2009 Ninth International Conference on Intelligent Systems Design and Applications

A 3D Lifting Based Method Augmented by Motion Compensation for Video Coding

Sedat Telceken[1], Sukru Gorgulu[2], Omer N. Gerek[2]
1 : Department of Electrical and Electronics Engineering, Anadolu University
2 : Department of Computer Engineering, Anadolu University
Iki Eylul Kampusu, 26470, Eskişehir, Turkey
phone: + (90) 2223213550 (6563), email: (stelceken, sgorgulu, ongerek)@anadolu.edu.tr

Abstract

This study introduces a spatio–temporal lifting based algorithm to be used in compression of video signals. The temporal correlation of consecutive frames causes temporal redundancies, which are subject to lifting-like motion predictive compression. Similarly, neighbouring pixels are correlated within each frame. A method that uses both correlations might be 3D lifting-based decomposition. In this study, block-based motion compensation is added to the classical 3D lifting method. Domain of motion compensation is first selected as free, and then reverse-symmetric. It is observed that reverse-symmetric motion compensation improves the performance of the prediction step in 3D lifting based coding.

1. Introduction

In this study, a video signal decomposition method that combines motion compensation and lifting based prediction techniques is examined. Following brief descriptions of these techniques, two alternative methods for the incorporation of motion compensation are proposed and their effects are discussed.

Due to the large amount of sample-wise redundancy, and deficiencies in the human visual system, video compression is generally performed in a lossy manner. The classical idea is to assign some frames within the video as "intra-type" and encode the temporal differences of other frames with respect to these intra-frames. To achieve a complete video codec, both the intra-frames and predicted frames should be quantized and entropy coded according to the desired compression ratio. For this particular work, it must be noted that a complete video coder algorithm is beyond the scope. Here, an intermediate step of 3D lifting based decomposition is taken as a starting point, and a marginal improvement is searched by incorporating block based motion compensation (MC) prior to the prediction step of the temporal lifting algorithm.

The reason of the work is motivated by the fact that both MC and temporal lifting prediction seek for minimization of the residual signal variance for predicted frames. Therefore their combined use must be justified by experimental work. Clearly, the direct application of temporal lifting does not take into account any kind of motion for the consecutive frames. Recently, for the temporal sub-band decomposition of video signals, Pesquet-Popescu and Bottreau improved the compression efficiency by using a lifting scheme with MC [1], [2], [3]. In those studies, both adaptive subband decomposition and linear subband decomposition methods were tested. In this study, a third decomposition method is preferred where the temporal decomposition is mixed with spatial information through the edge adaptation strategy of the 2D lifting method proposed in [4], [5]. In that method, the edge directions were used for the prediction direction of the lifting stage. Considering a slow motion within a video sequence, it is thought that a similar 3D gradient could also be incorporated into the prediction direction in 3D lifting, as explained in Sec.2.1.

The incorporation of motion compensation over the edge adapted temporal prediction is also tested according to two different assumptions. In the first assumption, the range and domain of the motion compensation according to the frame-to-be-predicted is selected as completely free, and the predicted frame location is determined according to these two matching blocks as explained in Sec. 2.2. In the second assumption, linear motion is assumed for three consecutive frames with a reference location of the predicted frame. Therefore, the motion for the backward frame must be exactly in the opposite direction for the forward frame. This method is

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.23

explained in Sec. 2.3. In Section 3, experimental works and results are presented.

2. Steps of Method

The 3D lifting scheme used in this study and two different MC methods included will be described in this section. Instead of focusing on the center pixels of blocks after block matching in motion compensation (as described in [1]), the entire block is dealt after block matching for directional spatio-temporal prediction.

2.1. 3D Lifting Scheme on Video

In a temporal lifting strategy, one stage of lifting starts by assigning even and odd numbered frames as predicted and intra-frames. In a complete system, the prediction error is also encoded according to a desired compression ratio. However, here, only the prediction success will be measured. Therefore, the amount of prediction error figures will be provided without actually compressing the prediction residual any further. The extreme case of total elimination of prediction residual will be provided as the decoding output. Therefore, while decoding, even-numbered frames of video signal are *only estimated* using polyphase decomposition analysis of odd-numbered frames according to the edge adapted lifting prediction as in [5].

Let $x_i[m,n]$ represent the pixel located at coordinates [m,n] of frame i, of video signal x. Let's consider a 3x3x3 cubic block of the 3D video signal so that the pixel to be estimated is located at the center of this block. The front 3x3 side of this cubic block belongs to frame (2i – 1), and the rear 3x3 side belongs to frame (2i + 1). Hence, between those sides, there exists an intermediate 3x3 layer which includes the center pixel to be estimated. These are illustrated in Figures 1, 2, and 3. A 3D illustration of these three 3x3 layers as one 3x3x3 block is illustrated in Fig. 4.

