insignificant pixel pass
- Insignificant set pass
- Refinement pass

Encoding process begins with the most significant bit plane found by scanning the wavelet coefficients and continues till a bit budget is met. The bit budget is nothing but the size of the compressed image file. As the partitioning rules in NLS are same as those of SPIHT the same output bits are produced but in a different order. Wheeler and William Pearlman in [8] have given results to prove that NLS performance is similar to that of SPIHT. The embedded coding property [8] makes NLS ideal for use at different rates.

The image compression system is tested using four basic test images Lena, Barbara, Boats and House at different resolutions. The image compression system works with any image resolution. Three levels of decomposition 3, 4 and 5 are used for testing the image compression system with Compression Ratios (CR) ranging from 80:1 to 8:1. The system is tested using wavelets like Haar, db4, la8, bi5/bi3 and bi9/bi7. The PSNR performance of the codec is superior at 5 levels of decomposition than at any other level. For all the wavelets an increasing trend is observed in the PSNR performance as the number of levels of decomposition increases. Hence we use 5 levels of decomposition when coding the I-frames. The bi9/bi7 wavelet gives good PSNR performance compared to other wavelets. Many numerical studies also have indicated that biorthogonal 9/7 wavelet gives best distortion rate performance for wavelet transform codes. Hence we use bi9/bi7 for compressing frames of the video.

Further the experimental results too indicated that bi9/bi7 yields better uniform PSNR performance compared to other wavelets. Hence we chose bi9/bi7 wavelet at 5 levels of decomposition for compression of I-frames. At a rate of 1 bpp (bits per pixel) the image after compression looks much like the original image. Hence for compression of I-frames we use bi9/bi7 wavelet at 5 levels of decomposition keeping the CR as 8:1. This CR can be varied but at 1 bpp the image is of high quality and ideal to function as a reference frame for coding other frames.

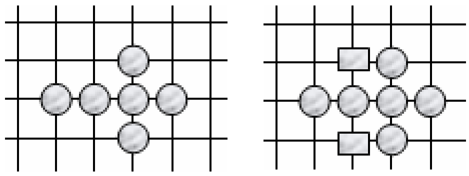
3. Video Compression

A set of images representing a moving scene is a set of temporal samples. In typical scene there will be lot of similarity between successive frames. This means we will have success in predicting one frame from the other. A reference frame is used and other frames are predicted from the reference frame. The reference frames is the intra coded I-frame and the predicted frames are the P-frames [1]. This temporal redundancy that exists between frames of a video sequence is exploited by video codecs. Video compression algorithms exploit the spatial redundancy and temporal redundancy.

Motion estimation is the process of obtaining the displacement of the block of pixels of the P-frame from the I-frame. This displacement is obtained by searching for the matching block which is done by Block Matching Algorithm (BMA). The displacements of the blocks are called the Motion Vectors. Motion vectors and the I-frame are used in producing the predicted version of the frame. Block Distortion Measure (BDM) gives the level of correlation between the blocks. We use the BDM Mean Absolute Difference (MAD) for our codec. We use a typical block size of 8 X 8 for all the three fields namely the Y, Cb and Cr. The difference between the predicted version of the P-frame and the original P-frame is called as the residual. We use the KCDS for motion estimation and KCDS finds the displacement of the blocks of the P-frame from the I-frame.

3.1. Kite Cross Diamond Search

KCDS [3] is an improved version of the CDS [2] and the Small-Cross-Diamond Search (SCDS) [5] algorithm. Unlike traditional block search algorithms KCDS uses an asymmetric Kite pattern, shown in Figure 2 so that the speed of search of stationary and quasi-stationary blocks is boosted. This makes the KCDS algorithm efficient for use with sequences with smooth and gentle motion. Experimental results show that KCDS achieves a 39 % improvement in speed compared to CDS for low motion videos [3]. Video conferencing sequences are best examples of video content with smooth and gentle motion. The kite search pattern and the biased corner pattern used in the algorithm is the main reason for the efficiency of the algorithm.



(a) Left Kite Search Pattern pattern. (b) KSP with Biased Corner.

Figure 2. KSP pattern.

The direction of the kite is ascertained by the initial direction of motion detected by the small cross pattern. The direction of Kite is nothing but the direction in which the dart points. KCDS has five execution steps and after five steps the motion vector corresponding to a block is obtained. This motion vector is calculated for all the blocks of the P-frame.

Once the motion estimation is done using the KCDS algorithm the predicted frame is reconstructed using the Motion Vectors and the I-Frame. The difference between the original P-frame and the predicted P-frame is called as the residual. This residual is transmitted along with the motion vectors to produce good quality predictive frames. The residual is also coded using NLS so that the size of the residual to be transmitted decreases. The rate at which the residual is coded is much less compared to the rate at which the I-frame is coded. The codec is tested at various rates of compression of the P-frame and generally as the rate of encoding increases the video quality also increases.

