

AN OPTICAL FLOW BASED MOTION COMPENSATION ALGORITHM FOR VERY LOW BIT-RATE VIDEO CODING

Shu Lin and Yun Q. Shi

Electrical and Computer Engineering Department,
New Jersey Institute of Technology, Newark, NJ 07102
email: shi@tesla.njit.edu

Ya-Qin Zhang

David Sarnoff Research Center, Princeton, NJ
email: zhang@sarnoff.com

ABSTRACT

In this paper, we propose an efficient compression algorithm for very low bit-rate video applications. The algorithm is based on (1) optical-flow motion estimation to achieve more accurate motion prediction fields; (2) DCT-coding of the motion vectors from the optical-flow estimation to further reduce the motion overheads; and (3) region adaptive threshold technique to match optical flow motion prediction and minimize the residual errors. Unlike the classic block-matching based discrete cosine transformation (DCT) video coding schemes in MPEG 1/2 and H.261/3, the proposed algorithm uses optical flow for motion compensation and the DCT is applied to the optical flow field instead of predictive errors. Thresholding techniques are used to treat different regions to complement optical flow technique and to efficiently code residual data. While maintaining comparable peak signal to noise ratio (PSNR) and computational complexity with that of ITU-T H.263/TMN5, the reconstructed video frames of the proposed coder are free of annoying blocking artifacts, and hence visually much more pleasant.

1. INTRODUCTION

The successful applications of ISO MPEG 1/2 and ITU-T H.261 [1][2][3][4][5] for video communications at relatively high bit-rates demonstrate that the block matching based motion compensation and DCT coding algorithm (BMDCT) works quite well at a bit-rate above 64 kbps for digital video coding [6][7][8]. However, it is well known that the block matching techniques have several serious drawbacks: unreliable motion fields in the sense of the true motion in the scene, block artifacts, and poor motion compensated prediction along moving edges [9]. At very low bit-rates (below 64 kbps), the block artifacts become severe and the

quality of the reconstructed images are degraded considerably. It is especially true in facial video coding, because most of the facial expressions involve nonrigid motion. Considerable research efforts have been made on very low bit rate coding, which result in the current H.263/TMN5 [5]. Despite its good performance, H.263 still suffers from the blocking effects, an intrinsic problem of the block-matching motion prediction and block DCT implementation.

Many techniques have been proposed to overcome the drawbacks of block matching techniques, such as overlapped windows, variable block size motion prediction. In this paper, we present a new algorithm as an improvement of H.263/TMN5. The proposed algorithm utilizes optical flow (dense motion field) instead of block matching to overcome the drawbacks of the traditional block matching and to provide more accurate predictions. We use DCT to code the high correlated optical flow vectors to reduce the overhead information. Thresholding techniques are developed to preprocess the image and to code the residual data.

2. NEW ALGORITHM

An overall block diagram of the proposed algorithm is shown in Figure 1. After pre-processing, optical flow is estimated from frames of a video sequence. These optical-flow vectors are divided block by block with each block of 8×8 . Those flow vectors which do not satisfy certain thresholds in the successive frame difference (SFD) are eliminated from coding. Those which pass the thresholds are then transformed into discrete cosine domain. The position information of these DCT coefficients are quantized, zig-zag scanned, and run-length coded followed by Huffman coding. The magnitudes of these DCT coefficients are Huffman coded. The predictive errors are thresholded first. They are then divided block by block in the same fashion. The

