Adaptive Multi-resolution Motion Estimation Using Texture-based Search Strategies

Jie Liu, Xianghu Ji, Chuang Zhu, Huizhu Jia, Xiaodong Xie, WenGao, *Fellow, IEEE*
Engineering Lab for Video Technology, School of EECS, Peking University
Email:{liuzimin, xhji, czhu, hzjia, xdxie, wgao}@jdl.ac.cn

***Abstract*—Motion estimation is the most complex module which contributes nearly 70% of computation resources in a hardware-based video encoder. This huge computational complexity limits the performance of HD video encoders in terms of encoding speed and power consumption. This paper presents a hardware oriented multi-resolution motion estimation algorithm using adaptive search strategies to reduce computational complexity. The spatial homogeneity and the temporal stationarity characteristics of video sequences are adaptively detected to determine search range and down-sampling rate. Homogeneous regions are detected by using Sobel edge operators and stationary regions are detected by using temporal information. These texture-based search strategies make motion estimation more concise under fixed computational complexity constraint. Additional computational cost originated by determining the search strategies can be neglected due to the simple addition and shift operations. Experimental results show that the proposed algorithm achieves better performance and reduces computation cost by 40% compared with previous works.**

***Index Terms* — Multi-resolution motion search, search strategies, homogeneity, stationarity, computational complexity constraint.**

I. INTRODUCTION

In many video coding standards, such as H.264/AVC [1], motion estimation (ME) plays a key role in the block-based hybrid coding framework. Integer motion estimation aims at reducing temporal redundancies between the current frame and the reference frame. There are some new tools such as variable block size motion estimation (VBSME), multiple reference frames for real-time motion estimation in high definition (HD) video encoder. As a result, the complexity and computation cost increase greatly.

Full-search block matching algorithm (FSBMA) [2] is widely used for hardware ME design due to its superior performance and high regularity. However, FSBMA needs lots of computation due to many candidate blocks to be matched. Many fast ME algorithm, including SEA [3] and DSA [4], were proposed to reduce high computation complexity. But these software-oriented approaches cannot be used in the hardware-based encoder. Multiresolution motion estimation algorithm (MMEA) [5]–[7] is developed with a coarse-to-fine search hierarchy. The MMEA is suitable for hardware implementation with its highly regular data flow. And, it can reduce the computational complexity by decreasing the number of computations.

However, traditional MMEA [5]–[7] only use fixed search range and the same down-sampling is applied to all the image area without discrimination. Thus, firstly, although it performs well for small and uniform motions, the resulting performance degradation is not negligible when the motion is complex. The basic idea of traditional MMEA is that potential match candidates are obtained from a large search area at the coarse level and the candidates become the search center in the lower fine levels. But for the sequences with complex texture, the MV in coarse search may be an incorrect result. It will yield search error directly passed on to the next level. If using large search range in fine level without down-sampling, these methods will have a high computational cost because the fine level contains large amount of calculation. Secondly, for flat region, since the texture in this region has similar spatial property, performance degradation caused by downsampling is negligible. Downsampled search at coarse level will be accurate in flat region. In this situation, downsampled search should be applied to reduce computational cost. Thirdly, when the current macroblock (MB) is not moving in adjacent video frames such as background, much computational complexity will be wasted since it also searches in the large search window.

In this paper, an adaptive multi-resolution motion estimation algorithm (AMMEA) by using texture-based search strategies is proposed. It makes search strategies customized for each block and reduces computational cost by 40% compared with traditional MMEA [5][6]. For different kinds of video sequences, it applies adaptive search range and down-sampling rate based on stationary and homogeneous features of current MB. The proposed algorithm use Sobel edge operators to detect homogeneous regions and stationary regions are detected by using temporal information.

The remainder of this paper is arranged as follows. Section II gives a brief introduction of the typical MMEA and its challenges. The AMMEA is proposed and explained in Section III. The experimental result are given in Section IV. Finally, conclusions are drawn in Section V.