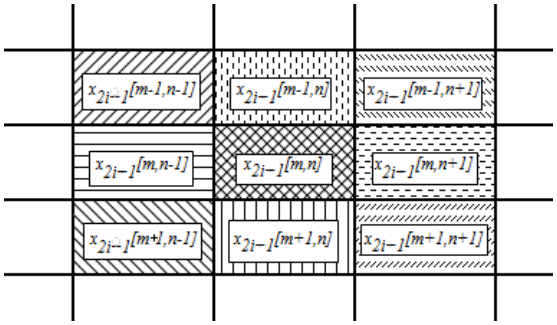

Figure 1: 3x3 front block side in frame (2*i* - 1)

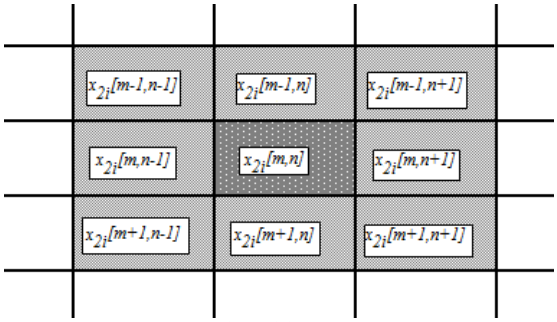

Figure 2: 3x3 intermediate block layer in frame (2*i*)

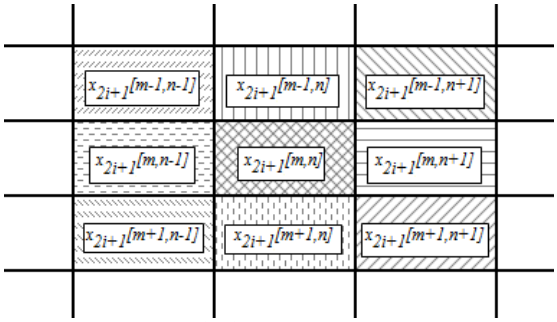

Figure 3: 3x3 rear block side in frame (2*i* + 1)

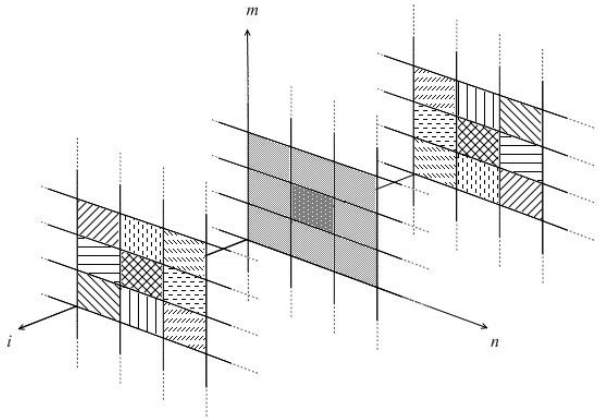

Figure 4: A 3-D view of block layers as one 3x3x3 block

The value of $x_{2i}[m,n]$ is estimated by the previous layer's and next layer's pixel couples which are centred around the $x_{2i}[m,n]$ pixel. We define 13 different gradient approximations around $x_{2i}[m,n]$. Four of these which are within the same frame, can be omitted. Remaining valid gradient approximations are given below:

$$\Delta_1 = |x_{2i-1}[m-1,n-1] - x_{2i+1}[m+1,n+1]|$$
$$\Delta_2 = |x_{2i-1}[m-1,n] - x_{2i+1}[m+1,n]|$$
$$\Delta_3 = |x_{2i-1}[m-1,n+1] - x_{2i+1}[m+1,n-1]|$$
$$\Delta_4 = |x_{2i-1}[m,n-1] - x_{2i+1}[m,n+1]|$$
$$\Delta_5 = |x_{2i-1}[m,n] - x_{2i+1}[m,n]|$$
$$\Delta_6 = |x_{2i-1}[m,n+1] - x_{2i+1}[m,n-1]|$$

$$\Delta_7 = |x_{2i-1}[m+1, n-1] - x_{2i+1}[m-1, n+1]|$$
$$\Delta_8 = |x_{2i-1}[m+1, n] - x_{2i+1}[m-1, n]|$$
$$\Delta_9 = |x_{2i-1}[m+1, n+1] - x_{2i+1}[m-1, n-1]|$$

