REGION BASED SEARCHING FOR EARLY TERMINATED MOTION ESTIMATION ALGORITHM OF H.264/AVC VIDEO CODING STANDARD

Mohammed Golam Sarwer and Q.M. Jonathan Wu

Computer Vision and Sensing System Laboratory
University of Windsor, Windsor, ON, Canada
sarwer@uwindsor.ca, jwu@uwindsor.ca

ABSTRACT

H.264/AVC is the newest international video coding standard which shows extremely higher coding efficiency compared to any other previous standards, but it brings high computational complexity. The variable block size motion estimation process is the most time consuming part of the H.264/AVC encoder. In order to reduce the computation of the full search motion estimation method, this paper presents a region based search algorithm. The entire search window is divided into several search regions based on the orientation of the previously calculated motion vector. The experimental results show that the proposed method reduced about 77% of motion estimation time with ignorable degradation of rate-distortion performance.

Index Terms— H.264/AVC, video coding, motion estimation, search region, rate-distortion

1. INTRODUCTION

The newest video coding standard H.264/AVC achieves [1] a significant improvement in coding efficiency compared to previous standards, such as MPEG-1/2/4 and H.261/H.263. But the computational complexity of the encoder drastically increased. Among the several new features which are introduced by H.264/AVC, the motion estimation (ME) process is highly computationally intensive and counts for about 80% of the total computation complexity [7]. Therefore, the development of efficient algorithms for the ME of H.264/AVC is one of the most challenging themes.

One way to reduce the complexity of ME process is to terminate the ME calculation early. Up to date, a number of efforts has been made to explore the early termination algorithms in motion estimation and mode decision for H.264/AVC video coding [2-5]. The well-known zero-motion detection algorithm [2] compares sum of absolute difference (SAD) of two blocks with a predefined threshold to determine the stationary block and then skips the remaining search points. A fast motion estimation algorithm with adaptive early termination threshold for the current accumulated partial SAD is proposed in [5]. Recently, improvement on fast motion estimation method adopted by H.264/AVC reference software is presented in [4]. The variable block-size zero motion detection and variable block-size best motion detection algorithms compare the rate-distortion cost function of the two blocks instead of the SAD to detect the zero motion blocks or best motion vector for the variable block-size H.264/AVC video coding standard [3]. However, this algorithm is using fixed threshold values that are difficult to determine good values for remaining the high accuracy with different sequences and parameters setting. This method searches either only one search point or all of the search points in the search range. Therefore, computation reduction is not so high. The efficiency of early termination algorithms is improved by determining the adaptive threshold based on the rate-distortion (RD) cost of highly correlated blocks [7].

In this paper, a region based searching method is presented to further reduce the computation of early termination method for motion estimation. In this technique, the entire search window is divided into several regions based on the direction of motion vector of highly correlated block and mode. The rest of the paper is organized as follows. In section 2, review of early termination method of H.264/AVC is explained. Section 3 describes the algorithm of proposed method. Simulation results are presented in section 4. Finally, section 5 concludes the paper.

2. EARLY TERMINATION ALGORITHM

The objective of the early termination is to decide whether a search point has meet the RD cost criterion so that the best search point for the current block can be terminated early without trying the rest of the search points. If the cost of a search point satisfies (1), this search point is taken as best search point and then remaining searches can be skipped [7].

$$J < T_m \text{, for } m=1, 2, ...7 \quad (1)$$

Where, J is the RD cost and T_m is the threshold for mode m. After searching each search point, the current RD cost J is compared against the threshold value of the specific block type and the search is terminated if this RD cost J is lower

than the threshold value. The more detail description and selection of adaptive thresholds are introduced in [7].

