Implementation of Wavelet-based Video Codec

Latt Latt Hlaing

Department of Electronic Engineering, Mandalay Technological University, Mandalay

Abstract - **Bringing video pictures into the digital age introduces one major problem. Uncompressed digital video pictures take up enormous amounts of information. Video compression (or video coding) is an essential technology for applications such as digital television, DVD-Video, mobile TV, videoconferencing and internet video streaming. An encoder converts video into a compressed format and a decoder converts compressed video back into an uncompressed format. Wavelets provide a mathematical way of encoding information in such a way that it is layered according to level of detail. These layers can be stored using a lot less space than the original data. The objective of this paper is to implement and evaluate the effectiveness of wavelet based video compression techniques using MATLAB. After DWT was introduced, run length codec algorithms was proposed to compress the transform coefficients. The image frame is represented as a stream of bits. The run length code is used to encode a string of pixels of the same colours by the number of repetitions in each string. The performance parameter such as Compression Ratio (CR) is evaluated based on the algorithm. Comparisons amongst the various wavelet names are carried out on the basis of calculated performance parameter. Mentioned technique is compatible for video formats like .MPEG, .AVI, .MOV etc making it more versatile.**

1. INTRODUCTION

With the increasing growth of technology and the entrance into the digital age, we have to handle a vast amount of information every time which often presents difficulties. Video compression is the process of encoding information using fewer bits. Compression is useful because it helps to reduce the consumption of expensive resources such as hard disk space or transmission bandwidth. The video is actually a kind of redundant data i.e. it contains the same information from certain perspective of view. By using data compression techniques, it is possible to remove some of the redundant information contained in images. Image compression minimizes the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. The reduction in file size allows more images to be stored in a certain amount of disk or memory.

The compression offers a means to reduce the cost of storage and increase the speed of transmission. Video compression is used to minimize the size of a video file without degrading the quality of the video. Over the past few years, a variety of powerful and sophisticated wavelet based schemes for image and video compression have been developed and implemented [1]. Wavelets are a mathematical tool for hierarchically decomposing functions. Wavelet-based coding provides substantial improvements in picture quality at higher compression.

The discrete wavelet transform (DWT) [1], [2] has gained wide popularity due to its excellent decorrelation property, many modern image and video compression systems embody the DWT as the intermediate transform stage. After DWT was introduced, run length codec algorithms was proposed to compress the transform coefficients as much as possible but a compromise must be maintained between the higher compression ratio and a good perceptual quality of image.

2. OVERVIEW OF THE VIDEO CODEC

Figure 1 shows the block diagram of the implemented video encoder and decoder. This section briefly describes each component of the encoder and decoder. Our coding scheme is basically a transform coder. The transform coder consists of the 2D discrete wavelet transform (DWT), and a lossless run length coding step which compacts the transform coefficients produced by the thresholding.

At the encoder, using the DWT, each video frame is decomposed into 10 frequency subbands. Then, each of the resulting subbands is encoded by means of an optimally designed uniform threshold followed by an optimally designed run length encoder. The output of the encoder is a bit stream consisting of the output of the run length encoders. The encoding method produce an efficient, compact binary representation of the information. The encoded bitstream can then be stored and/or transmitted.

At the decoder, the received bit stream is used to decode by run length decoder. A video decoder receives the compressed bitstream, decodes each of the syntax elements by run length decoder and extracts the information described above (transform coefficients). This information is then used to reverse the coding process and recreate a sequence of video images. Then, the inverse DWT (IDWT) is used to reconstruct each video frame. Finally the reconstructed frames are recombined to sequence of frames and output to video file.
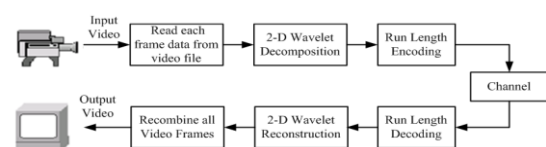

Figure 1. Video coding and decoding process

The video is represented as a sequence of frames, and each frame is treated as a two-dimensional array of pixels (pels). The color of each pel is consists of three components. Discrete Wavelet Transform is carried out by decomposing the image into four sub bands (LL, LH,

HL and HH) use separable wavelet filters and critically subs sampling the output. HH sub band gives the details of diagonal, HL sub band provides horizontal details and the LH sub band provides vertical details. The next coarser level of coefficients is obtained by decomposing the low frequency Sub band LL as shown in Figure 2.

