SHORT/LONG-TERM MOTION VECTOR PREDICTION IN MULTI-FRAME VIDEO CODING SYSTEM

Yi-Hon Hsiao, Tien-Hsu Lee, and Pao-Chi Chang

Department of Electrical Engineering, National Central University, Taiwan, R.O.C.

ABSTRACT

Multi-frame motion estimation is a novel technique to improve the video coding efficiency. However, the computational complexity in motion search is usually a bottleneck in the video coding system. According to the experiences of conventional techniques, good prediction in selecting an initial search point can further accelerate the convergence speed of fast motion search algorithms. In this paper, two motion vector prediction techniques are developed for the speed-up of multi-frame motion estimation. The proposed method can efficiently increase the prediction accuracy and reduce the search point up to 25% with the similar quality and bit-rate under the New Diamond Search framework in H.264.

1. INTRODUCTION

Multi-frame motion estimation extends the spatial displacement vector utilized in the block-matching video coding by permitting the use of more frames than the previously decoded one for the motion-compensated prediction. The use of multiple frames for the motion estimation in most cases provides significantly improved coding gain [1]. In H.264 [2], the multiple reference picture selection allows the encoder to store up to five previous frames for motion estimation. However, the computational complexity will increase dramatically. Numerous fast search algorithms in the single-frame motion estimation are proposed in the past. As we know, a good initial point can speed up the convergence velocity of fast search algorithms. Thus, we focus on discussing how to provide more precise prediction of an initial search point in the multi-frame condition. This is very useful for the real-time video conferencing applications.

Multi-frame motion estimation was proposed as a technique to improve the error-resilience of compressed video by Budagavi and Gibson [3], as incorporated into the Annex N of H.263 [4]. Then, multi-frame motion estimation for improving the compression performance was introduced by Wiegand *et al.* in 1997 [5]. This technique became part of Annex U "Enhanced Reference Picture Selection" of H.263 in 1999 [6], and is now an integral of the emerging JVT/H.264/AVC standard.

Most motion estimation strategies previously employed a fixed initial starting point, such as the center of the original search window. However, the spatial/temporal correlation in the motion vector fields is often quite high and can be used to predict a good initial search point close to the global minimum. The main idea of [7] selected a set of initial motion vector candidates from the spatially or temporally neighboring blocks and chose the best one as the initial estimation for further refinement. In [8], Duanmu *et al.* observed that there is a great deal of correlations between two neighboring reference frames. They utilized the neighboring reference frames to predict the initial search point in the multi-frame motion estimation. However, this algorithm can be improved to perform the prediction more precisely.

In the following section, the preliminary analyses are introduced first. Then, various motion vector prediction algorithms including the proposed ones are described in Section 3. Finally, we show the simulation results in Section 4 and give the conclusion in Section 5.

2. PRELIMINARY ANALYSIS

2.1. Rate-distortion model

The utilization of multiple reference frames for motion compensation provides significantly improved coding gain. However, the reference frame parameter has to be transmitted as side information and required additional bit-rate. To control the bit-rate budget, the rate-distortion model is usually utilized in the motion estimation process.

In H.264, the best matching motion vector is found by minimizing the following cost function:

$$J(\mathbf{m}, \lambda_{MOTION}) = SAD(s, c(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p}) \quad (1)$$

with $\mathbf{m}$ being the motion vector, $\mathbf{p}$ being the prediction for the motion vector, and λ_{MOTION} being the Lagrange multiplier. The term $R(\mathbf{m}\text{-}\mathbf{p})$ represents the bits used to encode the motion information only and obtained by

table-lookup. In this paper, we call this kind of prediction as the coding-based motion vector prediction and abbreviate it as CB-MVP. The SAD (Sum of Absolute Differences) is computed as the following equation.

$$SAD(\ m,n\ ) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |\ c(i,j) - s(i+m,j+n)\ | \qquad (2)$$