3. PROPOSED REGION BASED SEARCHING

The full search algorithm utilized in H.264/AVC applies a pixel-based spiral search method. The search area is centered at the same coordinates as the previous frame block and the search process of the search space moves outward in a spiral. This method suits video sequences with a motion vector distribution that is close to the search center, but for high motion sequences, the spiral search method is not more efficient [6]. This problem is addressed by dividing the search window into different regions and representing each region with a square of NxN size [6], with the size of the square related to the complexity reduction. Additionally, the early termination algorithm described in the previous section is based on the inter-mode correlation [7], so in order to apply the region-based search, a modification is necessary. Fig. 1 displays the 41 motion vectors of a low motion macroblock (MB) (Akiyo) and a high motion sequence (Stefan). All the motion vectors are more directionally oriented for the high motion sequences, so that dividing the search window into a square pattern is not as suitable as it is for low motion sequences.

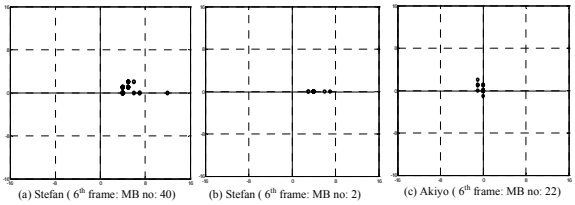

Fig. 1: 41 Macroblock motion vectors

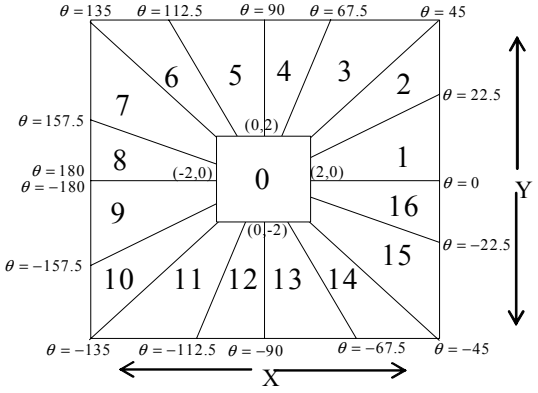

Fig. 2: Partitioned search range

Based on this observation, we have divided the search window into 17 different regions as shown in Fig. 2. Region 0 represents the low motion vectors. In this region, $X \le 2$ and $|Y| \le 2$. Regions 1-16 are based on the angle θ , which is defined as follows:

$$\theta = \tan^{-1}\frac{Y}{X}, \text{ where } |X| > 2 \text{ and } |Y| > 2 \quad (2)$$

The goal of this algorithm is to reduce the number of search points in the early termination method, so the starting search region is a key factor of the resulting computation reduction. After the search range is partitioned, an adaptive search starts from the most probable region. All the successive video frames are correlated in a temporal direction, so we can expect most of the picture movements to follow in the same direction. Table I gives the motion vector (MV) percentage for the previously calculated modes. This analysis collected data from five different types of video sequences (Foreman, Akiyo, Stefan, Mobile and Carphone) to produce the average results presented in Table I. We have seen that the MV of the 16x8, 8x16 and 8x8 modes correlate highly with the 16x16 mode, while the MV of the 4x8, 8x4 and 4x4 modes correlate more with the 8x8 mode of the current MB. Based on this analysis, Table II presents the most probable MV correspondence with different modes.

Table I: % of the event "MV at target mode=MV with previously calculated mode"

Target mode	Previously computed mode						
	Collocated MB in previous frame	Current MB in current frame					
	16x16	16x16	16x8	8x16	8x8	8x4	4x8
16x16	77.5	-	-	-	-	-	-
16x8	72.6	82.3	-	-	-	-	-
8x16	67.2	82.0	76.2	-	-	-	-
8x8	61.7	76.4	73.2	74.1	-	-	-
8x4	59.8	73.6	71.2	69.8	82.4	-	-
4x8	58.2	73.4	70.4	69.4	82.6	75.4	-
4x4	59.1	71.4	68.1	68.4	79.2	73.8	77.2

Table II: Selection of most probable MV

Mode	**Most Probable Motion vector**
16x16 mode	Motion vector of the collocated MB in the previous frame
16x8, 8x16 and 8x8 mode	Motion vector of 16x16 mode
4x8, 8x4 and 4x4 mode	Motion vector of 8x8 mode

In order to justify this method, we have done some experiments with different video sequences at different QP factors. The most probable region is the region corresponding to most probable MV. We have calculated

the percentage of the event "most probable region is the best region" and the results are presented in Table III. Approximately 95% of MVs fall within the most probable regions.

