TEMPORAL VIDEO COMPRESSION USING MODE FACTOR AND POLYNOMIAL FITTING ON WAVELET COEFFICIENTS

T. Nithyaletchumy Devi, W. K. Lim, W. N. Tan, Y. F. Tan, H. T. Teng
Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia

Y. F. Chang
Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman
46200 Petaling Jaya, Selangor, Malaysia

Keywords: Temporal video compression, Wavelet transform, Haar wavelet, Polynomial fitting.

Abstract: The core idea of this study is to build an algorithm that functions to compress video sequences. The mode value at every pixel along the temporal direction is calculated. If the frequency of the mode value satisfies a predetermined frequency, then the intensity values for entire entries at that particular pixel position will be changed to the mode value. The wavelet techniques will be applied to the pixels that do not satisfy the predetermined frequency and followed by a polynomial fitting method. For the purpose of compression, only the polynomial coefficients for pixels that do not satisfy the predetermined frequency, the mode values for pixels that satisfy the predetermined frequency and the corresponding pixel positions will be stored. To decompress, wavelet coefficients are estimated by the respective polynomials. The intensity values at the intended pixel position are obtained by inverse wavelet transform for pixels that do not satisfy the predetermined frequency. On the other hand, the stored mode values will be used to represent the intensity values throughout the time interval. This method portrays a prospect to achieve an acceptable decompressed video quality and compression ratio.

1 INTRODUCTION

On the whole, video compression denotes an act to represent the details of a video sequence by means of minimal data. Instead of transmitting all the images, a code of the image representation will be transmitted with a much smaller data size (Symes, P.D., 2001). Furthermore, data can be compressed before storage and transmission and decompressed at the receiver, besides increasing the bandwidth available (Symes, P.D., 2003). The two main classification of compression are the lossless compression and the lossy compression. These methods have two main strategies, namely, redundancy and irrelevancy respectively. In lossless compression, the concentration is on obtaining efficient ways of encoding the data. Additionally, there will be no information that is irretrievably lost in the process and it is exactly reversible. Whereas, lossy compression transforms the image to have simplified information and removes the data that we can't perceive in order to attain reduction in the file size. It is an irreversible process that permanently disposes some information. There are formats that allow compression to as little as 1% but too much compression may be dreadful as the changes becomes visible and observable which can also result to a video that can be hardly recognized (Dunn, R.D., 2002).

Wavelet has extensively inspired both image and video compression (Averbuch, A., 1996, Koornwinder, T.H., 1993). With this basis, our approach is to apply wavelet, to be precise, Haar, in the temporal direction. The quintessence of this study is to seek an algorithm to compress by extracting only the perceptible element, thus considerably reducing the data needed to be stored whilst maintaining the adequate image quality. Previously, the wavelet decomposition coefficients which were generated using the Haar wavelet (Figure 1) were applied on the pixel intensity values

Nithyaletchumy Devi T., Lim W., Tan W., Tan Y., Teng H. and Chang Y. (2009).
TEMPORAL VIDEO COMPRESSION USING MODE FACTOR AND POLYNOMIAL FITTING ON WAVELET COEFFICIENTS.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 160-165
DOI: 10.5220/0001787901600165
Copyright © SciTePress

at all pixel positions and were approximated by a polynomial of a fixed degree (T. Nithyaletchumy Devi et al., 2008). It is a hybrid method with an interest of wavelet related to a polynomial fitting.

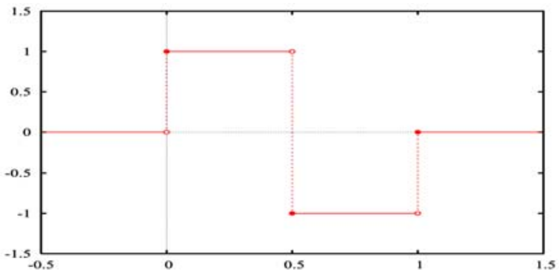

Figure 1: The Haar Wavelet $\psi(t)$.