2.2. Temporal relationship

Here, we will analyze the frame-by-frame relationship in the multi-frame motion estimation. In Fig. 1, each parallelogram represents a frame of video sequence. The arrow with hollow line is the time-direction of the video stream. The parallelogram with the dotted line represents the current frame in different periods. There are five parallelograms on the left of the parallelogram with the dotted line. The five parallelograms represent previous encoded frames stored in the memory. In multi-frame motion estimation, the current frame would compare with the five previous frames in the buffer one by one, i.e., from the delay time $\Delta=1$'s frame to $\Delta=5$'s frame. As the illustration in Fig. 1, we can find that the parallelograms with the same pattern are very similar and the $\Delta=2$'s frame in period T is indeed the $\Delta=1$'s frame in period $T-1$. If we can make use of the data having done before to predict the probable motion vector position, we may approach to the optimal position more accurately.
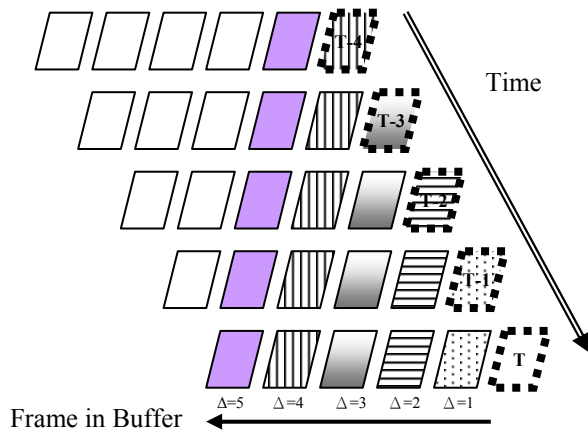

Fig. 1. The relationship of multi-frame motion estimation.

3. VARIOUS PREDICTION METHODS

The purpose of this paper is to discuss the issue about how to predict the motion vector more precisely in the multi-frame condition. The most important point, however, is how to find out the differences between single-frame and multi-frame motion estimation. Clearly, the temporal frame-by-frame relationship in the multi-frame condition is our starting point to predict the probable position. In this section, we will propose two prediction methods, i.e., short-term motion vector prediction and long-term motion vector prediction. Duanmu's algorithm will also be briefly described later for comparison.

3.1. Short-term motion vector prediction (ST-MVP)

Short-term motion vector prediction uses only information in the period of *T-1*. The corresponding motion vector difference between two successive images in the period of *T-1* is similar to that of period *T*. This relationship is depicted in Fig. 2. The bold line with double arrows illustrates the motion vector difference can be reused in the next period. Since this prediction method only utilizes the data of successive frames in the prior period, we name it as the short-term motion vector prediction (ST-MVP).
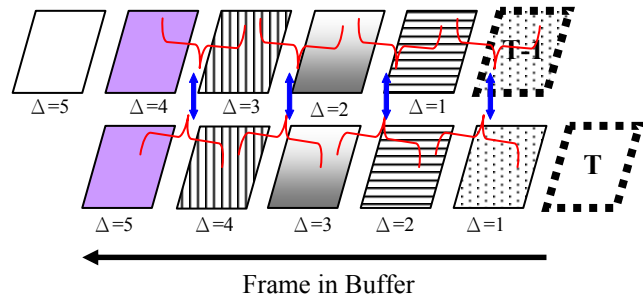

Fig. 2. The relationship of ST-MVP.

3.2. Long-term motion vector prediction (LT-MVP)

The same idea can be applied to the long-term history situation. In other words, we employ the information in different periods instead of the same period. According to Fig. 3, the squares with the same pattern are very similar. We can utilize the search results of $\Delta=1$ in different periods to be the motion vector differences between two successive frames in the current period. Since this prediction method utilizes the data in different periods, we name it as the long-term motion vector prediction (LT-MVP).
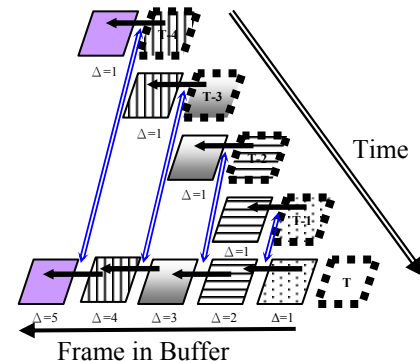

Fig. 3. The relationship of LT-MVP.

3.3. Continuous tracking motion vector prediction (CT-MVP) [8]

Duanmu *et al.* find that the motion vectors corresponding to two neighboring reference frames are very similar. They use this characteristic to develop the fast search algorithm. The main idea is that if one of two neighboring frame-dependent motion vectors is known beforehand, the other one can be estimated with very high degree of certainty by searching only a small area centered on the location corresponding to the known vector. In this paper, we name this idea as the continuous tracking motion vector prediction (CT-MVP). Although two continuous reference frames are very similar, there still exist some differences between two neighboring frames. In contrast, our proposed algorithm has considered such differences and performed the prediction more accurately.

