An Improved Variable-Size Block-Matching Algorithm with Minimal Error

Haifeng Wang, Hanqing Lu, Qinshan Liu

National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100080
Email: {hfwang, luhq, qsliu}@nlpr.ia.ac.cn

Abstract

In this paper, we proposed an improved "bottom-up" variable-size block matching (VSBM) method. The difference from most VSBM is that no threshold is needed for each frame in the proposed method, and we just keep all the motion vectors that result in the minimum matching error. A mode prediction method is put forward to speed up the motion estimation procedure. Experiments show its encouraging performance.

Keywords: VSBM, Threshold, Motion Estimation, Mode Prediction

1. Introduction

Motion Estimation (ME) and Motion Compensation (MC) are the critical components and also the most time consuming parts in video coding systems. Especially when multi prediction modes, multi reference frames and higher motion vector resolution are adopted, such as in H.264, the ME component can consume up to 60% with one reference frame, and up to 80% with five reference frames of the total H.264 encoding time [1]. In H.264, the exhaustive-search method is performed for ME with all the block-sizes defined in the standard, and the best one which minimizes the cost function is chosen[2]. This exhaustive method slows down the encoding procedure.

In order to reduce the complexity of VSBM, Chan[3] proposed a "top-down" approach, in which large blocks are matched first. If the SAD (Sum of Absolute Difference) of the larger blocks is larger than a predefined threshold, then these blocks are split into four smaller blocks. This process is repeated until the maximum number of blocks, or minimum errors are obtained. Finally, a remerging process is conducted to remerge those blocks that can not reduce SAD or improve image quality. In this "top-down" method, ME is re-done for four sub-blocks when a larger block was split into four sub-blocks, which makes the search procedure more complicated, even though some predict methods can fast the procedure.

Rhee[4] proposed a "bottom-up" VSBM algorithm, in which full-search is first done on minimum blocks. If the SAD of a block is smaller than a threshold, its corresponding motion vector (MV) is reserved in a candidate MV set. The neighbor blocks are merged or not depending on if they have the same MVs in the candidate MV sets. So the threshold is a very important parameter to this method. As we can see, the threshold determines which vectors are include+ed in the initial sets, which further determine the prediction precision. In Rhee's algorithm, the threshold is calculated using an iterative technique on a frame by frame base, which increases the computational complexity. Moreover, in real video sequences, the motion details of different parts in a frame are different, so if we use the threshold calculated on a frame base to determine whether or not the MVs of all blocks should be included in the initial sets, the prediction precision will be reduced because of the "majority effect". Tu[2] presented to calculate the threshold adaptively according to the information of the motion estimation, quantization parameter and rate distortion cost, but this method is also frame-based, so its ME process is also complicated.

In this paper, an improved VSBM method is proposed for motion estimation. In the proposed method, we do not need a threshold to determine which MV should stay in the candidate set. However, for each minimum block, we only keep MVs which result in the minimum SAD, so we can get the best predict precision while decrease the number of MVs using the advantage of VSBM. We use a macro-mode prediction method to further improve the performance of VSBM but still keep the full-search matching precision.

The paper is arranged as follows: In Section 2, we first introduce the theory of the "bottom-up" VSBM algorithm, and then propose some modifications to improve the performance of the VSBM. In Section 3, experimental comparison between the proposed method and Rhee's method are reported. Finally, conclusions are given in Section 4.

2. Improved VSBM Using Prediction

The heuristic algorithm proposed in [2] was based on the observation that the true motion cannot be obtained just using only local information for a moving object. Especially when motion estimation is performed by using small size block, such as the minimum 4×4 block adopted in H.264, the matching results may be very close to the global minimum, while not to the true motion of the object. However, the true motion of the moving object can be obtained by choosing the common "candidate" motion vectors in the neighboring blocks. A threshold, which is calculated on a frame by frame basis, is used to determine which vectors should be included in the candidate sets. As we all know, the motion levels of different parts in a frame are different. If the whole scene of one frame seems relatively stationary, the threshold will be large, for it is proportional to the minimum mean absolute matched error of the entire frame[4]. Thus, the number of candidate MVs is increased not only for the stationary blocks, but also for the blocks with detail motions, so blocks with detail motion information are more likely to be merged, which will decrease the prediction precision. Some researchers[5,6] proposed to use the Rate-Distortion optimization methods to get a tradeoff between rate and distortion, that is, making distortion minimum under a certain rate and vice versa.