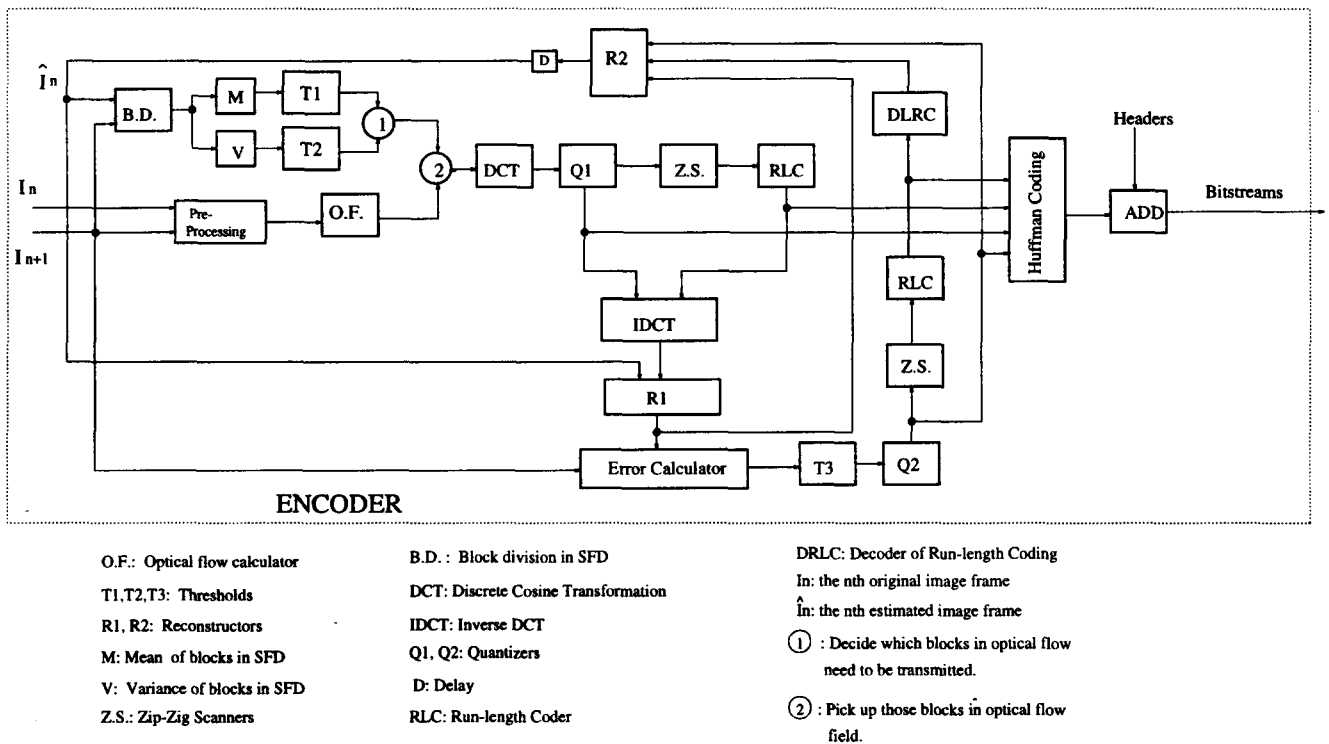


Figure 1: The encoder of the proposed algorithm

magnitude and position information of these non-zero residual data are treated in a similar way to that for the DCT coefficients of motion vectors.

2.1 Optical flow estimation

A preprocessing is to decide in which areas the optical flow vectors need to be calculated. To couple with H.263, the whole image is divided into blocks with the size of a block being 16 by 16. A measure of SFD of a block is defined as follows:

$$SFD_l = \|V_l(t) - V_l(t-1)\|, \quad (1)$$

where l is the index of the block, $V_l(t)$ represents a vector formed in a certain manner with all the pixels within the block in the image $I(x, y, t)$, and $\|\cdot\|$ means vector norm. The optical flow of block l will not be calculated except that its SFD_l is greater than a preset threshold T_{SFD} . Therefore T_{SFD} is an adjustable parameter of the algorithm. In our experimental work, the T_{SFD} is usually chosen such that only less than 25% blocks need optical flow determination. Concretely, we calculate SFD_l for each possible l , then arrange them, say, in a descending order. The T_{SFD} is chosen as such a value that there are only less than 25% of SFD values in the sequence are larger than it. The reason to do so is that there is usually only small motion experienced during the time interval between two

consecutive frames, and the change of brightness patterns mainly occurs around moving boundaries. For uniform or smooth regions in the image plane, there is no need to calculate flow vectors for video coding. In our experiments, this choice of T_{SFD} can produce both satisfactory reconstructed image quality and required bit-rate. This preprocessing saves not only huge computation, but also huge side information.

2.2 DCT coding of the motion vectors with thresholding

AR model and DCT It is considered in general that the motion vectors are transmitted as side information in motion compensated video coding schemes. However, when the very low bit-rate video coding is dealt with, the amount of the motion vector data becomes comparable with and even more than that of the error data. Therefore the motion vector coding for very low bit-rate becomes more important than that in the case of high bit-rate. Obviously, to transmit all the flow vectors needs many more bits since even after the preprocessing the number of motion vectors is much more than that in the BMDCT technique. The bits used to encode the optical flow substantially affect the transmission bit-rate.

How to code optical flow vectors? Let us consider human facial expressions. It is noted that the motion

of any point in a face is not free or independent of its neighboring points, it is constrained by some muscles and skin. That is, the motion of a point correlates with that of its neighborhood very closely. It is well known that the DCT works very efficient for highly correlated data. Hence, in the new algorithm we use the DCT to code optical flow vectors.

Using DCT to code optical flow vectors can also be justified from a theoretical analysis. That is, the probability density function (pdf) of the optical flow field associated with the Miss America sequence is verified to be very similar to that of the AR(1) model:

$$f_n = \rho f_{n-1} + \nu_n$$

where $\rho = 0.8$ and ν is a random variable, obeying Gaussian distribution with mean being 0 and variance 0.17.

