

Efficient Motion Estimation Using Spatial and Temporal Motion Vector Prediction

Ismail Ismaeil, Alen Docef, Faouzi Kossentini, and Rabab Ward

Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, BC, Canada V6T 1Z4
E-mail:{ismael, alen, faouzi, rababw}@ece.ubc.ca

ABSTRACT

This paper presents a motion estimation technique for the coding of moving video sequences that is based on spatial and temporal prediction. The motion vector of a moving object is tracked using spatial and temporal prediction and used as a starting point for the motion estimation search algorithm. The predicted motion vector is selected from several candidate motion vectors according to the block matching criterion. Experimental results show that spatio-temporal prediction reduces the number of computations performed by the motion search algorithm by 30% for MPEG-2 encoding and by 40% for H.263 encoding.

1. INTRODUCTION

In block based video encoders such as MPEG-2 [1] and H.263 [2], block matching algorithms (BMA) are widely used for motion estimation (ME). During encoding, motion estimation is usually the most computationally intensive task. The objective of motion estimation is to find the best match for a macroblock in the current picture from a set of candidate macroblocks in a restricted search area in the previously encoded frame. The sum of absolute differences (SAD) is the most widely used matching criterion. The full-search block matching algorithm (FS-BMA) guarantees that the minimum SAD value is obtained, but it is computationally expensive. Several fast algorithms have been proposed in the literature [3, 4, 5, 6] that achieve suboptimal compression performance with a significantly reduced computational complexity. The computational complexity is reduced by decreasing the number of search locations, and/or by using subsampled macroblocks to calculate the SAD for each location. A popular technique used to reduce the number of search locations is to compute a predicted motion vector (PMV) and use it as the starting point for the motion search algorithm. The performance of any predictive motion estimation algorithm is directly affected by the accuracy of the pre-

dicted motion vector. If the predicted motion vector is close to the optimum motion vector, then the motion vector search algorithm will end after a small number of SAD computations. On the other hand, if the predicted motion vector is far from the optimum motion vector, the motion estimation algorithm will either be trapped in a local minimum or the algorithm will spend many more computations to find the optimum motion vector. Extensive research has been conducted to develop techniques that increase the prediction accuracy of the motion vector [7, 8, 9, 10]. These prediction techniques exploit motion vector dependencies in the spatial and/or temporal domain. In this paper we present a motion estimation technique based on spatio-temporal motion vector prediction.

2. SPATIO-TEMPORAL PREDICTION METHOD

In moving video, the motion of an object has high correlation in the spatial and temporal domains. Many motion prediction methods that exploit this spatio-temporal correlation have been proposed in the literature [7, 8]. In this paper we combine both spatial and temporal prediction to obtain an estimate of the motion vector for the current macroblock. Next we discuss the spatial and temporal predictors and the PMV selection algorithm.

2.1. Spatial Prediction

A popular spatial predictor is the median of the three previously coded motion vectors corresponding to the macroblock to the left (A), above (B) and above right (C) as shown in Figure 1. The median predictor is very simple, yet its accuracy is quite high [12] for the sequences with a relatively uniform motion. Several spatial predictors are considered in this paper, using various macroblocks in the region of support depicted in Figure 1. The motion vectors corresponding to these

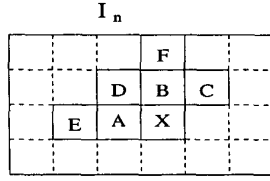


Figure 1: The region of support used for motion vector prediction.

macroblocks are considered as candidates for the spatial PMV. The block matching function is then computed for the candidate motion vectors as applied to the current macroblock. The spatial PMV is set to be equal to the candidate motion vector which yields the best match. Experimental results will show that the computational overhead of computing the extra SADs is offset by the increased MV prediction accuracy.

2.2. Temporal Prediction

For sequences with small and/or fast moving objects, the accuracy of the spatial predictor may deteriorate. To solve this problem, temporal motion prediction methods have been proposed that attempt to exploit the temporal correlation in the motion field. Most of the temporal prediction methods proposed in the literature compute the PMV as the motion vector of the macroblock in a previously encoded frame that has the same spatial location as the current macroblock. In this paper, we propose a temporal predictor based on a projection technique, and we compare its performance with the simple predictor described above.