Table III: Percentage of most probable region

Sequence	Percentage
Foreman	90.69 %
Akiyo	99.34 %
Mobile	98.30 %
Carphone	95.02 %
Claire	97.95 %
Silent	96.17 %
News	98.42 %
Container	99.38 %
Stefan	87.33 %
Average	95.84%

For high motion MB, the most probable region should be between regions 1-16, as shown in Fig. 2. The encoder starts to search from the most probable region and then moves to search region 0. The order of the remaining search regions is based on the angle between the most probable region and the other regions, which must be calculated by (3) beforehand:

$$\Delta\theta = |\theta_m - \theta_R| \quad (3)$$

for R=1,2,3 ……16, $R \neq$ most probable region

where θ_m is the direction of the most probable motion vector, and θ_R is the mean direction of region R, which is calculated as follows:

$$\theta_m = \tan^{-1}\left(\frac{MV_{my}}{MV_{mx}}\right) \quad (4)$$

$$\theta_R = \frac{\theta_{R(\max)} + \theta_{R(\min)}}{2} \quad (5)$$

Here, MV_{mx} and MV_{my} are horizontal and vertical component of the most probable motion vector found in Table II. For example, the most probable region is 3 and $\theta_m = 50^0$, followed by a search order of 3, 0, 2, 4, 1, 5, 16, 6, 15, 7, 14, 8, 13,9,12,10 and 11, which becomes slightly different when the most probable region is 0. In this case, the search starts from region 0 and the order of the remaining 16 search regions is calculated based on (3).

The process of constructing the region order for each mode of an MB in the current frame becomes an additional computation introduced by the region-based search. The value of θ_R in (5) is constant throughout the whole encoding process, and θ_R is calculated one time before encoding begins. Although the total number of modes in an MB is 7, the arc-tan operator in (4) should be calculated 3 times for each MB because MV_{mx} and MV_{my} in (4) are updated 3 times (clearly shown in Table II). Table I illustrates how about 77% of the MV for the 16x16 mode of the current MB is the same as that of the 16x16 mode for the collocated MB in the previous frame, while about 76% of the MV for the 8x8 mode is equal to that of the 16x16 mode from the current MB. Since the arc-tan operator is time consuming, we have used these observations to reduce the number of times the search order must be updated. This updated search order procedure is implemented as follows:

For each MB:

- Calculate (3) for 16x16 mode and store the search order in memory.
- Perform the 16x16 mode motion estimation.
- If the MV of the 16x16 mode does not equal that of the previous frame's 16x16 mode, update the search order for 16x8, 8x16 and 8x8 modes based on (3).
- Perform the ME for the 16x8, 8x16 and 8x8 modes.
- If the MV of the 8x8 mode does not equal that of the 16x16 mode, update the search order for the 4x8, 8x4 and 4x4 modes based on (3).
- Perform the ME for the 4x8, 8x4 and 4x4 modes.

End for

4. SIMULATION RESULTS

To evaluate the performance of the proposed method, JM 9.6 [8] reference software is used in the simulations and different types of video sequences are used as test materials. The simulation condition is: GOP structure IPPP, search range 16, RDO on, number of frames is 100 and frame rate is 30 fps. We chose seven sequences with motion activities varying from small, to large and the comparison results were tabulated in Table IV. Two different methods are compared with full search motion estimation technique. One is early termination method described in [7] and the proposed region based search method with early termination. PSNR and bit rate differences were calculated according to numerical averages between RD curves derived from the original and proposed algorithms, respectively [9]. In order to evaluate the ME time reduction, ΔT (%) is defined as follows:

$$\Delta T\% = \frac{T_{original} - T_{proposed}}{T_{original}} \times 100\% \quad (6)$$

where, $T_{original}$ denotes the ME time of the JM 9.6 encoder and $T_{proposed}$ is the ME time of the encoder in the proposed method.