Downsampling and Upsampling are widely used in image display, compression, and progressive transmission. Downsampling is the reduction in spatial resolution while keeping the same two-dimensional (2D) representation. It is typically used to reduce the storage and/or transmission requirements of images. Upsampling is the increasing of the spatial resolution while keeping the 2D representation of an image. It is typically used for zooming in on a small region of an image, and for eliminating the pixelation exact that arises when a low-resolution image is displayed on a relatively large frame. Run length coding is a proven technique for coding wavelet transform coefficients.
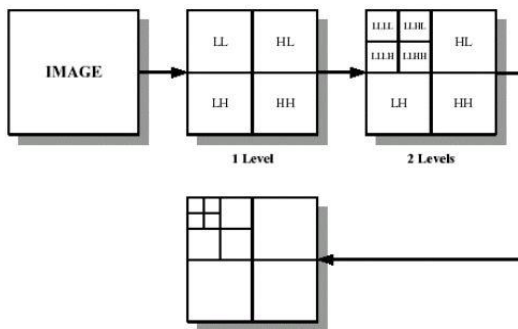

Figure 2. Decomposition of image frame from level 1 to 3

2.1. 2-D Discrete-Wavelet Transform

Figure 3 shows the 2-D DWT block of the encoder. The 2-D DWT block consists of three levels of decomposition as illustrated in Figure 3(a). Clearly, the specific decomposition used here results in 10 subbands. Each level of decomposition, represented by the operation A in Figure 3(a), is described further in terms of simpler operations in Figure 3(b). Specifically, A consists of low-pass and high-pass filtering (H and G) in the row direction and subsampling by a factor of two, followed by the same procedure on each of the resulting outputs in the column direction, resulting in four subbands.

The H and G filters ("Image Coding Using Wavelet Transform"[1]) are finite-impulse-response (FIR) digital filters. The specific input-output relationship for one level of DWT decomposition of a 1-D sequence X(n) can be represented as

$$X_1(n) = \sum_k h_1(2n-k)X(k)$$

$$X_h(n) = \sum_k g_1(2n-k)X(k)$$

(1)

in which $X_l(n)$ and $X_h(n)$ represent, respectively, the outputs of the low-pass and high-pass filters. The resulting

2-D subbands after the 2-D DWT operation are labeled subband1 through subband 10.

To reconstruct a replica of the image frame, the decode subbands are then fed into the 2-D IDWT block. Figure 4 shows the details of the 2-D IDWT operation. The 2-D IDWT block consists of three levels of reconstruction as illustrated in Figure 4(a).
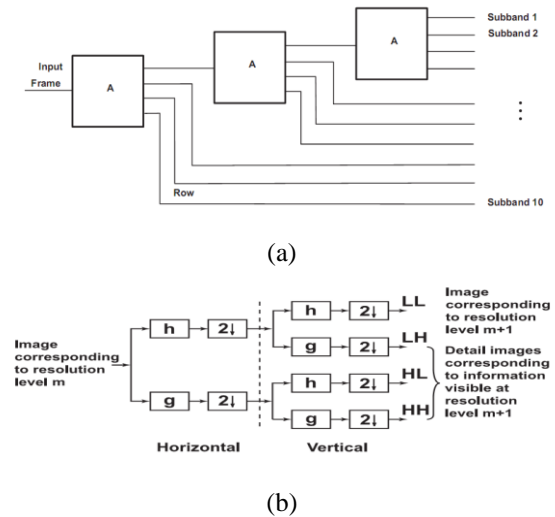

(a)

(b)

Figure 3. Discrete-Wavelet Transform

Each level of reconstruction, represented by the operation B in Figure 4, is described in terms of simpler operations in Figure 4(b). Specifically, B consists of up-sampling by a factor of two and low-pass and high-pass filtering in the column direction followed by the same procedure on the outputs of this process in the row direction, integrating four subbands into one wider band. The filters used for reconstruction ("Image Coding Using Wavelet Transform"[1]) are FIR digital filters. The specific input-output relationship for the reconstruction of the sequence X(n) is represented by

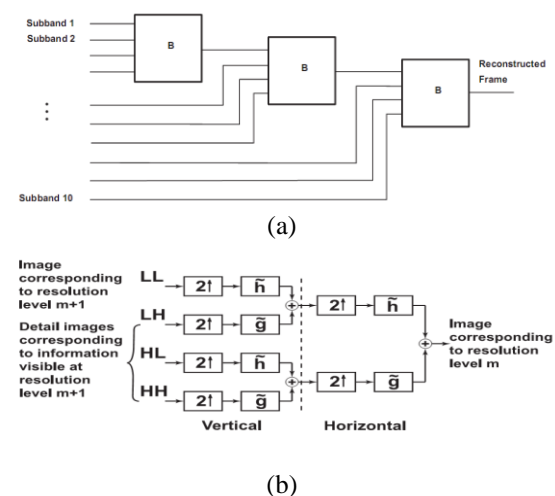$$X(n) = \sum_k h_2(2k-n)X_1(k) + g_2(2n-k)X_h(k) \quad (2)$$