Nine different $x_{2i}[m,n]$ estimation values can be obtained from these expressions, as listed below:

$$\hat{x}^1_{2i}[m,n] = (x_{2i-1}[m-1, n-1] + x_{2i+1}[m+1, n+1])/2$$
$$\hat{x}^2_{2i}[m,n] = (x_{2i-1}[m-1, n] + x_{2i+1}[m+1, n])/2$$
$$\hat{x}^3_{2i}[m,n] = (x_{2i-1}[m-1, n+1] + x_{2i+1}[m+1, n-1])/2$$
$$\hat{x}^4_{2i}[m,n] = (x_{2i-1}[m, n-1] + x_{2i+1}[m, n+1])/2$$
$$\hat{x}^5_{2i}[m,n] = (x_{2i-1}[m, n] + x_{2i+1}[m, n])/2$$
$$\hat{x}^6_{2i}[m,n] = (x_{2i-1}[m, n+1] + x_{2i+1}[m, n-1])/2$$
$$\hat{x}^7_{2i}[m,n] = (x_{2i-1}[m+1, n-1] + x_{2i+1}[m-1, n+1])/2$$
$$\hat{x}^8_{2i}[m,n] = (x_{2i-1}[m+1, n] + x_{2i+1}[m-1, n])/2$$
$$\hat{x}^9_{2i}[m,n] = (x_{2i-1}[m+1, n+1] + x_{2i+1}[m-1, n-1])/2$$

The value $\hat{x}^5_{2i}[m,n] = (x_{2i-1}[m,n] + x_{2i+1}[m,n])/2$ is the classical estimation which is obtained by the lifting prediction on the 3×3×3 cubic block. In case of no motion gradient, this value would give us minimum difference between estimated and actual pixel value.
However, if any motion gradient occurs near the centered pixel on the video sequence and if this causes a 3D edge within the cube boundaries, then another cross pixel prediction value may be closer to center pixel's actual value. Hence, pixel couples in the corresponding direction will give better results. This selection method introduces a direction adaptive filter, just like the 2-D version described in [4]. This adaptive filter still uses pixel couples for estimation and finds the direction that holds the pixel couples with minimum difference. Possible gradient approximation count is nine as defined above. Hence, calculation of the minimum one is not a computationally complex operation.

2.2. Application of MC to 3D Lifting Scheme Using Loose Central Pixel Location

The edge adaptive method described in Sec. 2.1 can govern subtle motions of 1 pixel per frame. On the other hand, movement of an edge formation in consecutive frames may exceed boundaries of 3x3 block area. In fact, for many typical video sequences, the amount of linear motion along duration of 3 frames exceeds the 3×3 block boundaries utilized for the edge adaptive lifting. In that case, quality of polyphase decomposition analysis is reduced. Therefore position of the 3×3 block on the $(2i\text{-}1)^{st}$ frame must be searched and properly aligned on the $(2i\text{+}1)^{st}$ frame. Then the pixel to be estimated must also be aligned properly on the $(2i)^{th}$ frame. Further application of polyphase decomposition on the 3×3×3 rectangular block would reduce prediction error.

Let us define an operator, $W_{i,j}$ for projection of motion compensated (MC) lifting scheme on video. $W_{i,j}$ operator connects *i-frame* to *j-frame* according to the selected MC scheme [6].

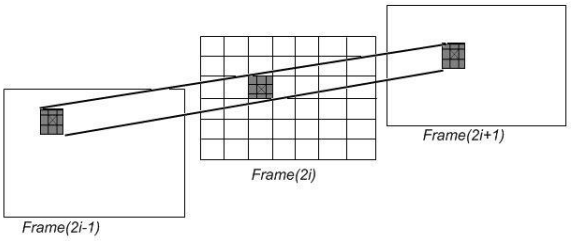

Figure 5: Locating estimated pixel after block matching

The prediction step of the temporal lifting can be described as follows:

$$\mathbf{h}[\mathbf{m},\mathbf{n}] = \frac{\mathbf{1}}{\mathbf{2}}\left(\mathbf{x_{2i+1}}[\mathbf{m},\mathbf{n}] - \mathbf{W_{2i-1,2i+1}}(\mathbf{X_{2i-1}})[\mathbf{m},\mathbf{n}]\right)$$
$$\mathbf{l}[\mathbf{m},\mathbf{n}] = \mathbf{x_{2i-1}}[\mathbf{m},\mathbf{n}] - \mathbf{W_{2i+1,2i-1}}(\mathbf{h})[\mathbf{m},\mathbf{n}]$$