In this study, the mode value at each pixel position along the temporal direction is calculated. Subsequently, the frequency of the mode value will be compared to a predetermined percentage. Having the acceptable value of frequency, substitutes all the intensity values at the pixel position along the temporal direction to have the mode value and stores a single mode value for the entire temporal direction. Pixel positions that did not satisfy the predetermined frequency will have wavelet applied on to obtain the wavelet decomposition coefficient using the Haar wavelet and followed by polynomial fitting of a fixed degree. Only the coefficients of the polynomials at apiece pixel along the temporal direction, the mode values for the pixels that satisfy the predetermined frequency and the corresponding pixel positions will be stored for the purpose of compression. The decompression is done by estimating the wavelet coefficients from the polynomials with the stored coefficients and retrieving the mode values at intended pixel position throughout the time interval.

Even though, there are a wide variety of popular wavelet algorithms such as the Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets which have the advantage of better resolution for smoothly changing motions but they are more expensive to calculate than the Haar wavelets (Kaplan, I., 2004). The Haar wavelet transform is theoretically trouble-free and speedy, exactly reversible and handles well over the edge effects that are a problem with other wavelet transforms.

The Haar wavelet's mother wavelet function $\psi(t)$ and its scaling function $\phi(t)$ are as described below. Without loss of generality, we shall use the same symbols for normalized wavelet and scaling functions.

$$\psi(t)=\begin{cases}1 & 0\le t<1/2,\\ -1 & 1/2\le t<1,\\ 0 & \text{otherwise}\end{cases} \quad \text{and} \quad \phi(t)=\begin{cases}1 & 0\le t<1,\\ 0 & \text{otherwise}\end{cases}$$

2 TEMPORAL COMPRESSION AND DECOMPRESSION

In order to understand ways to modify an image, it is only essential to master the way the computer stores the image. Consider a $M\times H\times T$ pixel gray scale image, where $M\times H$, the size of each frame and T, the total number of frames considered. The computer stores this image as a $M\times H$ matrix, with each elements ranging from 0 to 255. At this primary level, T is best sustained at powers of two (e.g. 2, 4, 8, 16 etc.) as to permit a straight forward distribution of data without any additional manipulation as far as wavelet is concern.

In this study, first and foremost task is to calculate the mode, denoted as q_{ij}, at each pixel location (i, j) along the temporal direction. Let fr_{ij} be the corresponding frequency of q_{ij}. For each (i, j), if $fr_{ij}\ge(p\%\times T)$, $0<p\le 100$, then the pixel location (i, j) will be stored in set S_1, else it will be in set S_2. For each pixel location (i, j) in set S_1, all intensity values along the temporal direction will be changed to q_{ij}. For each pixel location (i, j) in set S_2, Haar wavelet decomposition method will be applied. Illustrating this in a pictorial form (see Figure 2 below), the pixel positions in set S_1 are not shaded, whereas the pixel positions in set S_2 are denoted by the shaded boxes.

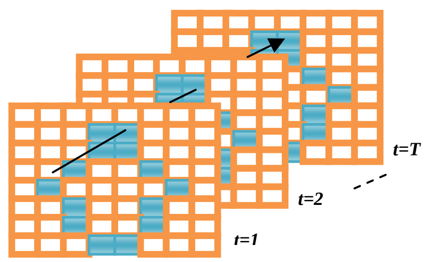

Figure 2: The pixel positions in set S_1 are not shaded, whereas the pixel positions in set S_2 are shaded.

