LEAST-SQUARES BASED SWITCHED ADAPTIVE PREDICTORS FOR LOSSLESS VIDEO CODING

Anil Kumar Tiwari and R. V. Raja Kumar

The LNM Institute of Information Technology, Jaipur and Indian Institute of Technology, Kharagpur

ABSTRACT

This paper presents a context-based predictive coding method for lossless compression of video. For this method, we propose a model to estimate level of activity in the prediction context of a pixel. This is measured in terms of slope and the same is optimally classified to results in a small number of slope bins. After finding the slope bins, we propose a LS based method to find switched predictors to be associated with the various bins. The set of the predictors are found on a frame-by-frame basis and when it is incorporated in CALIC frame work, the proposed method results in, on an average, a better compression performance than that is obtained using recently published methods - LOPT and M-CALIC. The proposed codec has higher coding complexity but much lower decoding complexity, which is necessary for real-time video decoding. The proposed method of coding, however, has much lower complexity as compared to the LOPT method, which has same order of high coding and decoding complexity.

***Index Terms*—** GAP, Slope Bins, Gradient, LS-based predictor.

1. INTRODUCTION

Efficient lossless image compression methods are in advanced stage whereas such methods for video are still growing. The state-of-art lossless image compression methods such as CALIC and JPEG-LS work on the context-based prediction method. In these methods, prediction is made using switched predictors and the switching principal is based on a measure of the characteristics of the neighbourhood pixels [1].

Recently, there has been an increased demand for lossless compression of video, such as medical imaging applications. Usually, medical images generate highly correlated sequences such as Computerised Axial Tomography (CAT) and Magnetic Resonance Images (MRI). Since single such image sequence is of very big in size (>100MB), compression is highly desirable.

There are a few methods such as [2] - [5] for lossless video coding. Method described in [2] finds difference frame, using present frame and motion compensated previous frame, and then JPEG2000 (in lossless mode) is applied on the difference frame. A context-based predictive coding method is reported in [3]. This method is based on a pixel-based coding mode selection - intraframe or interframe. In LOPT method [4] of video compression, decision of estimation of the LS based predictor, for $x(n)$, depends on the amplitude of the prediction error at the previous pixel location, $(n-1)$. If the amplitude is greater than a pre-set threshold value, the predictor is estimated for $x(n)$. Otherwise, the same previous predictor used at $(n-1)$ is used to predict the current pixel also. Pixels for which predictor is estimated are generally referred to as *edged* pixels. Motion-CALIC (M-CALIC) [5] is one of the recently reported lossless video coding method which uses previous two-frames for prediction. This method is based on block based feedforward type of prediction and the predictors are communicated to the decoder.

Low computational complexity is one of the most important requirements for video decoder as the same has to operate in real time. On the other hand the decoder based on the LOPT principal is required to estimate all those LS based predictors which were estimated (at the *edged* pixels) at the encoder end. Due to this requirement, decoder is computationally very complex and the same may fail to operate in real time, particularly if the video frames have large number of *edged* pixels.

In view of the low computational complexity requirement of video decoder, we propose to use a switched adaptive method of prediction. The proposed method classifies pixels of a given frame into a small number of classes, based on the activity level in their past neighborhood pixels. This is done using a set of pre-determined (obtained off-line) classification threshold values, which are robust as the same are obtained using a large number of video sequences. After the classification, a feedforward type of LS based predictor is estimated for pixels belonging to each of the classes. These predictors form a set of the switched predictors and the same are used in a feedback mode (switched predictors work in this manner). The set of the predictors are estimated, on a frame-by-frame basis, on the encoder side and the same are sent to the decoder. Therefore the predictor estimation cost is avoided on the decoder side. This ensures that the decoder is able to work in real time. However, this may be noted that M-CALIC estimates predictors on a block-by-block basis without employing any classification scheme for the pixels.