The temporal projection tracks the motion vector of a moving object from one frame to another. As shown in Figure 2, the motion vector V_{n-1} of macroblock B_{n-1} in the previously encoded frame I_{n-1} is projected on the macroblock B_n in the current frame I_n . The motion vector V_{n-1} is used as a candidate vector in the relevant macroblock B_n . The motion vector V_{n-1} will be used in the determination of the starting point for the motion estimation search algorithm. When projecting the motion vector determined for frame I_{n-1} to frame I_n , the motion vector may point to a 16×16 -pixel square that is not a macroblock, that is, its upper left corner does not have coordinates that are multiples of 16. The projected motion vector is then assigned to the macroblock in frame I_n that best overlaps the projected 16×16 -pixel square. Note that the overlap area for the closest macroblock contains at least 8×8 pixels.

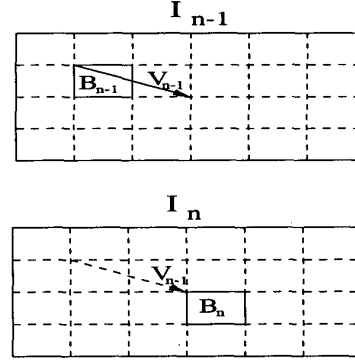


Figure 2: The temporal projection of a motion vector.

Ideally, one temporal PMV is obtained for each macroblock in the current frame. For some macroblocks, no temporal prediction is available and motion vector prediction relies on the spatial prediction only. Another irregular situation occurs when more than one motion vector points to the same macroblock in the current frame. In this case, we chose the motion vector based on a measure of its reliability measure, which is defined as follows. As shown in Figure 3, we assume that the motion vector V_{n-1} , predicted between the current frame I_n and the previous frame I_{n-1} , is reliable if $V_{n-1} \cong V_{n-2}$, where V_{n-2} is the motion vector detected between frames I_{n-2} and I_{n-1} . To estimate the reliability R of the predicted vector, we use the following equation,

$$R = \left(MSE(V_{n-2}, V_{n-1}) + \frac{1.0}{Overlap\ area} \right)^{-1}, \quad (1)$$

where $MSE(V_{n-2}, V_{n-1})$ is the mean square error between the vectors V_{n-2} and V_{n-1} . The overlap area is measured as the number of pixels common between macroblocks B_{n-1} and B_n if we assume that B_{n-1} is displaced by the vector V_{n-1} . The reliability measure will increase if the motion vector belongs to a moving object with constant speed. In the case where more than one motion vector points to the same macroblock in the current frame, we chose the one with the higher reliability measure R .

2.3. Predicted Motion Vector Selection

An inaccurately predicted motion vector can have a significant negative impact on the motion estimation algorithm. To choose between the temporal and spatial PMVs, we use the same procedure used to select

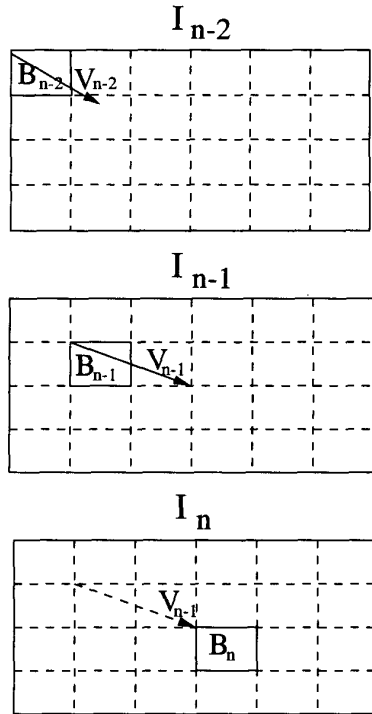


Figure 3: Motion vector reliability.

the spatial PMV described in section 2.1. If the temporal PMV exists, then its corresponding SAD is computed and compared to the SAD obtained by the spatial PMV. The PMV is chosen that yields the smallest SAD.

When a highly-optimized motion estimation algorithm is employed, the overhead of computing the SADs for PMV selection may become significant. In this case, a downsampled SAD can be computed for the purpose of PMV selection. The number of operations is reduced by 50% for a 2:1 vertically downsampled SAD and by 75 % for a 2:1 vertically and horizontally downsampled SAD. The effect of using a downsampled SAD on the PMV accuracy is evaluated in the next section.

3. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed motion vector prediction algorithm in the context of high-bit rate MPEG-2 encoding and low-bit rate H.263 encoding. The motion estimation algorithm used in this paper is described in detail in [12, 13]. This algorithm uses a floating center diamond-shaped search area and a computation-distortion stopping criterion.