Thresholding To further reduce motion vector data, we use thresholding technique at this stage as well. Most background of head-shoulder type of video frames such as in the Miss America and the Claire sequences is fixed. These regions need not to be transmitted in interframe coding. They can be copied from the previous frame to the current frame directly. To avoid complex segmentation and merging of these patches, the whole image is divided into fixed-size blocks (generally, 8×8 to cope with H.263). The mean and variance of the difference between estimated frame $\hat{I}_n$ and given frame I_{n+1} , named *SFD* in Figure 1, for those blocks whose optical flow has been estimated, are calculated. To decide which block's optical flow vectors need to be coded and transmitted, two thresholds, T_1 and T_2 , are set. If the mean of a block is less than T_1 and the variance is less than T_2 , then this block is assumed to be a non-motion block, all its contents can be replaced by the corresponding block in the previous frame and nothing is to be coded and transmitted. On the other hand, if the mean of a block is greater than T_1 or its variance is greater than T_2 , this block's motion vectors then need to be coded and transmitted. The DCT is applied to this 8×8 block. The coefficients of DCT are then quantized. And the positions of the nonzero coefficients are zig-zag scanned and run-length coded. Refer to [3] for details. The thresholds T_1 and T_2 can be viewed as adjustable parameters. Their selection will affect the quality of reconstructed frames and the bit-rate.

2.3 Adaptive coding of the residual pictures

Like all other motion compensated coding schemes, this algorithm also transmits error information. After we obtain the nonzero quantized DCT coefficients and their positions, we reorder these coefficients according to their positions, and the nontransmitted coefficients

are replaced with zeros. Then the inverse DCT is applied to recover the motion vectors. Using these motion vectors together with bilinear interpolation, if necessary, we estimate the current frame from the previous one. The difference between the estimated and actual current frames thus gives the predictive error. Because the pixel-based motion vectors rather than block-based motion vectors are used, the predictive errors are much less than that obtained by the BMDCT. The less the predictive errors, the lower the correlation they have, and the less effective the DCT, if applied, will be. For this reason, we do not apply the DCT to the error information. Instead, we transmit the errors directly. In order to use the bit-rate more effectively, another threshold T_3 is set so that only the predictive errors which are greater than T_3 will be quantized and transmitted. From experiments, it is observed that the predictive errors, needed to be transmitted in order to achieve good quality of the reconstructed image, are not scattered sparsely. Most of them are concentrated in the regions near the eyes and mouth. Using the 8×8 blocks defined above for optical flow vectors, the positions of these errors are zig-zag scanned and run-length coded in the same way as used for the nonzero coefficients of the DCT of the optical flow vectors. It is noted that T_3 is the only parameter left after T_{SFD} , T_1 and T_2 have been determined. Adjustment of this threshold T_3 may play a role to reach a desired compromise between the reconstructed frame and quality and bit-rate.

The values of both the quantized nonzero coefficients of the DCT of motion vectors and the quantized predictive errors, and the positions of these values are coded by variable length coding (Huffman codes) to further reduce the bit-rate.

3. EXPERIMENTAL RESULTS

In order to evaluate the performance of our new algorithm, we applied it to the Miss America sequence in the format of QCIF. The optical flow field is calculated with the modified Horn and Schunck algorithm[10]. The coefficients of DCT of motion vectors are quantized by 16 levels and the predictive errors are quantized by 32 levels. For 10 frames/s (30 frames/s, 2 frames are skipped for every 3 frames) interframe coding, the proposed algorithm achieves a bit-rate of 11 kbps and the PSNR of about 37.1 dB. Both are averaged for the first 36 frames of the sequence. These reconstructed frames are free of blocking artifacts. The 21st reconstructed frame is shown in Figure 2(b), the corresponding frame reconstructed with H.263/TMN5 is shown in Figure 2(c), and the original 21st frame in Figure 2(a). It is obvious that prominent blocking effects can be observed in the reconstructed frame

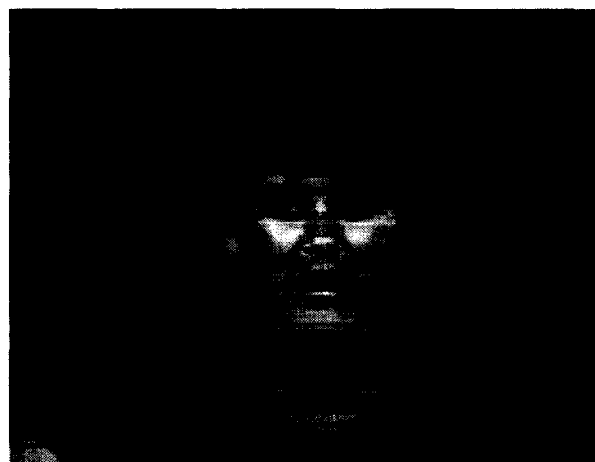
with H.263/TMN5. Similar results were achieved while working on the Claire Sequence.