Consider the pixels in S_2, identify the intensity values at the pixel position (i, j) at every frame as

$\{c_{ij}(t)\}$ and for each (i,j), perform Haar wavelet decomposition to the sequence $\{c_{ij}(t)\}$ for n levels, where $n=1,2,...,N$:

$$\sum_k a_{ij}(n,k)\phi\left(2^{N-n}t-k\right)+\sum_{q=N-n}^{N-1}\sum_k d_{ij}(N-q,k)\psi\left(2^q t-k\right)\in v_{N-n}+\sum_{q=N-n}^{N-1} w_q,$$

where

$$v_{N-n}=\left\{\sum_k a_{ij}(n,k)\phi\left(2^{N-n}t-k\right)\right\}$$

$$w_q=\left\{\sum_k d_{ij}(N-q,k)\psi\left(2^q t-k\right)\right\},\quad q=N-n,\dots N-1$$

are the orthogonal subspaces of the decomposition of our function space.

Two sets of coefficients, which are, the approximation coefficients $\{a_{ij}(1,t)\}$ and the detail coefficients $\{d_{ij}(1,t)\}$ will be generated as shown in Figure 3. Subsequently, the Haar wavelet decomposition is reiterated to the approximation coefficients sequence $\{a_{ij}(1,t)\}$ and obtain the sequences $\{a_{ij}(2,t)\}$ and $\{d_{ij}(2,t)\}$. This procedure is persisted for say, N number of times whereby N signifies the level of the wavelet decomposition and it is depending on our preference. This will affect the file size and quality which is the trade-off that we will have to decide on.

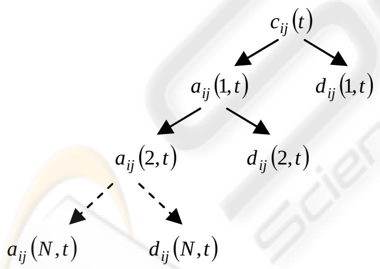

Figure 3: The wavelet decomposition tree.

By means of both the detail coefficients sequences $\{d_{ij}(1,t),\dots,d_{ij}(N,t)\}$ and the approximation coefficients sequence $\{a_{ij}(N,t)\}$, the original sequence $\{c_{ij}(t)\}$ can be reassembled. We notice that the detail coefficients include numerous values that are literally small, predominantly due to pixels comprising lesser movements along the temporal direction. Then, we calculate the cumulative values for the approximation coefficients and the absolute values of the detail coefficients to retain an increasing trend respectively. With that, we are able to fit the cumulative values of the respective coefficients using linear combinations of polynomials of degree $R-1$ as follows:

$$\sum_{r=1}^{R} b_{ij}^r t^{r-1} \qquad (1)$$

where b_{ij}^r are constants to be resolved for a preferred value of $r=1,2,\dots,R$.

For storage purpose, q_{ij} is stored for each pixel location in set S_1 and the corresponding values of b_{ij}^r for fitted polynomial on approximation and detail coefficients are stored for each pixel location in set S_2. For decompression purposes, the intensity values along temporal direction for each pixel location in set S_1 are assigned the values of q_{ij}. For each pixel location in set S_2, the cumulative values for both approximation coefficients and detail coefficients at any frame t, $t\in[1,T]$ can then be estimated from the polynomial in (1). Utilizing these estimated cumulative values, we are able to reconstruct the corresponding intensity values in a lossy manner.

3 RESULTS AND DISCUSSION

We tested the current study's method using a well-known video sequence, "Akiyo", with frame size of 176×144 pixels. This video was independently scrutinized through first 16 and first 32 frames for various *p* at different levels of wavelets and fitted on different degrees of polynomial. Outcome of each circumstance demonstrates the significance of the role played by numbers of frames, frequency of mode, level of the wavelet and degrees of polynomial. The ensuing data are then compressed using the discussed method, saved for decompression and consequently the peak signal to noise ratio (*PSNR*) (Taubman, D.S. et al., 2002) of each frame are computed using the following formula.

$$PSNR=10\log_{10}\left(\frac{255^2}{\text{Mean Square Error}}\right)$$

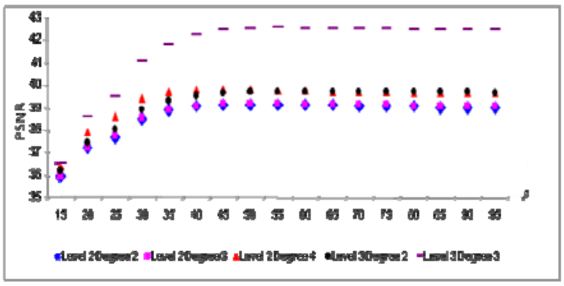

Figure 4: *PSNR* versus *p* using the first 32 frames at level 2 and 3 of wavelet decomposition and different degree of polynomial fitting.

