CODING THE DISPLACED FRAME DIFFERENCE FOR VIDEO COMPRESSION

Krishna Ratakonda, Seung Chul Yoon and Narendra Ahuja
Department of Electrical and Computer Engineering
Beckman Institute, University of Illinois, Urbana, IL 61801.
krishna@stereo.ai.uiuc.edu

Abstract

Popular techniques employed to code the Displaced Frame Difference (DFD) treat it no differently from an ordinary image for coding purposes. Since the DFD is generated by the process of motion compensation, such methods do not fully exploit the underlying redundancies. This paper is aimed at proposing a DFD coding method which exploits such redundancies while incurring negligible information overhead. The key idea is to predict locations of high DFD concentration which occupy small portions of the image and use this predicted information (which is also available to the decoder with out additional information transmission) to improve the quality of the decoded image. Two key features of the proposed approach are its compatibility with any transform based DFD coding scheme and negligible information overhead. Tests with a fully functional video coder show the efficacy of the proposed approach.

1 Introduction

Video compression methods usually employ a two step procedure to compress a video sequence. The first step involves prediction of the frame to be coded by using a motion compensation algorithm on the previous frame(s) which has been already coded (and is therefore available to the decoder). After subtracting the motion predicted frame from the frame to be coded, we are left with a residual image which is termed the Displaced Frame Difference (DFD). DFD is usually coded using a transform based technique [6]. The number of bits to be expended in this process is varied to meet the bitrate. Other techniques like Vector Quantization [3] which are normally used in still image compression may also be employed here but are not as popular. In this paper we assume that the DFD is coded using a discrete linear transform based technique (for example employing DCT or Wavelets). We will propose a novel scheme to improve the PSNR of the decoded frame while incurring negligible overhead. The scheme, as implemented in this paper, requires no major modification of an existing video coding setup. It is thus very attractive from a practical stand point.

2 Discussion

The key point which aids the proposed scheme is that we are compressing DFD generated by a *motion compensated video stream* and not still images (as generally assumed by most DFD coding schemes). Thus we expect that the residual image error to be concentrated in certain areas of the image. These areas can be predicted, given that we know the motion compensation scheme. For example, in a block based motion compensation scheme most of the error image energy is concentrated at either the block edges or at the places where the image edges were located in the previous frame. We conducted experiments on video sequences which suggested that 80-90% of the DFD energy is concentrated in *predictable locations* in the frame to be coded.

The DFD coding scheme proposed in this paper makes use of the above key observation to improve the performance of any transform based DFD coder. We first generate an image mask predicting the location of DFD energy concentration.

Now we assume that the DFD exists only in the predicted locations and is zero at other locations. So we are free to vary the values of the pixels at such locations, where DFD is predicted to be zero, to obtain transform coefficients which lead to a better quality of the coded image. Since the decoder can also generate the same mask, decoding is possible. It may be noted that we *do not need* any further information to be trasmitted in order to implement the proposed scheme.

The question to be answered is: how should the "free pixels" be chosen? For DCT or Wavelets, the free pixels need to be chosen so that most of the energy is concentrated in the low frequencies. In other words we would like to use the freedom given in the choice of the free pixels to maximally pack the energy in the low frequencies of the transform domain. This problem has been addressed recently by us in a general context [5]. We had proposed an iterative algorithm which leads to an optimal choice of the coefficients, which we will use in this paper.

We will be applying the algorithm in the context of a segmentation based video coding scheme [7]. The motion compensation scheme will be based on a recently developed multi-scale segmentation algorithm [1]. Comparative results for the DFD coding part employing a DCT based technique with and without the proposed addition are provided. The comparative results are obtained by employing a *fully functional video coding scheme* and thus reflect actual gains that may be expected in a practical implementation. *Section 3* is an outline of the proposed method in a general context as applicable to any transform based DFD coding scheme. *Section 4* provides a practical, implementation level example.

3 Proposed DFD Coding Scheme

The first step in implementing the proposed scheme would be to find the prediction mask locating the positions in the error image where the energy is concentrated. This depends on the motion compensation scheme employed. As already observed, for a block based motion compensation scheme the pixels that cause most of the error lie at:

1. The edges of the blocks with non-zero motion vectors.
2. The image edges in the previous decoded frame.

For a segmentation based motion compensation scheme [7] the error would be concentrated at:

1. The edges of the regions which moved.
2. The "holes" i.e., places vacated by regions which are not filled by other moving regions.
3. Image edges in the previous decoded frame.

Detection of edges in the previous decoded frame is implicit in a segmentation based motion compensation scheme. For a block based scheme we can employ a simple edge detector based on either thresholding the Laplacian or the Canny edge detector.

Once the mask predicting the error locations is found, we can obtain the optimal values of the "free pixels" by using an iterative algorithm. Two natural constraints which restrict the values of the free pixels are:

1. The high frequencies in the transform domain should be as small as possible.
2. The value of the non-free pixels in the spatial domain i.e., pixels where DFD energy is high are to be retained.

An algorithm which finds a solution that lies at the intersection of the above two constraint sets should provide a maximal packing of the DFD energy in the low frequencies. We further note that the above requirements can be formulated in terms of convex constraint sets [5, 2]. Hence the algorithm can be formulated using the oft used POCS formalism [4] and is guaranteed to converge to a solution. The projection operators on to the convex sets which we can use for this purpose are:

- Projection 1: The values of the non-free pixels are assigned their actual values in the spatial domain.