II. TRADITIONAL MMEA AND ITS CHALLENGES

The previous MMEA [5]–[7] includes three levels, as is illustrated in Fig. 1. In level 2 (most down-sampled), the search window is the largest and centered on original point (0, 0). In level 1, four search windows are centered on three candidates selected from level 2 and a candidate selected through predicted motion estimateon (PMV), respectively. The 4:1 down-sampling is adopted in level 1. Level 0 is a fine level without data subsampling and VSBME is used in this level. The supported block size is larger than or equal to 8x8.

The MMEA only has fixed search range and the same down-sampling for all blocks of a sequence uniformly. Fig. 2

978-1-4799-1291-9/14/$31.00 ©2014 IEEE

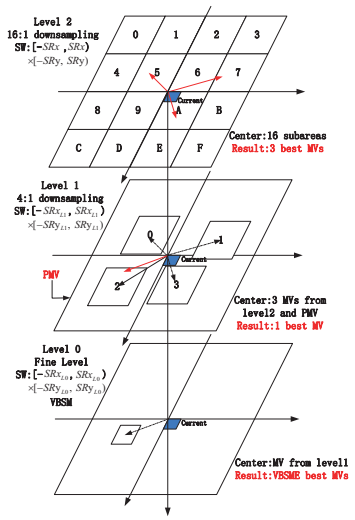

Fig. 1. Three level multi-resolution motion estimation

shows an example frame from the 720P sequence "Spincalindar". In Fig. 2(a), three best candidates after the level 2 search and PMV (0, 0) are chosen as the search center for level 1. The winner candidate of the level 1 is selected as the center for level 0 in Fig. 2(b). Fig. 2(c) shows that the final MV is achieved by using MMEA. However, the best MV should be (0, 0) according to FSBMA. For the current block with complex texture, the MV in coarse search is an incorrect result. And, it yields error match directly passed on to the next level. Therefore, performance degradation is inevitable in MMEA due to downsampling sequences with complex texture. Another problem is that the temporal stationary blocks are more likely to move around (0, 0) with a small range. But it also searches in the large search window and the much computational complexity will be wasted.

(a)Level 2 (b)Level 1 (c)Level 0

Fig. 2. The best MVs chosen after each level

III. Adaptive MMEA Using Search Strategies

In this section, we divide adaptive multi-resolution motion estimation algorithm into two parts to analyse. The first is the adaptive search strategies based on feature of block, the second is that the adaptive multi-resolution motion estimation algorithm using the search strategies.

A. Search Strategies Based on Feature of MB

In [8], it proposes a fast inter mode decision algorithm to decide the best mode in inter coding by using the spatial homogeneity and the temporal stationarity characteristics of video objects. Spatial homogeneity of a MB is based on the MB's edge intensity, and temporal stationarity is decided by the difference of the current MB and its co-located counterpart in the reference frame. In this paper, spatial homogeneity and temporal stationarity are applied to adjust search stretagies. Considering the stationary MB may move in small range, it is centered at the current MB position in the reference frame that is the origin (0, 0) and the search range(SR) will be set to be very small for reducing calculation. When MBs in the picture are considered as homogeneous blocks, performance degradation caused by downsampled searching at coarse level is negligible. Thus, it is reasonable to search in large window with down-sampling. In contrary, for nonhomogeneous blocks, if downsampling pixels are selected for final MV, the ME quality loss is inevitable. In this situation, it will abandon searching in coarse level. Then it will only search in fine level with a larger search range so as to achieve significant coding gain using VBSME.

Before the hierarchical ME process, there are two steps for stationary regions determination and homogeneous regions detection to determine search strategies. Stationary regions refer to non-moving regions in the temporal dimension. In this paper, we use stationary regions detection to determine search range. If current block is regarded as stationary region, we will only search around (0, 0) with a small search range. In natural video sequences, there are correlations between current frame and reference frame. A method for detecting stationary region is proposed using temporal information [8]. The difference between current MB and reference MB can be computed by using (1). Here C[i, j] and P[i, j] are respectively the luminance values in the current MB and reference MB. The image is 8 bit per pixel, and setting the threshold to 200 achieves good performance as suggested in [8].