4. CONCLUSION AND DISCUSSION

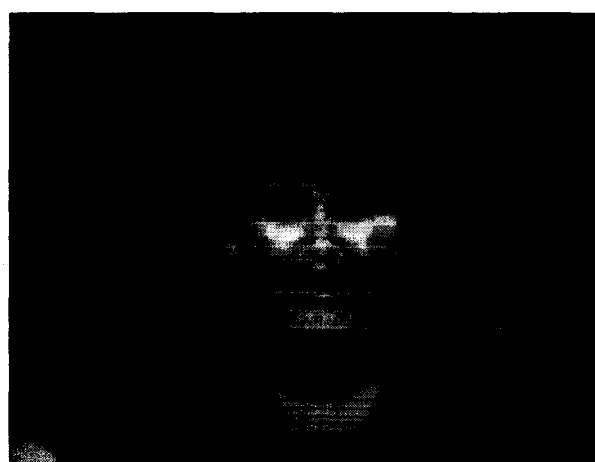
An efficient optical flow based motion compensation video coding scheme for very low bit-rate application is presented in this paper. The employment of the DCT to highly correlated optical flow motion vectors reduces data needed for motion vectors considerably. Adaptive thresholding techniques contribute to the coding efficiency of the algorithm as well. This algorithm enhances the reconstructed image quality significantly by eliminating the blocking artifacts resulted from block matching model. Experiments demonstrate that the proposed algorithm achieves superior performance than H.263 at very low bit-rate in terms of subjective evaluation. It can be widely used in various very low bit-rate applications such as video phone, teleconferencing, and wireless communication.

References

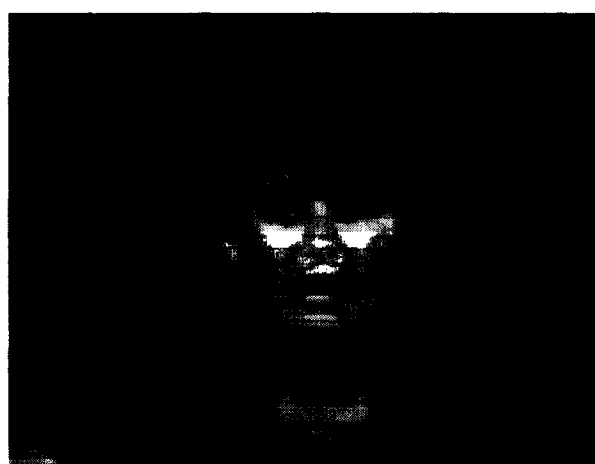
- [1] L. Chiariglione, "The development of an Integrated Audiovisual Coding Standard: MPEG," *Proceedings of the IEEE*, February 1995, vol.83, no. 2, pp. 151-157.
- [2] S. Okubo, "Reference Model Methodology - A Tool for the Collaborative Creation of Video Coding Standards," *Proceedings of the IEEE*, February 1995, vol.83, no. 2, pp. 139-150.
- [3] A. Soundararajan, "Digital Image and Video Standards," *Communications of the ACM*, April 1991, vol. 34, no. 4, pp. 47-58.
- [4] T. Ebrahimi and E. Resusens and W. Li, "New Trends in Very Low Bitrate Video Coding," *Proceedings of the IEEE*, vol. 83, no. 6, June 1995, pp. 877-891.
- [5] ITU-T Recommendation H.263, "Video Coding for Narrow Telecommunication channels at less than 64k-bits/s," (Draft) April 1995.
- [6] Y.-Q.Zhang, "Very Low Bit Rate Video Coding Standards," Invited Paper, *SPIE Visual Communications and Image Processing*, May 1995.
- [7] Y.-Q.Zhang, W. Li and M. Liou, "Special Issue on Advances in Image and Video Compression," *Proceedings of IEEE*, Feb.1995, vol.83, no. 2, pp. 135-138.
- [8] M. Kunt, "Special Issue on Advanced TV," *Proceedings of IEEE*, June 1995, vol. 83, no. 6, pp. 831-833.
- [9] F. Dufaux and F. Mescheni, "Motion estimation techniques for digital TV: A review and a new contribution," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 858-876, 1995.
- [10] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "Systems and Experiment Performance of Optical Flow Techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43-77, 1994.



(a) The 21st original frame of Miss America sequence



(b) The reconstructed 21st frame with proposed algorithm



(c) The reconstructed 21st frame with H.263/N

Figure 2: Miss America