Table IV: PSNR and bit rate comparison

Sequence (QCIF)	Δ PSNR in dB		$\Delta rate$ %	
	ET[7]	ET[7] + proposed	ET[7]	ET[7] + proposed
Akiyo	0.003	0.007	0.011	0.017
Claire	0.004	0.005	0.145	0.138
Carphone	0.009	0.011	0.215	0.353
Container	0.008	0.006	0.171	0.186
Foreman	0.003	0.002	0.368	0.004
Mobile	0.003	0.003	0.451	0.056
Stefan	0.016	0.025	0.598	0.566

Table V: Complexity (ΔT%) comparison

Sequence (QCIF)	ET[7]	ET[7] + proposed
Akiyo	75.62	81.76
Claire	76.38	80.34
Carphone	55.27	74.33
Container	68.17	76.44
Foreman	67.46	75.43
Mobile	62.45	73.93
Stefan	66.71	77.86
Average	67.44	77.15

Table V shows how the proposed algorithm yields 77% ME time savings on average, compared to a full search ME with a negligible PSNR reduction (0.009 db on average) and bit rate increment (0.32% on average). The RD performance of the proposed method is very closely matched with [7] but the proposed algorithm makes early termination method about 10% faster. For slow motion sequences like *Akiyo* and *Claire*, the proposed algorithm saves about 80% of ME time. The computation reduction is high because most of the motion vectors for these types of sequences are around the search center, and the encoder gets the best point after searching only a few of the points.

5. CONCLUSIONS

This paper develops a simple region based search algorithm in order to further reduce the complexity of early terminated motion estimation scheme. The entire search window was divided into 17 different regions. The search order of the region is selected based on the orientation of the highly correlated MV. The simulation results verified that the proposed technique is suitable for H.264/AVC full search motion estimations, and the proposed scheme allows most of the search points to be skipped in the early stages of the motion estimation process. The proposed method saves about 10% computation of early terminated motion estimation method.

6. ACKNOWLEDGEMENTS

This research has been supported in part by the Canada Research Chair Program, AUTO21 NCE, and the NSERC Discovery grant.

7. REFERENCES

[1] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Rec. H.264/ISO/IEC 14496-10* AVC, 2003.

[2] J. F. Yang, S. H. Chang, and C. Y. Chen, "Computation reduction for motion search in low rate video coders," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 12, pp. 948-951, Oct. 2002.

[3] Yang L., Yu K., Li J., Li S. "An effective variable block size early termination algorithm for H.264 video coding," *IEEE Trans. Circuits Syst. Video Tech..*, vol. 15, no. 6, pp.784-788, June 2005.

[4] Xiaozhong Xu and Yun He, "Improvement on fast motion estimation strategy for H.264/AVC," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 18, no. 3, pp.285-293, March-2008

[5] Esam A. Al Qarallah and Tian-Sheuan Chang, " Fast variable block size motion estimation by adaptive early termination," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 16, no. 8, pp.1021-1026, August-2006

[6] Soonjong Jin, Sang-Jun Park, and Jechang Jeong, "Adaptive fast full search Algorithm using partitioned region and optimized search order", *IEEE Transaction on Consumer Electronics*, vol. 53, no. 4, Nov. 2007, pp.-1703-1711

[7] M.G. Sarwer, T. M. Nguyen, Q.M.J. Wu, "Fast Motion estimation of H.264/AVC by adaptive early termination," *Proceedings of the 10th IASTED International Conference Signal and Image Processing* (SIP 2008), August 18-20, 2008, HI, USA, pp. 140-145

[8] http://iphome.hhi.de/suehring/tml/, JVT Reference Software version 9.6

[9] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," presented at *the 13th VCEG-M33 Meeting*, Austin, TX, April 2001.