$$Diff = \sum_{i=1,j=1}^{16,16} abs(C[i,j]-P[i,j]) \quad (1)$$

In addition, we also utilize homogeneous region determination to adjust search strategies. Homogeneous region refers to the regions having similar texture in the spatial domain. Edge information can represent texture complexity. According to the analysis on the texture complexity on image, homogeneous regions will be detected. As analysis in [8], there are many techniques for detecting edge information. Using the Sobel edge operators to obtain the edge information is a balance between computational expense and performance. The Sobel edge operators have two 3x3 convolution kernels to calculate approximations of the derivatives, one for horizontal changes, and another for vertical. For a pixel, in a luma picture, we define the corresponding edge vector, $\vec{D}_{i,j} = \{dx_{i,j}, dy_{i,j}\}$ as

$$dx_{i,j} = p_{i-1,j+1} + 2\times p_{i,j+1} + p_{i+1,j+1} - p_{i-1,j-1} - 2\times p_{i,j-1} - p_{i+1,j-1}$$
$$dy_{i,j} = p_{i+1,j-1} + 2\times p_{i+1,j} + p_{i+1,j+1} - p_{i-1,j-1} - 2\times p_{i-1,j} - p_{i-1,j+1} \quad (2)$$

where x and y represent the degree of difference in vertical and horizontal directions respectively. Therefore, the amplitude of the edge vector can be computed by

$$\mathrm{Amp}(\vec{D}_{i,j}) = |dx_{i,j}| + |dy_{i,j}| \quad (3)$$

The current block size is 16x16, so the homogeneity size of a block is the same. The sum of the amplitude of the edge vectors in the block is divided into three categories by two thresholds Thd1, Thd2. Three categories correspond to different search strategies. The details will be further discussed in Section III-B. The r and c are the indices of the

row and column of the block. The sum amplitude of one MB is represented as (4). The adaptive search strategies of MB(r,c) are defined as follows:

$$H(r,c) = \sum_{i,j \in N \times N} Amp(\vec{D}_{i,j}) \tag{4}$$

$$MB(r,c) = \begin{cases} \text{level 0:on, level 1:off, level 2:off,} & \text{if } H(r,c) >= Thd_1 \\ \text{level 0:on, level 1:on, level 2:off,} & \text{if } Thd_2 <= H(r,c) < Thd_1 \\ \text{level 0:on, level 1:on, level 2:on,} & \text{if } H(r,c) < Thd_2 \end{cases} \tag{5}$$

B. Adaptive Multi-Resolution Motion Estimation Algorithm

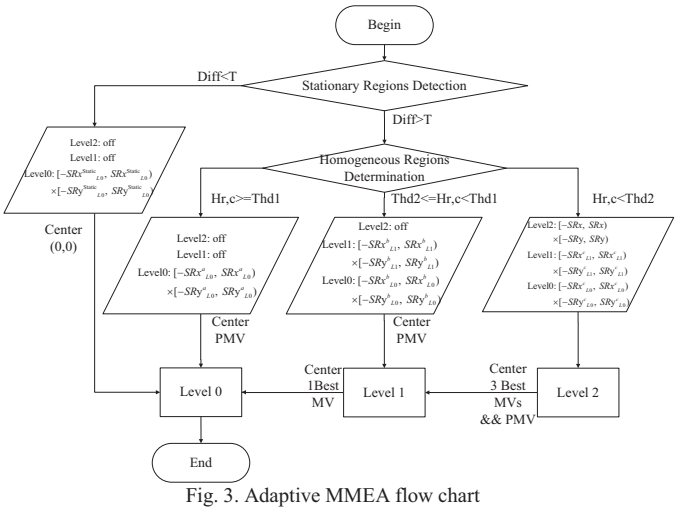

Fig. 3. Adaptive MMEA flow chart

Fig. 3. provides the algorithmic flow chart. To begin with, it will determine whether the current MB is stationary or not. If the current MB is stationary, the proposed algorithm adopts the search strategies that abandon searching at level 2 and level 1. Considering the stationary MB may move in small range, it is centered on (0, 0) and the SR is set to $[-SRx^{Static}_{L0}, SRx^{Static}_{L0}) \times [-SRy^{Static}_{L0}, SRy^{Static}_{L0})$ at level 0, as shown in Fig. 4(a). Further more, if it is not a stationary MB, we need to determine whether the current MB is homogeneous or not. As analyzed in Section III-A, there are two threshold Thd1 and Thd2 dividing the amplitude of the edge vectors into three categories (5). According to the experimental results, Thd1=20000 and Thd2=16000 will achieve good performance for all kinds of test sequences. The three categories are introduced as follows:

a) $H(r,c) >= Thd1$ The current MB is regarded as highly complex texture. In this situation, the resulting performance degradation is not negligible if downsampling is still used. Highly complex MB has a strong chance to be encoded using VBSME. VBSME is unsuited to be utilized at coarse levels. Therefore, it dosen't search at level 2 and level 1. Since the current MB is nonstationary, it is centered at PMV and the SR is set to $[-SRx^{a}_{L0}, SRx^{a}_{L0}) \times [-SRy^{a}_{L0}, SRy^{a}_{L0})$ at level 0, Fig. 4(b). This search strategies make sure that VSBME be used in finest level with larger SR.

b) $Thd2 <= H(r,c) < Thd1$ The current MB is regarded as modestly complex texture. Since level 2 is the coarest level that 16:1 downsampled from level 0, it only searches at level 1 and level 0. In level 1, it is centered at PMV and the SR is set to $[-SRx^{b}_{L1}, SRx^{b}_{L1}) \times [-SRy^{b}_{L1}, SRy^{b}_{L1})$. The 4:1 downsample is applied in this level. After the level 1 search, the MV with minimum cost is selected as the center for level 0 search window with the SR is $[-SRx^{b}_{L0}, SRx^{b}_{L0}) \times [-SRy^{b}_{L0}, SRy^{b}_{L0})$, Fig. 4(c).

c) $H(r,c) < Thd2$ The current MB is a homogeneous MB. In [8], the current MB will most probably be encoded using 16x16 mode. To a homogeneous MB, the resulting performance degradation caused by downsampling is negligible and the MV from the coarse level can be thought of relatively accurate. It is reasonable to use 16x16 mode at level 2 and level 1. In this situation, the search stategy is the same as three level MMEA [5][6] in Fig. 1. The only defference is that it is enough to search with a very smaller SR($[-SRx^{c}_{L0}, SRx^{c}_{L0}) \times [-SRy^{c}_{L0}, SRy^{c}_{L0})$) at level 0.

For different kinds of video sequences, the proposed algorithm uses flexible search range and down-sampling rate. Compared with previous works, the computational cost is reduced while maintaing a better performance. The details of the experimental result will be described in Section IV.

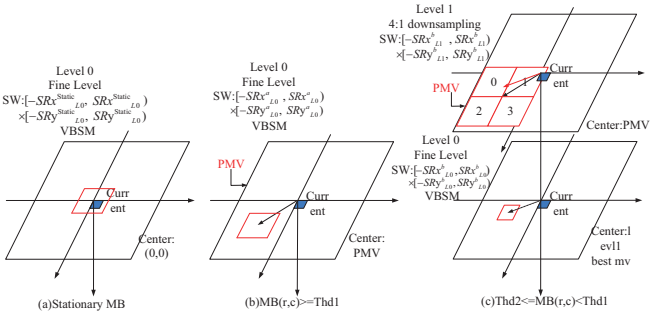

Fig. 4. Adaptive search strategies applied on MMEA

IV. Experimental Results

Our proposed algorithm was implemented to support AVS Jizhun profile. According to the specifications provided in [5][6], the test conditions are as follows: 1) Table I shows search window comparison. 2) SAD is used as the matching distortion criterion. 3) Reference frame number equals to 2. 4) VLC is enabled. 5) MV resolution is 1/4 pel. 6) GOP structure is IBBPB. 7) Inter block mode from 16x16 to 8x8. 8) The number of frames in a sequence is 32.