The MPEG-2 motion estimation uses the PMV as a starting point for a hierarchical search algorithm con-

sisting of a low-resolution ME, followed by normal resolution frame and field ME. Finally, for both frame and field ME, a half-pel refinement is performed. Several sequences with different motion types were encoded at a target bit rate of 3 Mbps. The sequences BUS and GARDEN have uniform motion, while the sequence NBA has complex motion. The sequence TABLE TENNIS has localized moving objects on a relatively constant background. The sequence BALLET has relatively slow motion.

In the first experiment we evaluated the performance of several spatial predictors with different regions of support. The results are compared with the median predictor used in [12]. Table 1 shows the average number of SADs versus the PSNR values for these predictors. We can see that the best performance is obtained by the spatial predictor that uses three previously encoded MVs. In this case the number of SADs is relatively small and the PSNR loss is insignificant. We can also see that our best spatial predictor outperforms the median predictor by up to 0.1 dB. In the second experiment we compare the spatio-temporal PMV with and without projection. Table 2 shows that using the projection method improves the coding performance by approximately 0.1 dB. More important, the spatio-temporal PMV leads to a reduction in the number of SAD computations of up to 40 % for a constant quality level. This is shown in Table 3 for all test sequences. The effect of using the downsampled SAD in the prediction mode selection is shown in Table 4 for both the vertically and vertically/horizontally downsampled SADs. The computational overhead due to the motion vector prediction algorithm is greatly reduced with minimal performance loss. Figure 4 shows the rate-distortion curves for MPEG-2 encoding using our PMV and the median PMV for the BUS sequence for a constant number of SAD computations. We can see that the proposed spatio-temporal prediction method outperforms the median prediction method by about 0.15 dB at all bit rates.

For H.263 encoding we used two QCIF size sequences: FOREMAN and NEWS. The sequence FOREMAN has complex motion and the sequence NEWS has slow motion. Table 5 shows that the spatio-temporal PMV reduces the SAD computations by 40 % as compared to the median PMV with no loss in performance. Figure 5 shows the rate-distortion curves for H.263 encoding using the proposed PMV and the median PMV for the FOREMAN sequence for a constant number of computations. We can see that the proposed spatio-temporal prediction method outperforms the median prediction method by about 0.25 dB at all bit rates.

4. CONCLUSIONS

This paper presents a motion vector tracking method for video sequences that is based on spatio-temporal prediction. The predicted motion vector is selected from several candidate motion vectors using the block matching algorithm. The experimental results show that the spatio-temporal prediction reduces the number of computations performed by 30% for an MPEG-2 encoder, and by 40% for an H.263 encoder while maintaining the same quality.

5. REFERENCES

- [1] ISO/IEC 13818-2—ITU-T Rec. H.262, *Generic Coding of Moving Pictures and Associated Audio Information: Video*. ISO/IEC, 1995.
- [2] ITU Telecom. Standardization Sector Study Group 15, "Document LBC-95 Working Party 15/1 Question 2/15," *Draft Recommendation H.263*, Leidschendam, 7 April 1995.
- [3] T. Zahariadis and D. Kalivas, "A spiral search algorithm for fast estimation of block motion vectors," *In Proc. European Signal Processing Conference (EUSIPCO-96)*, September 1996.
- [4] L. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *IEEE Transaction on Circuit and Video Technology*, Vol. 6, no. 4, August 1996.
- [5] Y. Lee, F. Kossentini, M. J. Smith and R. Ward, "Predictive RD-constrained motion estimation for very low bit rate video coding," *MPEG-4 Special Issue of the IEEE Transactions on Selected Areas in Communications*, 1997.
- [6] J. Mitchell, W. Pennebaker, C. Fogg, and D. LeGall, "MPEG video compression standard," *Chapman & Hall, International Thomson Publishing*, 1996.
- [7] H. Sonehara, Y. Nojiri, K. Iguchi, Y. Sugiura, and H. Hirabayashi, "Motion estimation method using the spatiotemporal characteristics of moving objects," *SMPTE Journal*, pp. 340-347, June, 1998.
- [8] M. Mattavelli, and G. Zoia, "Vector tracing techniques for motion estimation algorithms in video coding," *European Signal Processing Conference (EUSIPCO-98)*, August 1998.
- [9] D. Kim, J. Choi, and I. Kim, "Adaptive motion estimation based on spatio-temporal correlation," *Signal Processing: Image Communication*, Vol. 13, pp. 161-170, 1998.
- [10] B. Schwartz and K. Sauer, "Efficient block motion estimation using integral projection," *IEEE Transaction on Circuit and Video Technology*, Vol. 6, no. 5, pp. 513-518, October 1996.
- [11] J. Chalidabhongse, S. Kim, and C. Kuo, "Fast motion estimation by using spatio-temporal correlation of motion field," *in SPIE Proc. Visual Communications and Image Processing*, Vol. 2501, pp. 810-821, 1995.
- [12] Y. Lee, F. Kossentini, and R. Ward, "Efficient MPEG-2 Encoding of Interlaced Video," *Special issue of the Canadian Journal of Electrical and Computer Engineering*, Vol. 23, No. 1-2, pp. 61-67, June 1998.
- [13] M. Gallant and G. Cote and F. Kossentini, "Computation Constrained Block-Based Motion Estimation Algorithm for Low Bit Rate Video Coding," *Accepted for publication in IEEE Transactions on Image Processing*, march 1999.

