A FAST MOTION ESTIMATION ALGORITHM BASED ON MULTI-RESOLUTION FRAME STRUCTURE

Byung Cheol Song and Jong Beom Ra

Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
373-1 Kusongdong, Yusonggu, Taejon, Republic of Korea

ABSTRACT

We present a novel multi-resolution block matching algorithm (BMA) for fast motion estimation. At the coarsest level, a full search BMA (FSBMA) is performed for searching complex or random motion. Concurrently, spatial correlation of motion vector (MV) field is used for searching continuous motion. Here we present an efficient method for searching full resolution MVs without MV decimation even at the coarsest level. After the coarsest level search, two or three initial MV candidates are chosen for the next level. At the further levels, the MV candidates are refined within much smaller search areas. Simulation results show that in comparison with FSBMA, the proposed BMA achieves a speed-up factor over 710 with minor PSNR degradation of 0.2dB at most, under a normal MPEG2 coding environment. Furthermore, our scheme is also suitable for hardware implementation due to regular data-flow.

1. INTRODUCTION

The most popular method in eliminating temporal redundancy among successive video frames is motion compensated predictive coding, which is currently employed in video coding standards such as MPEG [1] and H.26x [2]. BMA is commonly used in video coding systems. Its goal is to find a block that is most similar to a current block within a pre-defined search area in a reference frame. To achieve this, FSBMA exhaustively evaluates all possible candidate blocks. Therefore, FSBMA is computationally intensive, and may not be practical for high-quality video coding applications such as MPEG2 or HDTV, where the search area should be sufficiently large so as to improve coding efficiency. As a result, many fast BMAs [3-11] have been developed to alleviate heavy computation requirements of FSBMA.

Fast BMAs can be categorized into a few groups according to their characteristics: BMA with unimodal error surface assumption such as LOGS and TSS [3-4], BMA based on pixel sub-sampling [5], hierarchical or multi-resolution BMA [6-8], and BMA using temporal/spatial correlation [9-10]. The first group of BMAs can be trapped in a local minimum with a large amount of matching errors because their basic assumption (unimodal error surface) rarely holds in practical video sequences. The second group of BMAs has a limitation in computation reduction, because excessive sub-sampling may hasten a local minimum phenomenon. The third group of BMAs works relatively well and provides fast computation. Usually, it only uses the information from coarser levels for MV refinement at finer levels without considering other useful information such as spatial/temporal correlations. So these BMAs do not always provide reliable performance. Finally, the fourth group of BMAs is based on a predictive search. It not only reduces its computational complexity by using spatial/temporal correlations, but also provides reliable performance for the video sequences having highly correlated motions. However, since the schemes assume that MV field always has strong correlation, their performance degradation becomes serious for video sequences where this assumption is not true. Moreover, due to their irregular data-flow, they are not suitable for hardware implementation.

To overcome drawbacks of the existing fast BMAs, we propose a new multi-resolution BMA using spatial correlation in MV field. In order to maintain data-flow regularity in computing sum-of-absolute-difference (SAD) at each level, an efficient SAD prediction method is presented. That is to say, instead of SAD of the current block, SAD of its shifted block is sometimes used. This brings two advantages: data-flow regularity of the reference frame and use of MV candidates without decimation at coarser levels. The accuracy of the proposed SAD prediction will be proven via simulation in Section 3. Furthermore, MV estimation at each level can be performed in integer pixel unit because of the efficiency of the SAD prediction method. So final MV refinement at full resolution level may be skipped without performance degradation. This provides additional complexity reduction. Therefore, the proposed BMA guarantees outstanding performance as well as high computational speed. In addition, it is suitable for hardware implementation due to its simplicity and data-flow regularity.

This paper is organized as follows: We propose a fast multi-resolution BMA in Section 2. Then simulation results are provided in Section 3. Conclusions are given in Section 4.