3.2. Proposed Video Encoder

Figure 3 is a simple block diagram representation of the proposed Video encoder. In case of an I-frame the frame is intra coded using the NLS algorithm and in case of a P-frame motion estimation is done and the residual is encoded. The feedback loop is same as the decoder and the only difference is that the residual will be added to the predicted version of P-frame instead of being subtracted. SPIHT is generally considered to be computationally complex and hence it has not been used for intra-coding of frames. But the NLS the implementation is easy and fast and is suitable for hardware implementation [8] and the disadvantage being an increase in memory requirement of about 56%. But this is a trade-off made for fast implementation of SPIHT. Further the motion estimation using KCDS boosts the speed of the encoder.

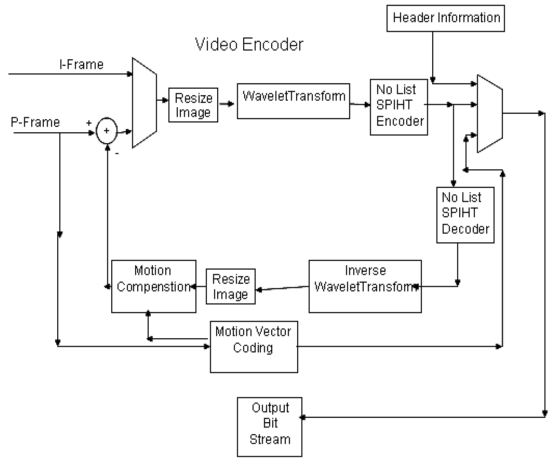


Figure 3. Proposed video encoder.

The video codec is ideally suited for sequences with low motion content but even with sequences that have high motion content there is at least an 8% improvement in the motion estimation phase with respect to the CDS algorithm [3]. The video encoder is tested using different video sequences. All sequences chosen contain a human face and are generally of the low motion content type. The codec is tested for both CIF and QCIF sequences indicated in Table I and the codec works well with any video resolution. The codec is tested for the first 120 frames of the sequences mentioned. Every eleventh frame is considered as a reference frame and the reference frame is change when there is a scene change i.e. when the PSNR performance falls below the tolerance level. Three compression ratios 80:1, 26.7:1, and 20:1 is used for the encoding of the residual and the I-frames are generally encoded at a CR of 8:1.

Table 1. Video sequences tested (120 frames).

Sequence	Resolution
Claire Akiyo Miss America Trevor Carphone	QCIF (176 X 144)
Foreman	QCIF (176 X 144) and CIF (352 X 288)

4. Results and Inferences

The intra-frame coding is done using the NLS coding on the wavelet transformed image using the bi9/bi7 wavelet at 5 levels of decomposition keeping the CR as 8:1. Processing is done by considering each frame to be individual images. The reference frames are intra coded and the predictive frames are predicted from the reference frames and only the error is coded using the NLS algorithm. We tested the video codec for 120 frames of each video. The PSNR performance is obtained by comparing the reconstructed frame from the original frame. The residual is coded using the

NLS on the wavelet transformed residual image using the bi9/bi7 wavelet at 5 levels of decomposition at different CR. The results for average PSNR of 120 frames of various sequences and also for a Group of frames were studied. The codec adapts well to scene changes as in case of Trevor sequence. Sequences with gentle motion like Miss America, Claire and Trevor yield best results. At a P-frame CR of 20:1 the compressed sequence approximates the original sequence. Sample reconstructed frames are shown in Figures 4, 5, 6, and 7.

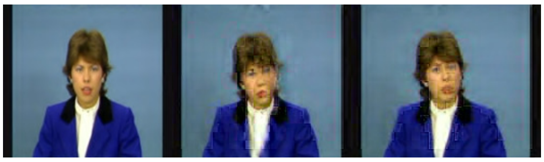


Figure 4. Reconstructed frames 1, 64 and 120 of Claire Sequence at I-frame rate 1bpp and P-frame rate 0.4bpp.

Figure 5. Reconstructed frames 1, 64 and 120 of Miss America Sequence at I-frame rate 1bpp and P-frame rate 0.4bpp.

Figure 6. Reconstructed frames 1, 64 and 120 of Trevor Sequence at I-frame rate 1bpp and P-frame rate 0.4bpp.

Figure 7. Reconstructed frames 1, 64 and 120 of grandma sequence at I-frame rate 1bpp and P-frame rate 0.4bpp.