where $x_{2i-1}[m,n]$ and $x_{2i+1}[m,n]$ represent two points which are settled on pixel coordinates [m,n] in two consecutive video frames. (Estimated $(2i)$st frame is out of calculation so that $(2i\text{-}1)$ and $(2i\text{+}1)$ are consecutive frames)
The $W_{i,j}$ operator is selected as the classical block matching, due to its simplicity. First, the position of 3x3 block on the frame $(2i\text{+}1)$ is marked on frame $(2i\text{-}1)$. Then the vicinity of the marked block is searched and the block that gives the closest match (in terms of squared difference) is selected. Finally, this selected block and the block from frame $(2i\text{+}1)$ are used to compute the center coordinates of the block in frame $(2i)$. This computation is based on construction of a straight line between frames $(2i\text{-}1)$ and $(2i\text{+}1)$ and finding the middle point of this line in 3D, as illustrated in Fig. 5.
When each 3x3 block from frame $(2i\text{+}1)$ to frame $(2i\text{-}1)$ is motion compensated, uncalculated pixel values can be left on frame $(2i)$. These pixel values are calculated without any MC directly using linear prediction. Thus pixels are filled by classical lifting interpolation [7], [8], [9], [10].

2.3. Application of MC to 3D Lifting Scheme Using Fixed Central Pixel Location

In the method described in previous section, internal 3x3 slide of 3x3x3 cubic block is located by matching of front side in frame (2i-1) to the rear side in frame (2i+1), and the value of central pixel ($x_{2i}[m,n]$) of this internal layer is estimated as illustrated in Fig. 6. Due to the complications that arise when a central coordinate is never rendered by this method, an alternative method is developed which makes sure that each and every pixel of the prediction frame is processed. In this new method, the location of internal layer is kept as reference point and during block matching search, matched blocks are moved in symmetrically opposite directions. During matching operation, sum of squared differences within a block is used as the matching metric. The symmetric block-wise search operation is performed for each and every pixel of the prediction frame, $(2i)$.

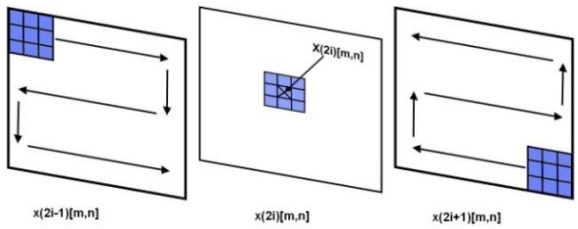

Figure 6: Locating source blocks in block-matching

Several search block sizes (3x3, 5x5, 7x7,...) were experimentally tested to achieve an empirically good performance. Results of different block sizes are also given in this work.

3. Experimental Work and Results

Experiments are applied on gray level (Y channel only) videos with CIF(352x288) size. Typical videos used in experiments are ***bus.yuv***, ***coastguard.yuv***, and ***container.yuv***. Although the methods are applied to longer sequences, in order to indicate the results in a compact form, prediction errors corresponding to the first 5 frames of each video are given here. The results are listed in Table 1.

In Table 1, mean squared errors (MSE) of predicted frames reconstructed using methods described in Sections 2.2 and 2.3 are compared. Implementations are temporarily restricted to only Y channels of videos. Hence comparison results show performances of methods on intensity levels. Only one level of lifting decomposition is made, therefore the prediction error results correspond to even-indexed frames of the temporal polyphase decomposition.

Table 1 – MSE percent rates of Loose Central Pixel Location (2.2) and Fixed Central Pixel Location (2.3) methods

	bus		coastguard		Container	
Frame No	Method 2.2	Method 2.3	Method 2.2	Method 2.3	Method 2.2	Method 2.3
1	---	---	---	---	---	---
2	**1,36**	**0,016**	**0,32**	**0,019**	**0,32**	**0,0003**
3	---	---	---	---	---	---
4	**1,43**	**0,017**	**0,3**	**0,019**	**0,32**	**0,0002**
5	---	---	---	---	---	---

Obvious performance improvements in terms of reduced MSE values are achieved by the method 2.3 (as compared to method 2.2). The improvement is also visually apparent when the two prediction frames are compared in Fig. 7.

Figure 7 – *Left : Frame reconstructed using method* 2.2 *– Right: Frame reconstructed using method* 2.3

The results in Table I and Fig. 7 were obtained with a motion compensation block size of 3x3. In addition to these results, detection of movements on video frames can be tried over various block sizes and their performances can be compared. As a preliminary comparison, method 2.3 is run on bus.yuv video signal with three different block sizes (3x3, 5x5, 7x7) and also three different search pixel sizes (2, 4, 6). Search pixel size is defined as the distance from the center pixel of block. MSE values of reconstructed even frames indicate that 3x3 block size and search size 4 pixel gave better results for this particular video sequence.