Figure 4 shows the *PSNR* versus *p* for *p* ranging from 5 to 95, for 32 frames using Haar wavelet decomposition level 2 and 3 at degree 2, 3 and 4. Prominently, for both 16 and 32 frames, different degrees used for polynomial at level 2 wavelet decomposition does not bring significant improvement in the *PSNR*, whereas for 32 frames, different degrees used for polynomial at level 3 shows a more significant improvement in the *PSNR*. Perhaps, this is due to the fact that a degree 3 polynomial fit exactly to the 4 data obtained on the approximation coefficients.

The *p* chosen is vital as it ascertains the stringency on the minimum frequency that needed to be obtained to segregate between pixels that go through the wavelet decomposition process and the pixels that will have all intensity values surrogated with the mode value. From Figure 3, the optimum *p* lies between 45 to 55 for the tested set of parameters.

The degree chosen at each scenario is also important because as the level of wavelet and degree of polynomial increase, the number of data concerned in generating the approximation coefficients should possibly consist of at least the minimum number of data needed to evaluate the coefficients on that particular degree of polynomial. This is an essential rule of thumb for analyzing rationales. For an instance, applying a level 3 wavelet using the first 16 frames, it is not suitable to use a degree 3 polynomial for its fitting as there will only be two approximation coefficients, conversely, we can do so using the first 32 frames. Alternatively, in order to use a level of wavelet decomposition that is higher than the number of data obtained at the approximation coefficients, we may chose to apply a lower degree of polynomial to the approximation coefficients and a higher degree of polynomial to the detail coefficients. This allows us to evaluate the scenarios at a higher level but of course with a natural increase in the file size bearing in mind that higher polynomials produce more coefficients.

The minimum, maximum and mean values of *PSNR* of each scenario are tabulated in Table 1 (using the first 16 frames) and Table 2 (using the first 32 frames) in comparison with our previous study's method (T. Nithyaletchumy Devi et al., 2008) and current study's method, with *p* = 25, 50 and 75 . In the previous method, the wavelet decomposition coefficients were generated from the pixel intensity values at all pixel positions. Then, these coefficients were estimated by a polynomial of a fixed degree at every pixel positions. The current method on the other hand, applies polynomial fitting only at pixel positions whereby the predetermined frequency is not satisfied. By design, this reduces the amount of information to be saved which has a direct impinge on file size and compression ratio.

Information in the tables includes the following:

Previous : Results using the previous study's method (wavelet applied to all pixel intensity values)

Current : Result using the current study's method (conversed method)

p : Frequency of mode to be at least *p*% of the total number frames involved

Frames : Number of frames used (16 and 32 frames)Level : The wavelet decomposition level (level 2 and 3)

Degree : The polynomial degree (degree 2, 3 and 4)

File Size : Size of the information needed to be stored

Average *PSNR* : *PSNR* of each scenario

CR : Compression ratio

Table 1: Minimum, maximum and mean values of *PSNR* using the first 16 frames using *p* = 25, 50 and 75 at different levels of wavelet and degrees of polynomial.

Frames	Level	Degree	File Size (KB)				Average PSNR				CR			
			Prev.	Current			Prev.	Current			Prev.	Current		
				p=25	p=50	p=75		p=25	p=50	p=75		p=25	p=50	p=75
16	2	2	152,150	21,356	45,977	95,099	44.57	45.73	48.41	47.52	2.67	18.99	8.82	4.26
16	2	3	200,087	22,163	52,843	115,376	45.03	45.91	49.25	48.33	2.03	18.30	7.67	3.51

Table 2: Minimum, maximum and mean values of *PSNR* using the first 32 frames using p = 25, 50 and 75 at different levels of wavelet and degrees of polynomial.