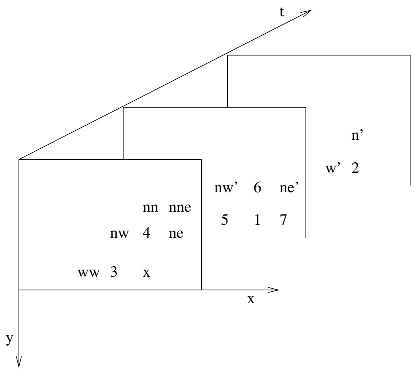

Fig. 1. A predictor structure that uses pixels at various numerals in the current and motion compensated previous two frames.

This paper is organised as follows. Section 2 briefly discusses about the LS based predictor estimation mechanism for video. Section 3 discusses about the proposed method of switched adaptive prediction and Section 4 discusses about the proposed method of finding switched predictors. To show the effectiveness of the proposed method, some simulation results are given in Section 5 and concluding remarks are made in Section 6.

2. THE LS BASED PREDICTOR

Pixels can be predicted using a linear combination of their past neighbourhood pixels in the current frame and *near* pixels in the motion compensated previous frames. After the motion compensated previous frames are obtained and predictor structure, such as the one proposed in Fig. 1, is decided, prediction coefficients for a pixel can be estimated by minimising the prediction error energy in a small training window (preferably half-plane training window above the current pixel). In the proposed structure, we are considering pixels in the two previous frames, besides pixels in the current frame, because of expectation that good amount of redundancy can be removed when prediction is made using pixels at various numerals shown in the Fig 1. Training window consists of a small number of past pixels in the current frame. A predictor that is estimated using this criterion of energy minimisation is known as the Least Squares (LS) based predictor.

Expression for the predicted value of a pixel, x, using a predictor of order p, can be given as shown below.

$$\hat{x} = lround \sum_{k=1}^{p} a(k)x(k) \tag{1}$$

where $a(k)'s$ are the prediction coefficients and $x(k)'s$ are the intensity value of the pixels located at various spatial locations, k, shown in Fig. 1. The LS based predictor can be found by solving the following *normal* equations.

$$\begin{bmatrix} R(1,1) & . & R(1,p) \\ . & . & . \\ R(p,1) & . & R(p,p) \end{bmatrix} \begin{bmatrix} a(1) \\ . \\ a(p) \end{bmatrix} = \begin{bmatrix} R(0,1) \\ . \\ R(0,p) \end{bmatrix} \tag{2}$$

where $R(m,n)$ are the elements of the covariance matrix, shown on the left side of (2). These elements can be computed as described below.

$$R(m,n) = \sum_{m,n \in W} x(m)x(n).$$

Here, W is the training window.
Once the predictors are obtained, prediction can be made using (1) and the corresponding prediction error can be entropy coded.

Use of the LS based predictor, estimated for each and every pixel in a video sequence, is expected to result in a very good compression ratio. But decoder will have to repeat the same method of predictor estimation at each and every pixel. This makes decoder very complex and due to this reason the decoder may fail to operate in real time. In view of the same we propose switched adaptive prediction method for lossless video compression.

3. THE PROPOSED SWITCHED ADAPTIVE PREDICTION METHOD

In this method, the unknown pixels are predicted using some useful characteristic features of their past neighbourhood pixels. One such feature is the slope (how the intensity values are changing). The unknown pixel is expected to fall in the same slope rule and therefore it is a very useful information for prediction.

Usually, in this prediction scheme, the slope is classified into a small number of slope bins because slope space is usually large [1]. After the classification, a different predictor is associated with each of the bins. Once slope is classified and predictors are associated with the various bins, prediction scheme for unknown pixels can be given as below.

1. Find the slope bin to which the unknown pixel belongs.
2. Use the associated predictor with the bin, found in step 1, to predict the pixel.

Therefore, slope estimation, its classification and the associated predictors are important requirements in this method of prediction. These issues are discussed in the following subsection.