2. A FAST MULTI-RESOLUTION BLOCK MATCHING ALGORITHM (MRBMA)

For the last several years, the multi-resolution schemes have been actively studied for various application areas including motion estimation. In conventional multi-resolution BMAs schemes, each frame is decomposed into a few resolutions. The multi-resolution nature provides hierarchical MV field obtained at different resolutions. Then, several MV candidates are estimated at the coarsest level, and they are refined in the following finer levels based on MV information obtained at coarser levels. Recently, a fast BMA combining multi-resolution, spatial, and temporal correlation properties, has been presented [8]. It provides relatively good performance in comparison with other fast BMAs. Nonetheless, there are two drawbacks in this

BMA. It is not suitable for hardware implementation due to its irregular data-flow. Also, it shows significant performance degradation for video sequences having complex or random motion. This is because temporally adjacent MVs as MV candidates are used after decimation at each coarser level. In general, MV precision has a major effect on prediction performance.

In this section, we propose a novel multi-resolution BMA called MRBMA, which provides high computational speed and high performance concurrently. Furthermore, it is suitable for hardware implementation.

2.1 Frame Structure for MRBMA

Only the frame structure for frame-based motion estimation will be described hereafter because the same rule can be applied to field-based motion estimation. We assume that L is the number of levels; and the level numbers are ordered from 0 to L-1, where levels 0 and L-1 represent the coarsest and finest levels, respectively.

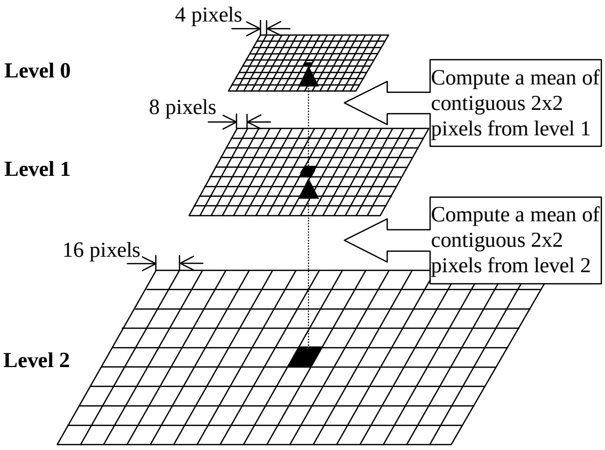


Figure 1. Multi-resolution frame structure for the reference frame (L=3). Each rectangle denotes a macroblock (MB).

MRBMA adopts a multi-resolution frame structure for the reference frame. Each pixel of its coarser level image is obtained by computing a mean of contiguous 2x2 pixels from its previous finer level image as depicted in Fig. 1.

MRBMA adopts a block mean pyramid [11] as the frame structure for the current frame (see Fig. 2). Unlike the multi-resolution frame structure, the frame and MB size of the block mean pyramid are constant at each level, while a pixel value is obtained similarly.

Note that the computational overhead for the current frame structure consists of 3 additions and 1 shift per pixel, which are approximately equivalent to 2 matching operations. So this overhead is minor in comparison with the total complexity of motion estimation. On the other hand, the complexity to compose the multi-resolution frame structure for the reference frame is much smaller than this.

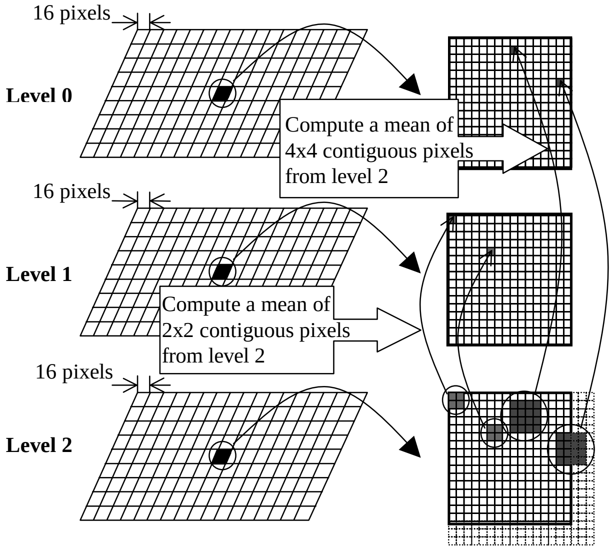


Figure 2. Block mean pyramid frame structure for the current frame (L=3). Each rectangle denotes a MB.

2.2 A Framework of MRBMA

In MRMBA, a full search is performed at level 0 as in the multi-resolution BMA of [8]. From the perspective of level L-1, the full search at level 0 is equivalent to examining regularly sub-sampled MVs within a whole search area Ω_S ($(2N_S+1)\mathrm{x}(2N_S+1)$), which is usually useful for searching random or complex motion. During the full search at level 0, four MVs (V_m^{cn}'s for $1 \le m \le 4$) of causally neighbors of the current MB $B_{(i,j)}$, whose upper-left corner has the coordinate (i, j), are checked as possible initial MV candidates (see Fig. 3).