The search pixel size could also be adjusted adaptively with possible increase in performance. However, that would increase the complexity of the algorithm and

complicate the idea presented herein. Therefore, the adaptation idea is left out of the scope of this paper.

Table 2 – MSE percent rates of various block sizes in block matching method.($x10^{-2}$)

	bus.yuv								
	Block Size = 3x3			Block Size = 5x5			Block Size = 7x7		
Search Size\ Frame No	2	4	6	2	4	6	2	4	6
1	---	---	---	---	---	---	---	---	---
2	**1,60**	**1,60**	**1,66**	**1,60**	**1,64**	**1,67**	**1,62**	**1,63**	**1,67**
3	---	---	---	---	---	---	---	---	---
4	**1,71**	**1,69**	**1,71**	**1,74**	**1,74**	**1,78**	**1,75**	**1,76**	**1,81**
5	---	---	---	---	---	---	---	---	---

As a conclusion, experimental results indicate that the usage of method 2.3 gives a better performance of encoding and decoding as compared to the alternative motion compensation method which inherently misses a fraction of pixels in the prediction frame as "empty", yielding a necessity to make temporal linear interpolation. It is also observed that the 3D edge gradient method marginally outperforms the direct temporal prediction in the lifting stage. This is a natural result due to the fact that, when there is no temporal gradient, the proposed method already includes the direct temporal prediction as its subset. Nevertheless, without the incorporation of motion compensation, neither classical 3D temporal lifting, nor the temporal edge gradient based method performs well for video sequences with apparent motion vectors of length larger than 1 pixel / frame. Further and thorough analysis is necessary for the verification of the empirical selection of parameters such as motion compensation block size, motion search window size, 3D gradient search cube size, etc.

4. References

[1] B. Pesquet-Popescu, V. Bottreau, "Three-Dimensional Lifting Schemes for Motion Compensated Video Compression", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, pp. 1793-1796, Salt Lake City, UT, 2001.

[2] G. Pau, C. Tillier, B. Pesquet-Popescu, "Optimization of the Predict Operator in Lifting-Based Motion Compensated Temporal Filtering", Visual Communications and Image Processing, vol. 5308 of Proceedings of SPIE, pp. 712–720, San Jose, Calif, USA, January 2004.

[3] G. Pau, C. Tillier, B. Pesquet-Popescu, H. Heijmans, "Motion Compensation and Scalability in Lifting-Based Video Coding," Signal Processing: Image Communication, Vol. 19, Issue 7, pp. 577-600, August 2004.

[4] O.N. Gerek, A.E. Cetin, "Adaptive polyphase subband decomposition structures for image compression," Proceedings of the IEEE Transactions on Image Processing, Vol. 9, No.10, pp. 1649-1660, October 2006.

[5] O.N. Gerek, A.E. Cetin, "A 2-D Orientation-Adaptive Prediction Filter in Lifting Structures For Image Coding," Proceedings of the IEEE Transactions on Image Processing, Vol. 15, No.1, pp. 106-108, January 2006.

[6] A. Secker, D. Taubman, "Motion-Compansated Highly Scalable Video Compression Using An Adaptive 3D Wavelet Transform Based On Lifting," Proceedings of the Image Processing, 2001. Proceedings. 2001 International Conference on Vol. 2, pp.1029 – 1032, 7-10 October 2001.

[7] L. Luo, F. Wu, S. Li, Z. Zhuang, "Advanced Lifting-Based Motion-Threading (MTh) Technique for the 3D Wavelet Video Coding," Visual Communications and Image Processing. Conference, Lugano, ITALY (08/07/2003) 1988, vol. 5150 (3), pp. 707-718.

[8] A. Wang, Z. Xiong, P.A. Chou, S. Methora, "Three Dimensional Wavelet Coding of Video With Global Motion Compensation," Proceedings. DCC'99, pp. 404-413, 1999.

[9] J. –R. Ohm, "Three Diemnsional Subband Coding with Motion Compensation," IEEE Transactions on Image Processing, Vol. 3, No.5, pp. 559-571, September 1994.

[10] L. Luo, J. Li, S. Li, Z. Zhuang, Y.-Q. Zhang, "Motion compensated lifting wavelet and its application in video coding," IEEE Intl. Conf. on Multimedia and Expo(ICME2001), Tokyo, Japan, August 2001.