Table I
Search Window Comparison with Traditional MMEA

Designs	MMEA[5][6]	Proposed algorithm			
		(c)	(b)	(a)	Static
Level2	$SRx = 128$ $SRy = 96$	$SRx = 128$ $SRy = 96$	Off	Off	Off
Level1	$SRx_{L1} = 8$ $SRy_{L1} = 8$	$SRx^{c}_{L1} = 8$ $SRy^{c}_{L1} = 8$	$SRx^{b}_{L1} = 16$ $SRy^{b}_{L1} = 16$	Off	Off
Level0	$SRx_{L0} = 10$ $SRy_{L0} = 10$	$SRx^{c}_{L0} = 4$ $SRy^{c}_{L0} = 4$	$SRx^{b}_{L0} = 8$ $SRy^{b}_{L0} = 8$	$SRx^{a}_{L0} = 12$ $SRy^{a}_{L0} = 12$	$SRx^{Static}_{L0} = 12$ $SRy^{Static}_{L0} = 12$

In Table I, the search range and down-sampling rate of the proposed algorithm is observed through experimental results to achieve a better performance while reducing computation cost by 40% compared with [5][6]. Keeping fixed computation cost, AMMEA adjustable search range depends on different search strategies in Table I. According to fully re-configurable

parallel processing element (PE) array structure [5], there are totally 64×2 parallel four-pixel PEs in this architecture. The computational cost to determine search strategies is negligible. The total cycle consumption for three levels are given as follow:

$$T_{IME}=\frac{(2\times SRx/4)\times(2\times SRy/4)}{4\times 4}+(2\times SRx_{L1})\times(2\times SRy_{L1}) \quad (6)$$
$$+(2\times SRx_{L0})\times(2\times SRy_{L0})$$

Table II shows the total cycle consumption of the proposed algorithm and the comparison with previous works.

TABLE II
IMPLEMENTATOIN COST COMPARISON WITH PREVIOUS ME

Designs	Proposed	MMEA [5][6]	Huang [9]	Liu [10]	Chen [11]	Deng [12]
Video Spec.	1080P@ 30fps	1080P@ 30fps	720P@ 30fps	1080P@ 30fps	720P@ 30fps	SD@ 30fps
Ref.Number	2	2	4	1	1	1
Search Range	256×192	256×192	128×64	196×128	128×64	65×65
Number of PE	512	512	N/A	2048	128×8	16×16
Data Latency(Cycles)	576/512	848	N/A	960	1536	5216
Working Frequency(MHz)	150	220	108	200	108	260

Different sequences under different resolutions (1080P, 720P, D1, CIF) are chosen for the test. Every resolution selects some sequences with different characteristics. These features include not only complex texture and high motion, but also homogeneous region and small motion. The PSNR degradation of proposed algorithm and [5][6] compared with FSBMA are shown in Table III. The metrics is BD-PSNR using four QPs(24, 28, 32, 36). According to the last column of Table III, the proposed algorithm achieves better performance as [5][6]. From Table III, we find that the feature of these sequences with complex texture and small motion, such as "Spincalendar" and "Fireworks", have high PSNR gain. For high motion sequences, the proposed algorithm still have better performance than [5][6], such as "BasketballDrive" and "Tractor". Meanwhile, Fig. 5. shows the PSNR curves for different resolutions.

TABLE III
THE PSNR PERFORMANCE COMPARISON

Resol-ution	Sequence	Complex Texture	Simple Motion	Complex Motion	Prop-osed (dB)	[5][6] (dB)	Diff (dB)
1080P	Fireworks	√	√		-0.10	-0.17	0.07
	BasketballDrive			√	-0.06	-0.07	0.01
	Tractor			√	-0.06	-0.07	0.01
	Cactus			√	-0.01	-0.02	0.01
	MobcalYer		√		-0.00	-0.01	0.01
	Crowdrun	√	√		-0.00	-0.04	0.04
720P	Spincalendar	√	√		-0.07	-0.13	0.06
	Sheriff		√		-0.00	-0.01	0.01
	City	√	√		-0.03	-0.07	0.04
	Optis		√		-0.00	-0.01	0.01
D1	Mobilecalendar	√	√		-0.05	-0.14	0.09
	Flowergarden		√		-0.00	-0.05	0.04
CIF	Mobile	√	√		-0.00	-0.18	0.18
	Foreman		√		-0.04	-0.11	0.07
	Kiel	√		√	-0.05	-0.21	0.16
	Crew		√		-0.02	-0.07	0.05

V. CONCLUSION

In this paper, a hardware orirnted fast motion estimation algorithm is proposed by using MB's texture and stationarity characteristics to determine search strategies. The proposed algorithm and architecture of IME also can be used for H.264/AVC and HEVC. Compared to FSBMA, the proposed algorithm reduce the computational complexity with a negligible average PSNR loss of 0.03 dB. It has better performance and reduces computation cost by 40% compared with other hierachical motion search algorithm. The proposed algorithm reachs a good balance between computational complexity and performance.