(a)

(b)

Figure 4. Inverse Discrete-Wavelet Transform

2.2. Run Length Encoding

RLE is a natural candidate for compressing graphical data. A digital image consists of small dots called pixels. Each pixel can be either one bit indicating a black or white dot or several bits indicating one of several colors or shades of gray. We assume that these pixels are stored in an array called bitmap in the memory. Pixels are normally arranged in the bit map in scan lines. So the first bit map pixel is the dot at the top left corner of the image and the last pixel is the one at the bottom right corner. Compressing an image using RLE is based on the observation that if we select a pixel in the image at random there is a good chance that its neighbors will have the same color. The compressor thus scans the bit map row by row looking for runs of pixels of same color.

Example, the grayscale bitmap-

12,12,12,12,12,12,12,12,35,76,112,67,87,87,87,5,5,5,5, 5,5,1- - - - - - - -

Compressed Form-

9,12,35,76,112,67,3,87,6,5,1- - - - - - - - - - - - - - -

2.3. Coding Algorithm

In video compression, each frame is an array of pixels that must be reduced by removing redundant information. Standard video is normally about 30 frames/sec, but studies have found that 16 frames/sec is acceptable to many viewers, so frame dropping provides another form of compression.

When information is removed out of a single frame, it is called *intraframe* or *spatial* compression. But video contains a lot of redundant interframe information such as the background around a talking head in a news clip. *Interframe* compression works by first establishing a key frame that represents all the frames with similar information, and then recording only the changes that occur in each frame. The key frame is called the "I" frame and the subsequent frames that contain only "difference" information are referred to as "P" (predictive) frames. A "B" (bidirectional) frame is used when new information begins to appear in frames and contains information from previous frames and forward frames.

An interframe codec is one which compresses a frame after looking at data from many frames (between frames) near it, while an intraframe codec applies compression to each individual frame without looking at the others. The interframe compression provides high levels of compression but is difficult to edit because frame information is dispersed. Intraframe compression contains more information per frame and is easier to edit. Freeze frames during playback also have higher resolution. In this work intraframe compression technique is used.

2.3.1. Intra-frame Coding.

The term intra-frame coding refers to the fact that the various lossless and lossy compression techniques are performed relative to information that is contained only within the *current frame*, and *not relative* to any other frame in the video sequence. In other words, no temporal processing is performed outside of the current picture or frame. Non-intra coding techniques are extensions to these basics. Intra coding is very similar to that of a JPEG still image video encoder, with only slight implementation detail differences. Inter frame has been specified by the CCITT in 1988-1990 by H.261 for the first time. H.261 was meant for teleconferencing and ISDN telephoning.
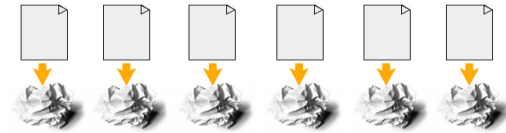

Figure 5. Intra-frame coding

Data is usually read from a video camera or a video card. The coding process varies greatly depending on which type of encoder is used (e.g., JPEG or H.264), but the most common steps usually include: transformation (e.g., using a DCT or wavelet) and Run Length encoding.

2.3.2. Inter-frame Coding.

An inter frame is a frame in a video compression stream which is expressed in terms of one or more neighbouring frames. The "inter" part of the term refers to the use of *Inter frame prediction*. This kind of prediction tries to take advantage from temporal redundancy between neighbouring frames allowing to achieve higher compression rates.
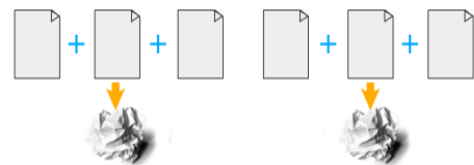

Figure 6. Inter-frame coding

3. PROPOSED ALGORITHM

Figure 7 illustrates the processes involved in the proposed video codec. Video compression using wavelet transform is performed with following steps:

- Load video file : Use the "VideoReader" function with the read method to read video data from a file into the MATLAB® workspace.
- Extract frames from video file
- Apply 2D forward wavelet transform using different wavelets.
- Encode wavelet coefficients.
- Decode wavelet coefficients.
- Perform inverse wavelet transform using different wavelets to reconstruct each frame
- Recombine all frame as a sequence of frames
- Playing video
- Save video file
- Calculate and displays uncompressed video file size, compressed video file size and compression ratio.
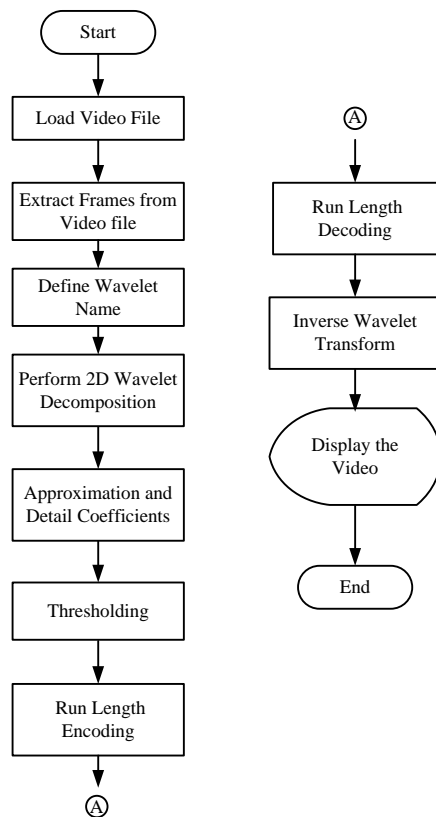
Figure 7. Flowchart of video codec

3. SIMULATION RESULTS

In order to investigate the performance of the proposed compression technique, MATLAB program is applied. Test the video file (*.avi) with the following Properties

- VidFormat = RGB24
- Number of Frames = 121
- Video frame Height = 288
- Video frame Width = 512
- frame_rate = 30
- BitsPerPixel=24
- Video length= 4.033 sec

The results are obtained through simulation by MATLAB. The experimental data is used to analyze and test the performance of coding scheme.

Figure 8 illustrates the extracted video frame. Each frame is decomposed as describe in Figure. The wavelet coefficients are decoded by run-length encoding method. After transmission, the encoded or compressed frames are decoded. Finally the decoded data must be reconstructed to get the video frames. Figure 10 shows the original and reconstructed video frame for frame No. 1.
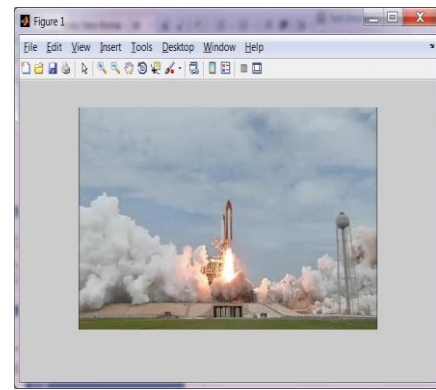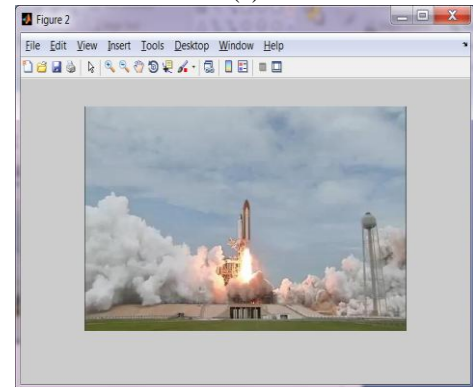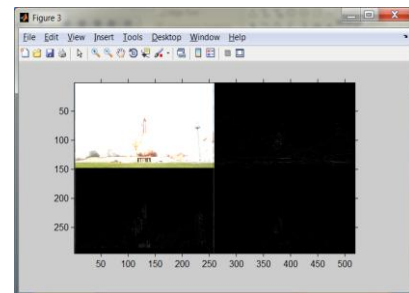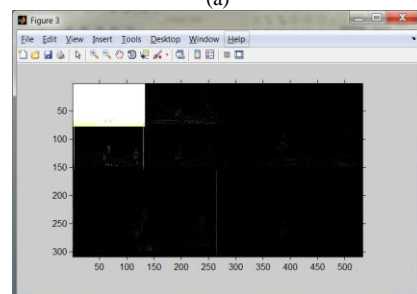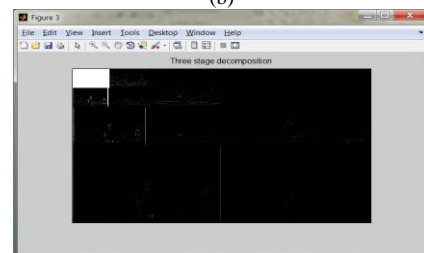