Sequence	BUS		TABLE TENNIS	
	SADs	PSNR	SADs	PSNR
spatial PMV				
ABCDEF	60.54	31.08	56.96	28.56
ABCDE	59.60	31.06	55.97	28.56
ABCD	58.72	31.06	55.08	28.54
ABC	57.39	31.05	54.10	28.54
AB	57.09	30.99	53.94	28.50
A	57.85	30.41	52.72	28.29
B	59.30	29.98	52.88	28.24
median	57.38	30.96	56.98	28.41

Table 1: Average number of SAD operations per macroblock and the PSNR for several spatial predictors.

Sequence	BUS		TABLE TENNIS	
	SADs	PSNR	SADs	PSNR
spatial PMV				
projection	35.13	30.97	33.29	28.48
no projection	35.27	30.91	34.12	28.40

Table 2: Average number of SAD operations per macroblock and the PSNR for several spatial predictors.

Sequence	Spatio-Temporal		Median	
	SADs	PSNR	SADs	PSNR
BUS	36	30.97	57	30.96
BALLET	37	44.20	53	44.24
GARDEN	46	29.18	51	29.16
T. TENNIS	34	28.48	57	28.42
NBA	47	28.58	61	28.59

Table 3: Average number of SAD operations per macroblock and the PSNR for the two methods.

Sequence	BUS		TABLE TENNIS	
	SADs	PSNR	SADs	PSNR
MB resolution				
full resolution	40.26	31.03	36.31	28.52
subsampling by 2	36.29	31.01	36.31	28.50
subsampling by 4	35.13	30.97	33.29	28.48

Table 4: Average number of SAD operations per macroblock and the PSNR using downsampled macroblocks used to compute the SAD.

Sequence	Spatio-Temporal		Median	
	SADs	PSNR	SADs	PSNR
FOREMAN	23.50	29.64	38.60	29.63
NEWS	22.68	29.86	40.8	29.88

Table 5: Average number of SAD operations per macroblock and the PSNR for the two methods.

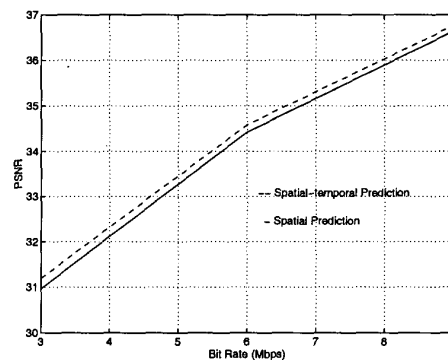


Figure 4: The performance of the two methods for the BUS sequence.

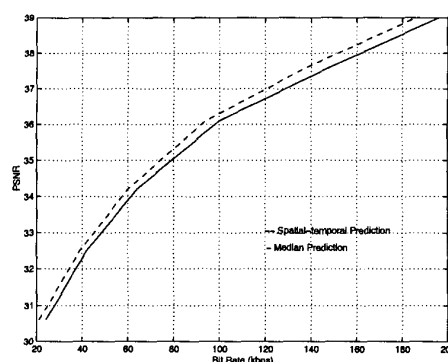


Figure 5: The performance of the two methods for the QCIF FOREMAN sequence.