Frames	Level	Degree	File Size (KB)				Average PSNR				CR			
			Prev.	Current			Prev.	Current			Prev.	Current		
				p=25	p=50	p=75		p=25	p=50	p=75		p=25	p=50	p=75
32	2	2	191,075	43,544	113,516	145,631	38.48	37.65	39.10	39.03	4.24	18.63	7.14	5.57
32	2	3	249,605	49,742	140,281	182,559	39.16	37.88	39.86	39.77	3.25	16.30	5.78	4.44
32	2	4	311,226	58,736	172,329	224,951	38.95	38.55	39.79	39.65	2.61	13.81	4.71	3.61
32	3	2	198,073	44,222	115,799	142,197	39.23	39.98	39.66	39.66	4.09	18.34	7.00	5.70
32	3	3	256,594	51,492	143,126	185,594	41.69	39.46	42.48	42.45	3.16	15.75	5.67	4.37

Generally, having not much of movements involved, engaging any polynomial would present a well-fitted line. Studying the results obtained, using our conversed method by means of 16 or 32 frames seem to have an evidently improved result in the file size compared to using previous method. Also, even though using more frames allows us to obtain a tolerable file size and an acceptable compression ratio, but the trade-off for the image quality would be less efficient and not worth the compromise as the diminution in the image quality is quite prominent. Having compared the obtained *PSNR* values with other methods used in video compression (Duanmu, C. J., 2006, Liang, J. et al., 2005, Lin, K. K., et al., 2004, Zadeh, P. B., et al., 2008), the results, as far as the *PSNR* values are concerned, they are comparable and are in a very comfortable and acceptable range.

The mode factor portrays an obvious reduction in file size as the frequency reduces. Although the deviation in the *PSNR* as the frequency changed is not to a great extent, but the file size is notably altered. It also shows that, the reduction or increment in the frequency will only result to a certain level of improvement in the *PSNR*.

Typically, the value of *PSNR* is proportional to the degree of polynomial and the level of wavelet applied. At most events, as the degree of polynomial increases at every level, the image quality improves, as far as the mean value of *PSNR* is concerned. In addition, higher level of wavelet decomposition allows enhanced analysis on the details of the motions involved. For that reason, at every degree, as the level is extended, the image quality is also constantly improved. Even though higher degree of polynomial and higher level of wavelet decomposition engages more space, an advantageous extent of improvement is preserved. Additionally, increase in the number of frames yield to decrease in the values of *PSNR*. On the whole, the proposed method emerges to evidently boast positive upshot as the qualities of the images are relatively elevated while delivering better representations of the original images.

4 CONCLUSION AND SUGGESTIONS

Figures 5(a) shows the original images from the “Akiyo” video sequences at frames 1, 5, 8, 10, 12 and 16. Figures 5(b) and 5(c) below shows selected frames (frames 1, 5, 8, 10, 12 and 16) of the decompressed images from the “Akiyo” video sequences using the first 16 frames with level 2 wavelet decompositions and degree 3 polynomial fitting on both previous and proposed method

As for the findings and analyses, the range of *PSNR* acquired using the first 16 frames results the uppermost value of *PSNR* seeing that lesser points will have lesser deviation as far as accuracy is concerned. Using 32 frames may grant a sensibly reduced file size with a reasonable compression ratio, but a massive concession on the image quality has to be acknowledged. Nonetheless, engaging 16 frames distributes the most rationale results with a balanced trade-off as far as efficiency and quality is concerned. On average, using $p = 50$ appears to have a higher *PSNR* but the compromise in file size is too massive.

Regardless of the method used, the highest value of *PSNR* is obtained when fewer frames are considered. Even so, the desired option will be the one with a good trade-off between the file size and the *PSNR* as they have a vast impact on the storage efficiency and the image quality respectively. The proposed method of polynomial fitting applied to the wavelet coefficients of the relevant pixels produced a fine outcome with anticipated level of efficiency as far as compression is concerned.