3.1. Estimation of the Slope

Slope information about an unknown pixel can be found by measuring the variations in the intensity value of the pixels in

Table 1. Slope Bins found using frames of a large number of video sequences

Input s	Bin
$s > 28$	1
$s < -28$	2
$12 < s \leq 28$	3
$5 < s \leq 12$	4
$-28 \leq s < -12$	5
$-12 \leq s < -5$	6
$-5 \leq s \leq 5$	7

the current frame and in the past frames. We propose to use following models to measure such variation in x, y and in t directions (ref. Fig 1).

$$\begin{aligned} d_x &= |x(3) - ww| + |x(4) - nw| + |ne - x(4)| + \\ &\quad |x(1) - x(5)| + |x(6) - nw'| + |x(6) - ne'| \quad (3) \\ d_y &= |x(3) - nw| + |x(4) - nn| + |ne - nne| + \\ &\quad |x(5) - nw'| + |x(1) - x(6)| + |x(7) - ne'| \quad (4) \\ d_t &= |x(3) - x(5)| + |x(4) - x(6)| + |nw - nw'| + \\ &\quad |ne - ne'| + |x(1) - x(2)| \quad (5) \end{aligned}$$

Because the expression for d_x and d_y has six absolute summation terms whereas d_t has only five such terms, the normalised slope can be expressed as

$$s = \frac{d_x}{6} - \frac{d_y}{6} - \frac{d_t}{5} \quad (6)$$

After finding slope expression as given in (6), same can be classified as described below.

3.2. Slope Classification

As mentioned before, the slope space to which pixels belong is large. Therefore a large number of predictors associated with each slope values are to be be estimated at the encoder end and the same are to be sent, as overhead information, to the decoder. Because of these requirements, the estimation cost of the predictors and the associated overhead information is high. Also, it will be a time consuming process in searching a particular predictor while coding and decoding of video frames. Due to these problems, we optimally classified the slope and obtained the classifying bin boundaries as given in Table 1. We have also experimented by increasing the classification levels and found that number of bins more than seven do not yield any significant improvement in the performance. Details of the classification scheme can be found in [1].

4. LEAST-SQUARES BASED METHOD OF FINDING SWITCHED PREDICTORS

Once the bin boundaries are known, predictors to be associated with each of the bins can be found as described below.

- Pixels of the current frame can be classified according to the slope bins given in Table 1.
- Find LS based optimal predictor (feedforward type) for all the pixels belonging to a bin i. This predictor can be found using (2). While using (2), covariance matrix elements, $R(m, n)$, should be computed using all such pixels. Let this predictor be denoted by p_i.
- Similarly, find the LS based predictors for other bins. This results in the required set of switched predictors.

4.1. Computational Complexity

The proposed method and LOPT requires following computations in their prediction schemes.

1. The proposed method
   (a) The method requires computations associated with estimation of the covariance matrix. The computations depends on the order the predictor and the number of the pixels used for its estimation.
   (b) As every pixel in a given frame belongs to an unique slope bin, the sum of the covariance matrix estimation costs for the pixels belonging to the various bins is the same as the cost for the covariance matrix estimation for the whole frame.
   (c) As there are only seven proposed slope bins, given in Table 1, the same number of the matrix inversions are needed to estimate the prdictors.
   (d) The prediction coefficients obtained in c) can be quantized and stored as header information with the frame and same will be sent to the decoder.
2. The LOPT method
   (a) This method requires covariance matrix estimation for all the edged pixels. The covariance matrix is estimated, for all such pixels, using pixels in a training window. If there are $k\%$ of the edged pixels in a frame of size $W \times H$, there will be a total of $N = (k \times W \times H)/100$ number of edged pixels in it. The estimation of the covariance matrix, for all such pixels, uses a total of $N \times N_W$ number of pixels, where N_W is the number of pixels in the training window. If there are only 5% edged pixels and for $N_W = 70$ [4] and frame of size 720×576 involves more than 14

Lacks (0.05×720×576×70) of pixels for the estimation of the covariance matrix. On the other hand our method uses only 720×576 number of pixels for the same purpose.