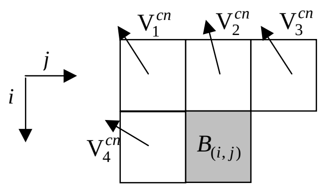


Figure 3. Geometry of causally adjacent MBs.

They are useful for searching continuous motion [10]. One salient feature of MRBMA is examining the four V_m^{cn}'s at level 0 without MV decimation. This is possible because MRBMA adopts the above-mentioned frame structures. Note that in the conventional multi-resolution BMA, it is impossible to examine the V_m^{cn}'s at level 0 without MV decimation, or the V_m^{cn}'s will become meaningless if the decimation is performed at level 0 [8].

In order to predict the SAD of $B_{(i,j)}$ corresponding to V_m^{cn}, MRBMA employs the SAD of $B_{(i+\alpha,j+\beta)}$ somewhat shifted from $B_{(i,j)}$ (see Fig. 4). In Fig. 4, pixel coordinates in the reference frame are described from the perspective of the current frame. For example, if the coordinate of a V_m^{cn} is (p, q), α and β are defined at level l as

$$\alpha = \left\lceil \frac{p}{2^{L-l-1}} \right\rceil \cdot 2^{L-l-1} - p, \quad \beta = \left\lceil \frac{q}{2^{L-l-1}} \right\rceil \cdot 2^{L-l-1} - q \qquad (1)$$

In (1), $\lceil A \rceil$ denotes the smallest integer that is larger than or equal to A.

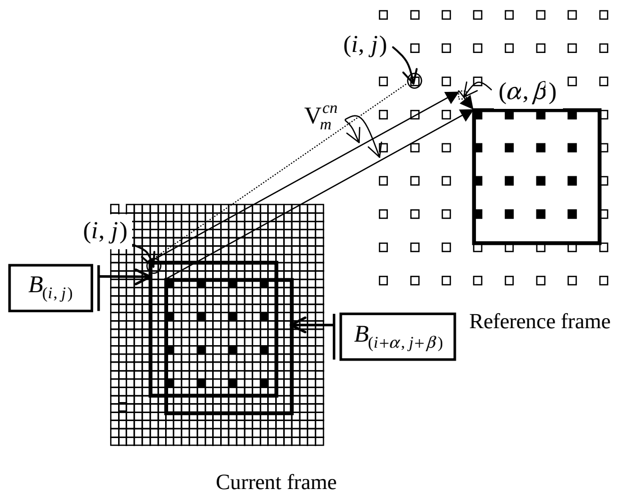


Figure 4. A V_m^{cn} search at level 0 in case of L=3. Black colored pixels are to be matched each other in SAD

The proposed SAD prediction is much more accurate than the SAD prediction by MV decimation because MV precision is guaranteed and most pixels within $B_{(i,j)}$ are involved in $B_{(i+\alpha,j+\beta)}$ due to relatively small α and β (i.e., $0 \le \alpha, \beta \le 3$). The superiority of this prediction will be proven via simulation in Section 3.

During the full search, SAD computations corresponding to the MVs decimated from the V_m^{cn}'s are skipped. Instead, the proposed SAD predictions are performed at those locations. This is because a V_m^{cn} has a higher probability of being a global minimum than does its decimated MV candidate and irregular data-flow of the reference frame can be avoided. Therefore, the proposed SAD prediction scheme provides regular data-flow of the reference frame as well as good prediction performance.