Nonetheless, there exist certain limitations to this conversed method where it is only more suitable for video sequences with minimal motions and minor changes in the background. Example of such application is the storage of surveillance camera footage or a closed-circuit television. This study is still in the ground work and has heaps of rooms for incessant advancement and enhancement. In future, resolving a suitable level of wavelet and identifying an appropriate degree of polynomial along with suitable polynomials for different pixels throughout the frames may be considered. Also, development of a model to obtain the optimal p for any domain of dataset can be deemed. Color video sequence can also be given a thought.

Figure 5(a): Images of the frames 1, 8, 12 and 16 using the first 16 frames of the Akiyo video sequences.

Figure 5(b): Images of frames 1, 8, 12 and 16 using the first 16 frames using the previous method.

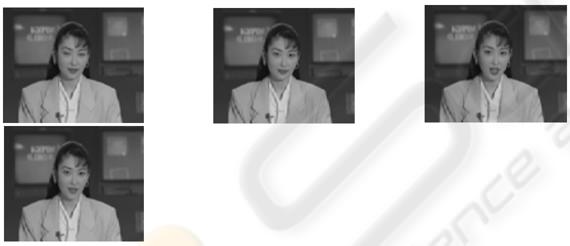

Figure 5(c): Images of frames 1, 8, 12 and 16 using the first 16 frames using the current method at p=25.

REFERENCES

Averbuch, A., Lazar, D. and Israeli, M., 1996. Image Compression Using Wavelet Transform and Multiresolution Decomposition. *IEEE Trans. on Image Processing, Vol. 5, Issue 1*, 4 –15.

T.Nithyaletchumy Devi, Lim W.K., Tan Y.F., Tan W.N., Teng, H.T. and Chang, Y.F., 2008. Video Compression Using Temporal Polynomial Fitting on Wavelet Coefficients. *16th Int. Symposium of Science and Mathematics (SKSM).*

Duanmu, C. J., 2006. Fast Scheme for the Thee-step Search Algorithm by the Utilization of Eight-bit Partial Sums. *49th IEEE Int Midwest Symposium on Circuits and Systems. In MWSCAS'06, Vol. 2,* 128-131.

Duanmu, C. J., 2006. Fast Scheme for the Four-step Search Algorithm in Video Coding. *IEEE Int Conference on Systems, Man and Cybernetics. In SMC'06, Vol. 4,* 3181-3185.

Dunn, J.R., 2002. Faster Smarter Digital Video, *Microsoft Press.*

Kaplan, I., 2004. *Applying the Haar Wavelet Transform to Time Series Information,* http://www.bearcave.com/misl/misl_tech/wavelets/haar.html

Koornwinder, T.H., 1993. Wavelets: An Elementary Treatment of Theory and Applications, *Singapore: World Scientific.*

Liang, J., Tu, C., Tran, T. D., 2005. Optimal Block Boundary Pre/Postfiltering for Wavelet-based Image and Video Compression. *IEEE Trans. on Image Processing, Vol. 14, Issue 12,* 2151-2158.

Lin, K. K., Gray, R. M., 2004. Wavelet Video Coding With Dependent Optimization. *IEEE Trans. on Circuits and Systems for Video Technology, Vol. 14, Issue 4,* 542-553.

Symes, P.D., 2001. *Video Compression Demystified,* McGraw Hill.

Symes, P.D., 2003. *Digital Video Compression*, McGraw Hill.

Taubman, D.S., Marcellin, M.W., 2002. *JPEG2000: Image Compression Fundamentals, Standards and Practice,* Kluwer Academic Publishers.

Zadeh, P.B., Buggy, T., Akbari, A.S., 2008. Statistical, DCT and Vector Quantization-based Video Codec. *IET Image Processing, Vol. 2, Issue 3,* 107-115.