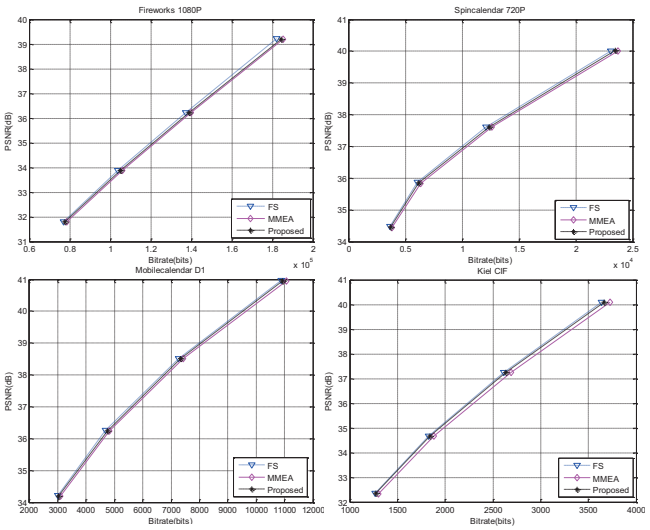

Fig. 5. The PSNR curves of the proposed algorithm versus FSBMA and MMEA

REFERENCES

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

[2] T.C. Chen et al, "Analysis and Architecture Design of an HD720p 30 Frames/s H.264/AVC Encoder," *IEEE Trans. Cir. Syst. Video Tech.*, vol. 16, no. 6, pp. 673-688, June 2006.

[3] X. Q. Gao, C. J. Duanmu, and C. R. Zou, "A multilevel successive elimination algorithm for block matching motion estimation," IEEE Trans. Image Process., vol. 9, no. 3, pp. 501 –504, Mar. 2000.

[4] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 9, no.2, pp. 287–290, Feb. 2000.

[5] X. H. Ji, C. Zhu, H. Z. Jia, et al, "A Hardware-Efficient Architecture for Multi-Resolution Motion Estimation Using Fully Reconfigurable Processing Element Array,"in Proc. ICME, Jul. 2011, pp. 1–6.

[6] H. B. Yin, H. Z. Jia, H. G. Qi, X. H. Ji, et al, "A Hardware-Efficient Multi-Resolution Block Matching Algorithm and Its VLSI Architecture for High Definition MPEG-Like Video Encoders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1242–1254, Sept. 2010.

[7] X. Bao, D. Zhou, and S. Goto, "An advanced hierarchical motion estimation scheme with lossless frame recompression for ultra high definition video coding," in Proc. ICME, Jul. 2010, pp. 820–825.

[8] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, andC. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 953–958, Jul.2005.

[9] Y.-W. Huang, T.-C. Chen, et al, "A 1.3 TOPS H.264/AVC single-chip encoder for HDTV applications," in IEEE ISSCC Dig.Tech. Papers, pp128-129, Feb.2005.

[10] Z.Y. Liu, Y. Song, M. Shao, et al, "HDTV 1080P H.264/AVC encoder chip design and performance analysis," *IEEE J. Solid-State Circuits*, vol. 44, no. 2, 816 pp. 594–608, Feb. 2009.

[11] T.-C. Chen, S.-Y. Chien; Y.-W. Huang, et al, "Analysis and architecture design of an HDTV720p 30 frames/s H.264/AVC encoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 673–688, Jun. 2006.

[12] L. Deng, W. Gao, M. Z. Hu, Z. Z. Ji, "An efficient hardware implementation for motion estimation of AVC standard," *IEEE Trans. Consumer Electron.*, vol. 51, no. 4, pp. 1360-1366, Nov. 2005.