In this paper, we propose a method of VSBM. We seek for the minimum error using VSBM, so we do not need a threshold to determine which motion vector should be included in the initial set, but contain all the motion vectors resulting in the minimum prediction error. This method is equivalent to set the minimum prediction error of each block as the threshold. After we get the initial sets for all mini-blocks, a merging process begins. Figure 1 shows the partition of a macro-block defined in H.264 draft.

a	b
c	d

Fig.1: The partition of macro-block (or sub-block)

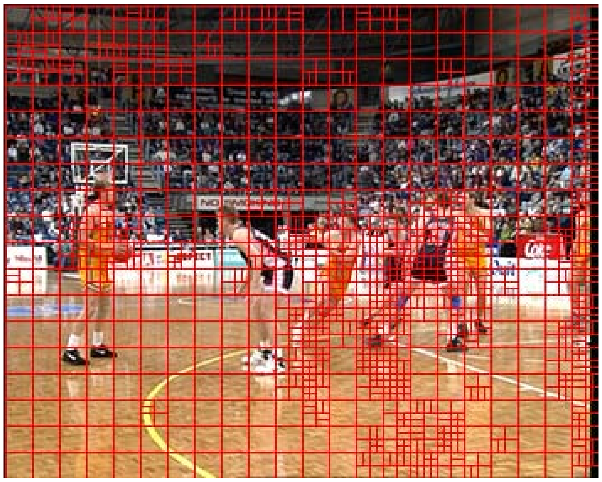
The merging procedure performed is the same as that used in [2]. As shown in fig.1, if block "a" can merge with b and block c can merge with d, or if block a can merge with c, and block b can merge with d, then rectangle block mode is used. If the four blocks can be merged together, then a larger square block mode is used. Different from [2], we use the minimum 4×4 size blocks in H.264 as the initial block size, not 8×8 size blocks.

Most of the VSBM algorithms suffer from intensive computational load, which is resulted from a bottom-up style motion vectors field building[7]. In this paper, we propose a macro-block mode prediction to reduce the computational complexity. It is intuitive that in VSBM, larger or macro blocks are used to represent those stationary parts or parts with global motion, while smaller size blocks are used to represent parts with detail motion information. In our experiments, we found that these macro-blocks show great correspondence, while smaller or minimum blocks may change a lot according to the content of the video. You can see this from Fig.2, in which the macro-block mode shows great correspondence.

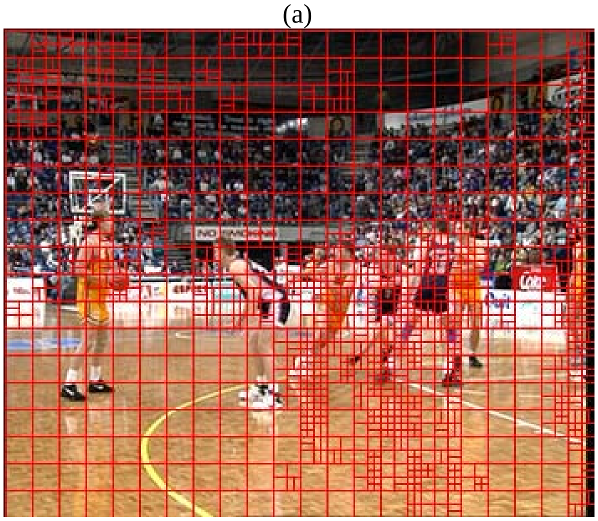
So in our mode prediction procedure, we only predict those macro-block mode, such as 16×16 size block in H.264, from the previous frame. For every macro-block in current frame, if the mode of the corresponding block in the previous frame is macro-block mode, we use (MV_x, MV_y) and (0,0) as the predict MVs, and the average of SAD of that block is used as the threshold to determine whether the current block should use one of the predict MVs or have to undergo the whole VSBM process. When dealing with macro-blocks, there is a problem of "majority effect", that is, the local information is obscured by the average effect of the large size block. So we split the macro-block into four sub-blocks when applying prediction. If the average prediction error of every sub-block is smaller than that of the corresponding macro-block in the previous frame, then prediction mode is selected, otherwise the VSBM process is directly performed.