After we perform the full search at level 0, we adopt multiple-candidates strategy at level 1. However, too many MV candidates at each level make hardware implementation difficult along with increase in complexity. So we choose two MV candidates having minimum SADs at level 0. The first candidate $\mathrm{V}_1^{\mathrm{L0}}$ is chosen among the four V_m^{cn}'s and the second one $\mathrm{V}_2^{\mathrm{L0}}$ among the remaining MV candidates. Here superscipt 0 means level 0. Finally, after local searches around the two initial estimates, one MV candidate V^{L1} for a local search at level 2 is found. Here local search area Ω_{L1} is set to $(2N_{\mathrm{L1}}+1)\times(2N_{\mathrm{L1}}+1)$ (3x3 or 5x5 in our experiment), which is dramatically small compared to Ω_{S}. Note that every level search of MRBMA can be performed in integer pixel unit by using the proposed SAD prediction method, while existing multi-resolution BMAs can do local search in integer pixel unit only at level L-1. This is another salient feature of MRBMA. Therefore, a local search at level L-1 often gives a minor effect on overall performance of MRBMA. So we can skip the local search at level L-1 without performance degradation. Also, the skipping results in additional complexity reduction. Simulation results in Section 3 will support this fact.

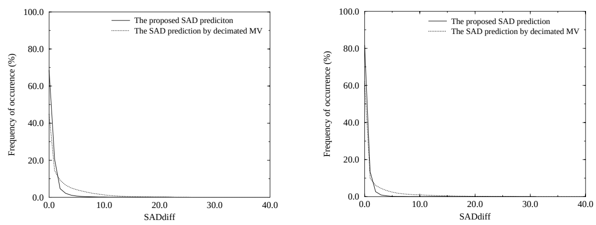
At level 2, we find a final MV V^{L2} by doing a local search around V^{L1}. Here local search area Ω_{L2} is set to $(2N_{\mathrm{L2}}+1)\times(2N_{\mathrm{L2}}+1)$ (3x3 in our experiment). As mentioned previously, this final local search can be skipped.

3. SIMULATION RESULTS

For the experiment, the first 100 frames of four MPEG2 video sequences are used; “cheer leaders (cheer),” “car (car),” “football (foot),” “flower garden (flower).” Each frame has a resolution of 720 pixels x 480 lines as in the CCIR 601 format. Let MRBMA-*mLnS* denote an *n*-step MRBMA with *m* level. Normally *n* can be set to *m* or *m*-1. For example, MRBMA-3*L*2*S* denotes an MRBMA without final local search in case of *L*=3.

3.1 The accuracy of the proposed SAD prediction

In order to test the performance of the SAD prediction scheme proposed in Section 2, we compute predicted SADs and compare them with the exact ones. First, a full search is performed at level 2 to obtain the exact MV of each MB. Then two types of SADs are computed at level 0; one by the proposed SAD prediction scheme and the other corresponding to the MV decimated from a V_m^{cn}. Finally, we compare the two SADs with the exact SAD at level 0. Fig. 5(a) provides the experimental result. In Fig. 5, $\mathrm{SAD}_{\mathrm{diff}}$ denotes the absolute difference between two corresponding SADs, which is divided by MB size of 256. To show the result more clearly, the graphs are displayed only below $\mathrm{SAD}_{\mathrm{diff}}$ of 40, not 255.



(a) (b)

Figure 5. Accuracy test of the proposed SAD prediction scheme in case of *L*=3 (a) at level 0 and (b) at level 1.

We can observe that the SAD obtained by the proposed prediction scheme is more accurate than the one from the decimated MV. Fig. 5(b) depicts the result at level 1. As shown in graphs, the tendency is similar regardless of the level. From these results, we can confirm that the proposed SAD prediction is very reliable and good enough to use for MV estimation.

3.2 Performance evaluation of MRBMA based on an MPEG2 video encoder

We select parameters of N = 12 (GOP size), M = 2 (frame distance between P-frames), and T = 6 Mbps (target bit rate). Ω_S is commonly set to 127 x 127 pixels for P-frames and 63 x 63 pixels for B-frames. The performance of MRBMA is evaluated in terms of average peak-to-peak signal-to-ratio ($PSNR_{avg}$). Performance comparisons in terms of $PSNR_{avg}$ are given in Table I. MRBMA provides almost the same $PSNR_{avg}$ as that of FSBMA. It is especially noticeable that regardless of local search areas, MRBMA gives no PSNR degradation for flower sequence.

In Table II, we compare computational complexity of MRBMA with that of FSBMA for various search areas. We examine the number of operations per MB as a complexity measure. Then the speed-up factor is given as the ratio of the complexity of FSBMA to that of each scheme.

From the tables, it clearly shows that MRBMA always provides almost the same performance as that of FSBMA with considerable complexity reduction. And also, it provides consistent performance throughout all test sequences.