- Projection 2: The high frequencies (definition depends on the transform being used) are put to zero in the transform domain.

The convex sets related to projections 1 and 2 may not intersect. In such a case the solution obtained will lie in the convex set related to projection 1 and will be as close as possible (in a mean square sense) to the convex set in Projection 2, thus minimizing the high frequency content.

4 Application in a video coder

In order to test the proposed method, we use a segmentation based motion compensation scheme [7]. The underlying segmentation is obtained using a multi-scale segmentation technique [1]. The base DFD coding scheme used for comparison purposes is due to Yokoyama [6]:

1. Partition the error image into 8x8 blocks on a regular grid.
2. Order the blocks in descending order of the residual error.
3. Send as many blocks as the bit rate permits using a DCT based coding scheme for each block.

The prediction mask for the proposed method is obtained as described in the previous section. The projection operators used are outlined in figure 1. Convergence for the iterative algorithm usually occured within 2-3 iterations. It was found that due to prediction errors, the proposed scheme might be worse than actual DCT for some blocks. We therefore selected the best of the two schemes for each block. This requires an additional 5-20 bits to be sent for each frame, which is negligible.

Figure 2 shows the improvement obtained by the proposed method over the base scheme per block for the table tennis sequence. Both the base scheme and the proposed method use the same uniform quantization strategy. The encoded bits per frame is constrained so that both the schemes perform at the same bit rate. It may be noted that the blocks with the same block number are identical for both schemes within the same graph. However the block numbers are not correlated between the two graphs. No refresher frame was sent in the simulations and the blocks are taken from the first 7 frames (approximately) for step size 16 and the first 5 frames (approximately) for step size 32. It may be observed that the average advantage due to the proposed method decreases as block number increases. This occurs due to the fact that prediction degrades as the frame number (and hence block number) increase due to lack of refresher frames.

References

[1] N. Ahuja. A transform for the detection of multi-scale structure. *IEEE Trans. PAMI*, December 1996.

[2] H. H. Chen et al. A block transform coder for arbitrarily shaped image segments. *IEEE International Conference on Image Processing*, 1:85–89, 1994.

[3] D. P. deGarrido et al. Conditional entropy-constrained vector quantization of displaced frame difference subband signals. *Proceedings IEEE ICIP*, pages 745–749, 1994.

[4] B. Javidi et al. *Real time optical information processing.* Academic Press, 1994.

[5] K. Ratakonda and N. Ahuja. Discrete multi-dimensional linear transforms over arbitrarily shaped supports. *To be published in ICASSP 97.*

[6] Y. Yokoyama et al. Very low bit rate coding using arbitrarily shaped region-based motion compensation. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(6):500–507, 1995.

[7] S. C. Yoon, K. Ratakonda, and N. Ahuja. Region based video coding using a multiscale image segmentation technique. *To be published in Proceedings ICIP-97.*

(a)	(b)
11110000	11111000
11100000	11000111
11000000	11111000
10000000	11111100
00000000	11111000
00000000	11111110
00000000	00111100
00000000	00000111

Figure 1: Projection operators for the proposed method on an 8x8 image block: (a) put all coefficients with mask value zero to zero in the DCT domain. (b) constrain the values of the pixels with mask value 1 to the actual DFD values in the spatial domain.

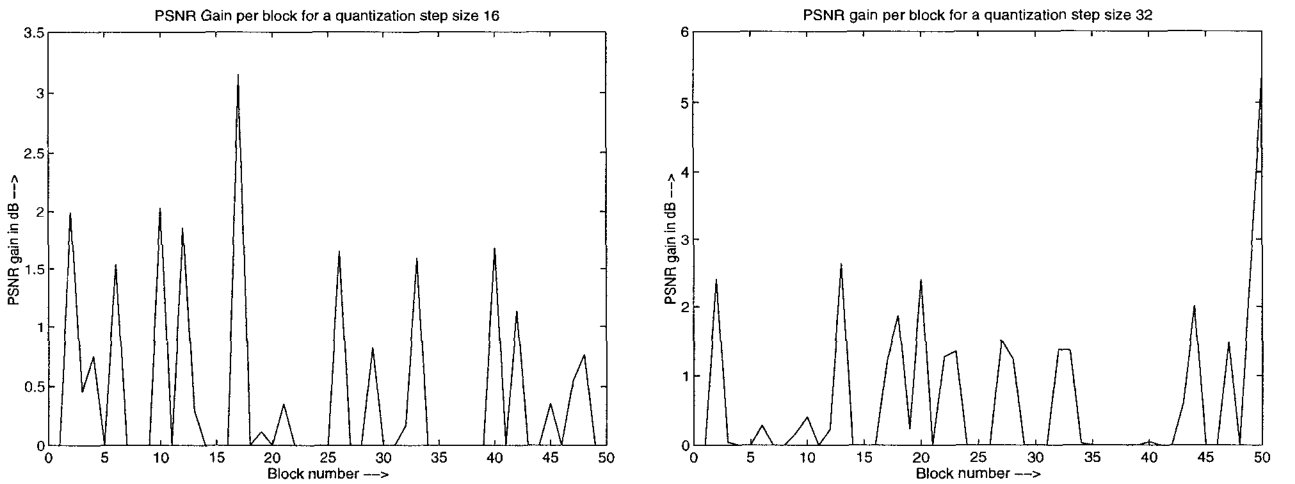

Figure 2: Comparative Results: The PSNR gain of the proposed method over normal DCT based coding scheme per block number. Block Numbers are not correlated between the two graphs.