(b) This method requires N number of matrix inversions. For $k = 5$ and frame of the size 720×576 requires more than 20 thousand matrix inversions. On the other hand our method requires only 7 number of the matrix inversions as there are only seven number of the proposed slope bins.

(c) Decoder based on LOPT method requires all these computations, which makes the decoder very complex. In many applications, however, decoder simplicity is highly desirable.

5. SIMULATION STUDIES

The proposed method of video coding is simulated using many video sequences and the results are compared against LOPT and Motion-CALIC (M-CALIC). The first four CCIR standard video sequences (dimensions 720 × 576) are the same as used in [4] while last six CIR sized sequences are the same that are used [5]. While coding a frame, motion compensated previous two frames are found and then a set of switched predictors, with the structure shown in Fig. 1, are estimated. The switched predictors are used in CALIC frame work to compress the frame. This is done with each of the frames of the video sequences. Motion vectors corresponding to the two previous frames are found using a block size of 16 × 16 and search range of ± 8 pixels [4],[5]. The motion vector field was compressed using JPEG-LS method and this overhead is approximately 0.04 bpp. As there are seven number of switched predictors of order seven each and we quantised the coefficients using 8 bits, the overhead corresponding to the set of predictors is 392 number of bits (8 × 7 × 7) per frame. These overheads are included in the results given in Table 2. From these results, it can clearly be seen that the average performance of the proposed method is significantly better than the other two methods.

6. CONCLUSIONS

This paper presents two pieces of work, first - a model is proposed to measure the level of activity in the neighbourhood pixels and then the same is classified into a small number of bins. The second work corresponds to LS based method to find switched predictors to be associated with each of the bins. Application of the LS based method of finding switched predictors for each frame results in, on an average, better performance than that are obtained by using LOPT and M-CALIC methods. The proposed prediction method is significantly faster than that of the prediction method used in LOPT. The switched predictors obtained at the encoder end will be sent as over head information to the decoder. Because of the availability of the predictors to the decoder, it works very efficiently as compared to the decoder based on the LOPT method which needs to estimate the LS based predictors. Estimation of LS based predictor at decoder may make it practically unviable for it to work in the real time, particularly if the video sequence has large number of edged pixels.

Table 2. Compression Performance of the Proposed Method. Avg (com) corresponds to the average bit rate for the last six sequences for which results with LOPT and M-CALIC are available

Video Sequence	LOPT-3D	M-CALIC	Proposed
Mobile	4.17	-	4.06
Renata	4.32	-	4.41
Flowers	4.1	-	4.03
Tennis	4.33	-	4.26
Salesman	4.32	3.33	3.58
Container	3.51	3.16	3.13
Tempete	4.53	4.45	4.08
Kitchgrass	4.22	4.11	3.95
Sean	3.40	2.94	2.83
Silent	3.55	3.13	3.04
Avg.	4.05		3.74
Avg (com)	3.92	3.52	3.44

7. REFERENCES

[1] Anil Kumar Tiwari and R. V. Raja Kumar, "Optimal slope bin classification in gradient adjusted predictor for lossless compression of medical images," in *Proceedings of the IEEE ICIP*, Genova, Italy, 11 - 14 September 2005.

[2] S. Imaizumi, A. Takagi and H. Kiya, "Lossless inter-frame video coding using extended JPEG2000 http://www.kmutt.ac.th/itc2002/cd/pdf/19_07_45/fa1_pj/7.pdf," .

[3] K. H. Yang and A. F. Farid, "A context-based predictive coder for lossless and near-lossless compression of video," in *Proceedings of the IEEE International Conference on Image Processing*, September 2000, vol. I, pp. 144–147.

[4] D. Brunello, G. Calvagno, G. A. Mian and R. Rinaldo, "Lossless compression of video using temporal information," *IEEE Transactions on Image Processing*, vol. 12, pp. 132 – 139, FEBRUARY 2003.

[5] E. S. G. Carotti, J. C. De Martin, "Motion-compensated lossless video coding in the CALIC framework, http://multimedia.polito.it/files/carotti_isspit2005.pdf," 2005.