Table I. Average PSNR comparison between MRBMA and FSBMA in an MPEG2 video encoder.

ALGORITHM		SEQUENCES			
		car	cheer	foot	flower
FSBMA		39.0	31.3	33.7	31.8
MRBMA	3L3S (N_{L1}=2, N_{L2}=1)	38.9	31.2	33.6	31.8
	3L3S (N_{L1}=N_{L2}=1)	38.9	31.2	33.6	31.8
	3L2S (N_{L1}=2)	38.8	31.2	33.6	31.8
	3L2S (N_{L1}=1)	38.8	31.2	33.5	31.8
	4L3S (N_{L1}=N_{L2}=1)	38.8	31.2	33.5	31.8

Table II. Comparison of computational complexity in terms of speed-up factor.

MRBMA	SPEED-UP FACTOR				
	Ns= 7	Ns= 15	Ns= 31	Ns= 63	Ns= 127
3L3S (N_{L1}=2, N_{S2}=1)	7	27	83	169	227
3L3S (N_{L1}=N_{S2}=1)	11	40	111	194	237
3L2S (N_{L1}=2)	11	43	116	197	238
3L2S (N_{L1}=1)	29	92	178	231	249
4L3S (N_{L1}=N_{L1} =1)	24	101	387	1222	2585

4. CONCLUSIONS

We propose a new fast multi-resolution BMA (MRBMA) that maintains good performance with very high computational speed even for large search areas. Via computer simulation, the performance evaluation is done by applying MRBMA to an MPEG2 video encoder, with a large search area of 127x127 pixels in case of P-frames. For instance, MRBMA-4L3S retains outstanding performance with little PSNR degradation of 0.2dB at most. Its computational complexity is no more than 0.14% of that of FSBMA. Since MRBMA has a relatively simple structure, further reduction of computation can be achieved by incorporating a thresholding technique [8] in each level search, if a small amount of performance degradation is tolerable. Therefore, MRBMA can be very useful for real-time video encoding applications. It should also be noted that since MRBMA can maintain regular data-flow through the entire search procedure, it is suitable for hardware implementation.

5. REFERENCES

[1] MPEG2, “Information technology-generic coding of moving pictures and associated audio,” Tech. Rep., ISO/IEC 13818-2, Committee Draft, Mar. 1994.

[2] CCITT Study group XV, “Draft revision of recommendation H.261-Video codec for audio visual services at px64 kbps,” Temporary Document 5-E, July 1990.

[3] J. R. Jain and A. K. Jain, “Displacement measurement and its application in inter-frame image coding,” *IEEE Trans. Commun.*, vol. COM-29, pp. 730-741, 1981.

[4] T. Koga, K. Iinuma, A. Hirano, Y, Iijima, and T. Ishiguro, “Motion compensated inter-frame coding for video conferencing,” *Proc. Nat. Telecommunication Conf.*, pp. G5.3.1-5.3.5, Nov. 29-Dec. 3, 1981.

[5] B. Liu and A. Zaccarin, “New fast algorithms for the estimation of block motion vectors,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 3, pp. 148-157, Apr. 1993.

[6] M. Bierling, “Displacement estimation by hierarchical block matching,” *Proc. SPIE Visual Communications and Image Processing ’88*, vol. 1001, pp. 942-951.

[7] K. M. Nam, J. S. Kim, R. H. Park, and Y. S. Shim, “A fast hierarchical motion vector estimation algorithm using mean pyramid,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 5, no. 4, pp. 344-351, Aug. 1995.

[8] J. Chalidabhongse and C.-C. J. Kuo, “Fast motion vector estimation using multiresolution-spatio-temporal correlations,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 7, no. 3, pp. 477-488, June 1997.

[9] S. Zafar, Y. Q. Zhang, and J. S. Baras, “Predictive block matching motion estimation for TV coding— part I: Inter-block prediction,” *IEEE Trans. Broadcasting*, vol. 37, no. 3, Sep. 1991.

[10] B. C. Song and J. B. Ra, "Hierarchical block matching algorithm using partial distortion criterion," *Proc. SPIE Visual Communications and Image Processing*, vol. 3309, 1998.

[11] W. Li and E. Salari, “Successive elimination algorithm for motion estimation,” *IEEE Trans. Image Processing*, vol. 4, no.1, Jan. 1995.