Blocking artefacts are nearly absent as Wavelet transform is used and the errors do not reinforce one another but are distributed throughout the frame. This absence of block based transform makes the codec produce a better picture quality. It is evident from Figure 8 that the Miss America sequence which has smooth and gentle motion than all other sequences has higher PSNR values which indicate that the encoder performs well with such sequences. The picture quality of Miss America sequence is excellent compared to other sequence due to the presence of more number of stationary blocks. The rates indicate the bits per pixel (bpp) used for encoding the residual image. The codec works well with the sequences that contain stationary and quasi-stationary blocks. The block size and the search window can be altered to make the codec perform effectively with fast motion sequences too but the speed of motion estimation is affected.

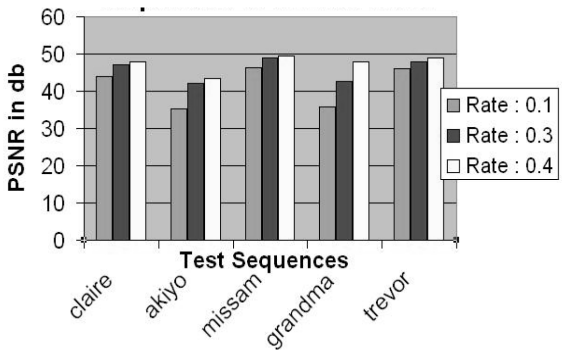


Figure 8. PSNR performance of the codec for various Video sequences.

5. Conclusion

SPIHT works well with natural images [1] and the images produced by SPIHT are recognizable even at a low rate of 0.1bpp. Higher rates yield better image qualities and the images encoded at a rate of 1bpp nearly equal the original image.

KCDS is the motion estimation algorithm used and it provides a performance improvement of 39% for sequences with smooth and gentle motion compared to CDS. This makes the Codec ideal for implementation to video conferencing which has very less motion content. The embedded coding property of SPIHT makes the codec produce video at any rate desired. The codec also works with sequences of any resolution and there is absence of blocking artefacts as wavelets are used. This makes the codec ideal for implementation to video conferencing.

References

[1] Koga T., Linuma K., Hirano A., Lijima J., and Ishiguro T., "Motion Compensated Inter Frame Coding for Video Conferencing," *in Proceedings* of *Natural Telecommuncation Conference*, New Orleans, pp. G5.3.1-G5.3.5, 1981.

[2] Lam W., Po M., and Cheung H., "A New Cross-Diamond Search Algorithm for Fast Block Matching Motion Estimation," *in Proceedings of IEEE Transaction on Neural Networks and Signal Processing*, pp.1262-1265, 2003.

[3] Lam W., Po M., and Cheung H., "A Novel Kite Cross Diamond Search for Fast Block Matching Motion Estimation," *in Proceedings of the IEEE*

Transaction on Circuits and Systems, pp.729-731, 2004.

[4] Pearlman A., "Trends of Tree-Based Set Partitioning Compression Techniques in Still and Moving Image Systems," *in Proceedings of Picture Coding Symposium*, Korea, pp. 1-8, 2001.

[5] Po M. and Cheung H., "A Novel Small-Cross-Diamond Search Algorithm for Fast Video Coding and Video Conferencing Applications," *in Proceedings of the IEEE Transaction on Image Processing*, USA, pp. I-681, 2002.

[6] Said A. and Pearlman W., "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchal Trees," *in Proceedings of IEEE Transactions on Circuits and Systems for Video Technology*, USA, pp. 243-250, 1996.

[7] Shapiro M., "Embedded Image Coding Using Zero Trees of Wavelet Coefficients," *in Proceedings of IEEE Transactions on Signal Processing*, pp. 3445-3462, 1993.

[8] Wheeler W. and Pearlman A., "SPIHT Image Compression without Lists," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2000)*, Istanbul, Turkey, pp. 769-772, 2000.

[9] Radhakrishnan S. and Subramaniam J., Novel Image Compression Using Multiwavelets with SPECK Algorithm," *in International Journal of Information Technology,* vol. 5, no. 1, pp. 45-51, 2008.

Sudhakar Radhakrishnan received the BE in electronics and communication engineering from Madras University in the year 1990, and ME communication systems from Thiagarajar College of Engineering, Madurai, in the year 2001. He has completed his PhD from Anna University.

Guruprasad Subbarayan received the BE in electronics and communication engineering from PSG College of Technology in the year 2007. He is currently working in the network infrastructure team of an IT firm. His areas of interest include digital signal processing, digital image processing, and embedded systems.

Karthick Vikram received the BE in electronics and communication engineering from PSG College of Technology in the year 2007. He is currently working in an IT projects team in TCS. His areas of interest include digital signal processing and computer networking.