3.4. A typical example of ST-MVP, LT-MVP, and CT-MVP

To clarify the description, a typical example is given and drawn in Fig. 4. It is assumed that we would like to search the motion vector within Δ=3 in the period of *T*. MV1 is the motion vector within Δ=2 in the period of *T*. MV2 and MV3 are the search results within Δ=2 and Δ=1 individually in the period of *T-1*. MV2' is the search result within Δ=1 in the period of *T-2*. In ST-MVP, we use the difference of motion vectors between Δ=2 and Δ=1 in the period of *T-1* as the motion vector prediction in the period of *T*. The final ST-MVP's result is the sum of MV1 and the difference between MV2 and MV3. In LT-MVP, we use the motion vector within Δ=1 in the period of *T-2* as the motion vector prediction in the period of *T*. The LT-MVP's result is the sum of MV1 and MV2'. In CT-MVP, only the search result of neighboring reference frames is utilized in the motion vector prediction. Consequently, the CT-MVP's result is MV1.
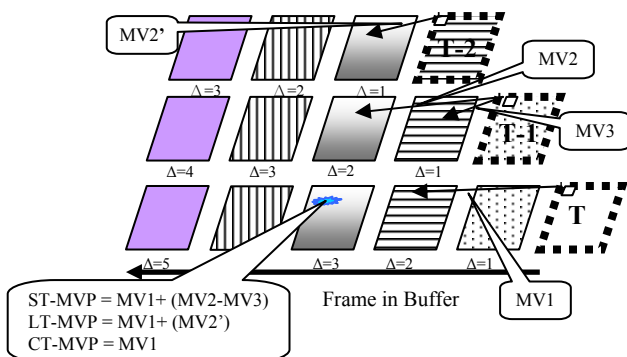

Fig. 4. An example of ST-MVP, LT-MVP, and CT-MVP.

3.5. Amendment to ST-MVP/LT-MVP

Even though ST-MVP and LT-MVP can achieve good performance, they may increase bit rates slightly while being tied with fast search algorithms. In order to remove this weakness, we can further combine CB-MVP with ST-MVP/LT-MVP to amend the proposed algorithms. We take the LT-MVP as an example and demonstrate the amendment in Fig. 5. We name this hybrid prediction procedure as the amendment to long-term motion vector prediction, and abbreviate it as ALT-MVP.

The procedure is described as follows. After executing the LT-MVP, we calculate and compare the costs of CB-MVP and LT-MVP results. Then the final prediction result is chosen as the position with the lower cost. This additional decision process will increase the complexity slightly, but it can solve the problem of bit-rate increase.
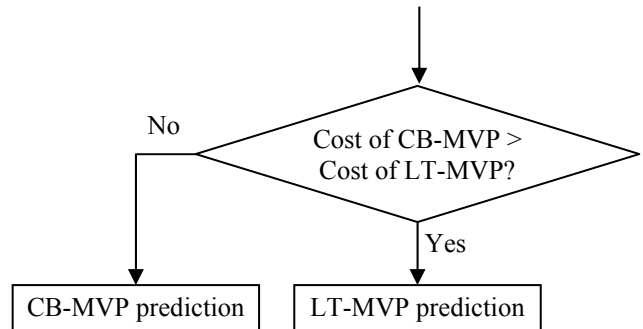

Fig. 5. The procedure of ALT-MVP.

4. SIMULATION RESULTS

The performances of various prediction methods can be verified by the prediction distance measurement. The prediction distance is defined as the following equation:

$$\text{Distance}(p, f) = \sqrt{(p_x - f_x)^2 + (p_y - f_y)^2} \qquad (3)$$

with p being the prediction result of the motion vector, and f being the result of full search. The experimental results are obtained from the H.264 JM6.1 encoder and depicted in Fig. 6 (Center represents the prediction that is just the center of the original search window). Typical test sequences, such as Akiyo, Carphone, Coastguard, News, Foreman, and Mobile, are used in the experiments. The picture format is QCIF. We can find that the proposed algorithms have better prediction performance than CT-MVP and CB-MVP in the distance measurement. By the way, there is less significant performance distinction between ST-MVP and LT-MVP. In our tests, LT-MVP is slightly better than ST-MVP in five video sequences. They are Akiyo, Carphone, Coastguard, Foreman, and Mobile.