Figure 8. Frame 1 and 2 of video file

Figure 9. Wavelet decomposition for one frame (a) Level 1 (b) Level 2 (c) Level 3
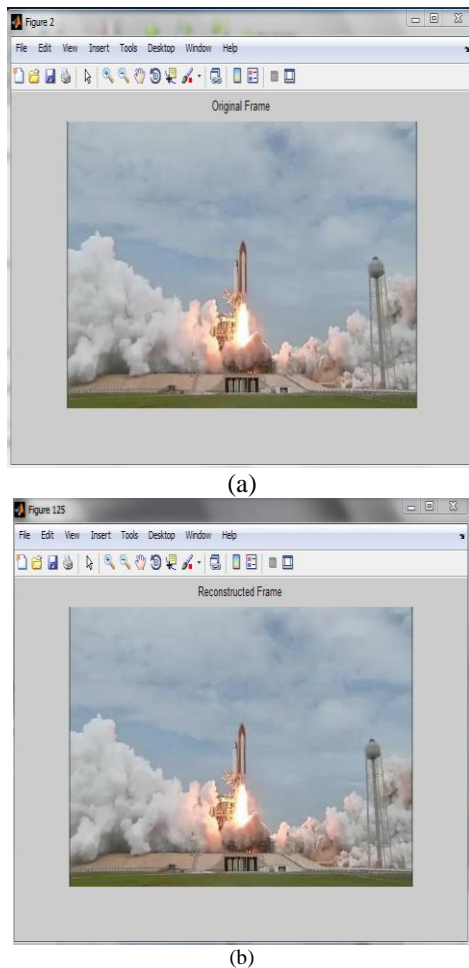
(a)

(b)

Figure 10. Comparison between Original and Reconstruction for frame 1

A number of quantitative parameters can be used to evaluate the performance of the coder, in terms of reconstructed video quality after compression scores. The compression ratio of the proposed video codec is compared by using various wavelet names. The compression ratio (CR) is defined as the ratio of the original video file size to the compressed file size.

$$\text{Compression Ratio} = \frac{\text{Original FileSize}}{\text{Compressed FileSize}}$$

Table 1. Performance test with various wavelet names

Video properties	Wavelet Name	Compression Ratio
Video format=*.avi nFrames = 121 frame_rate = 30 vidlength= 4.033 sec	Sym4	1.6765
	Sym3	1.6115
	Db4	1.6272
	Haar	1.4314

The performance parameter such as Compression Ratio (CR) is evaluated based on the algorithm. Comparisons amongst the various wavelet names such as sym4, sym3, db4 and haar are carried out on the basis of calculated performance parameter. The "sym4" wavelet achieved the best Compression Ratio.

4. CONCLUSION

This paper presents a wavelet-based video codec, its fast implementation, and its compression efficiency. For video codec, Intra-frame coding technique is implemented in this work. By using the DWT, each video frame is decomposed into 10 frequency subbands. The run length encode is used to scan the bit map row by row looking for runs of pixels of same color. Each of the resulting subbands is encoded and decoded. Finally, the inverse DWT (IDWT) reconstructs each video frame. By using the "sym4" wavelet, the compression ratio is the best.

REFERENCES

1. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform", IEEE Trans. On Image Proc., Vol. IP-1, pp. 205 220, Apr. 1992.
2. N. Treil, S. Mallat and R. Bajcsy, "Image Wavelet Decomposition and Application," GRASP Lab 207, University of Pennsylvania, Philadelphia, Technical Report MS-CIS-89-22, April 1989.
3. T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits System Video Technology, pp. 243-250, 2003.
4. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on Image Processing, vol. 1, no. 2, pp. 205-220, 1992.
5. CAPON, J. (1959). A probabilistie model for run-length coding of pictures. IRE Trans. On Information Theory, IT-5, (4), pp. 157-163.
6. APOSTOLOPOULOS, J. G. (2004). Video Compression. Streaming Media Systems Group. http://www.mit.edu/~6.344/Spring2004/video_compression_2004.pdf (3. Feb. 2006)
7. J.-R. Ohm, "Advances in scalable video coding," Proceedings of The IEEE, vol. 93, no. 1, pp. 42-56, Jan. 2005.
8. J.-R. Ohm, M. van der Schaar, and J. W. Woods, "Interframe wavelet coding – motion picture representation for universal scalability", Signal Processing: Image Communications, vol. 19, no. 9, pp. 877-908, Oct. 2004.
9. R. Klepko and D. Wang, "A real-time wavelet-based video decoder using SIMD technology," submitted to SPIE Conference on Real-Time Image Processing V, 26-31 January 2008, San Jose, California, USA.