Good prediction in selecting an initial search point can further accelerate the convergence speed of fast search algorithms. We apply the New Diamond Search (NDS) [9] with different prediction methods to show performance comparisons in terms of the search point and bit-rate. Test results are shown in Fig. 7 and Fig. 8 and represented by the percentage to the one that simply applies single-frame NDS with the center initial search point to all multi-frames. The PSNR values are not compared here because they are nearly the same. As analogized to Fig. 6, the simulation results show that LT-MVP can improve the search efficiency up to 25% while retaining the same video quality. Although LT-MVP increases the bit-rate slightly while comparing with CB-MVP, ALT-MVP can amend this problem successfully with moderate search point increase.

5. CONCLUSION

Good prediction in selecting an initial search point can further accelerate the convergence speed of fast search algorithms. In this paper, two prediction schemes with the amendment are developed for the multi-frame motion estimation. The proposed prediction method can easily cooperate with various fast search algorithms to further reduce the search point efficiently.

6. REFERENCES

[1] K. Dovstam, "*Video Coding on H.26L,*" Master Science Thesis of Royal Institute of Technology, Ericsson Research.

[2] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

[3] M. Budagavi and J. Gibson, "Multi-frame Block Motion Compensated Video Coding for Wireless Channels," *in Thirtieth Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, pp. 953-957, Nov. 1996.

[4] ITU-T Recommendation H.263, "*Video Coding for Low Bit-rate Communication,*" Annex N, 1997.

[5] T. Wiegand, X. Zhang, and B. Girod, "Block-Based Hybrid Video Coding Using Motion-Compensated Long-Term Memory Prediction," *in Proc. of the Picture Coding Symposium*, Berlin, Germany, pp. 153-158, Sept. 1997.

[6] ITU-T Recommendation H.263 "*Video Coding for Low Bit-rate Communication,*" Annex U, 1999.

[7] B. Zeng and R. Li, "Optimization of Fast Block Motion Estimation Algorithm," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 833-844, Dec. 1997.

[8] C. Duanmu, M. O. Ahmad, and M. N. S. Swamy, "A Continuous Tracking Algorithm for Long-Term Memory Motion Estimation," *ISCAS*, vol. 2, pp. 356-359, May 2003.

[9] S. Zhu and K. K. Ma, "A New Diamond Search Algorithm for Fast Block Matching Motion Estimation," *ICICS*, vol. 1, pp. 9-12, Sept. 1997.
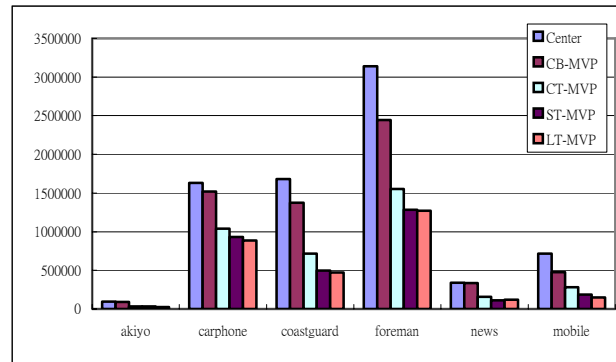
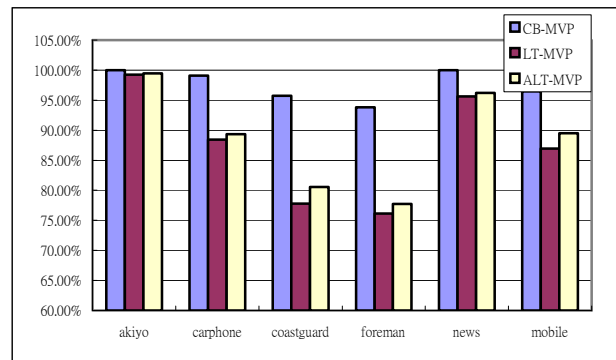Fig. 6. Distance between prediction position and motion vector obtained by full search.

Fig. 7. Search point comparison for various motion vector prediction algorithms.
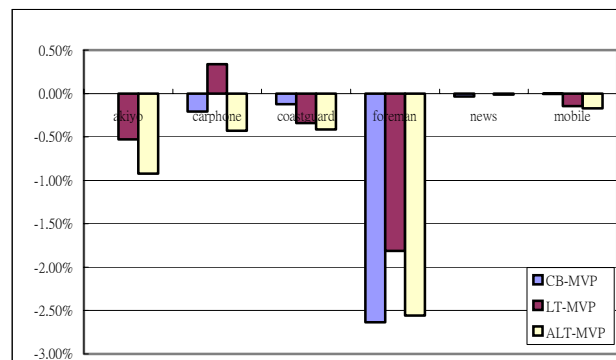

Fig. 8. Bit rate comparison for various motion vector prediction algorithms.