The improved VSBM algorithm in this paper can be summarized as follows:

1. If the mode of the corresponding macro-block in the previous frame is macro-block mode, go to 2; otherwise go to 3.
2. Split the current macro-block into four sub-blocks, and get the average of the SAD calculated with motion vector (MV_x, MV_y). Denote the current average of the SAD as SAD_gc and the SAD of macro-block in the previous frame as SAD_pre. For each sub-block, if SAD_gc < SAD_pre, (MV_x, MV_y) is set as the MV of current block and go to 4. Otherwise use the data predicted with (0,0) to update SAD_gc. If SAD_gc is still smaller than SAD_pre, (0,0) is set as the MV of current block and go to 4, else go to 3.
3. Perform the "bottom-up" VSBM.
4. Deal with the next macro-block.



(a)

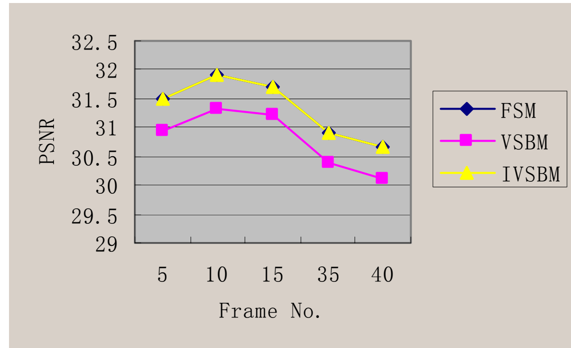


(b)

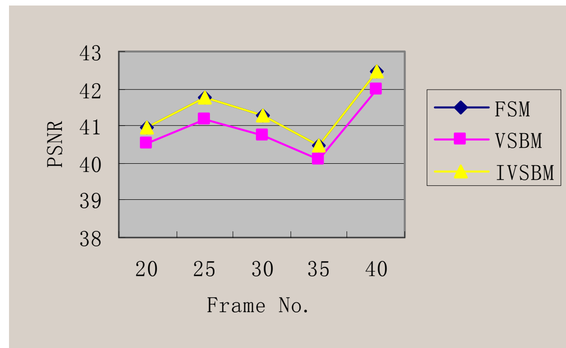
Fig.2: The mode selection results of applying "bottom-up" VSBM algorithm on basketball.yuv (CIF). (a). the result of predicting the 7^{th} frame from the 5^{th} frame; (b). the result of predicting the 9^{th} frame from the 7^{th} frame.

3. Simulation Results and Analysis

The simulation program is written in C++ and the source code is complied on Visual C++ 6.0 platform. Two sequences, i.e., basketball.yuv and akiyo.yuv, with CIF format are used to compare the proposed method with FSM (Fixed-Size Matching) and VSBM[4] and the improved VSBM we proposed in this paper. In the following, we denote the proposed method as IVSBM for simplicity. The search range we used is 7, only to speed up the test period.



(a)



(b)

Fig.3: PSNR comparison: FSM, VSBM and IVSBM (a). basketball sequence (b). akiyo sequence

We use the original *n-2*-th frame as the reference frame of the *n*-th frame, not the frame reconstructed after transformation, quantization and inverse transformation and quantization. The PSNR results are shown in Fig.3. We see that IVSBM can achieve exactly the same PSNR performance as FSM, which is 0.5db above Rhee's "bottom-up" VSBM. In some frame, the IVSBM algorithm can even achieve a bit better prediction precision than FSM algorithm. The splitting macro-block into four sub-blocks mode prediction method prevents the displacement error from increasing.

Fig.4 shows the number of blocks relationship between the three methods, and we can see that for CIF sequences, the number of blocks needed in IVSBM algorithm is 600 to 1200 more than that of VSBM.

The speed performance depends on the content of the video. If mode prediction is used, we only need to perform 4 to 8 times SAD calculation comparing with 225 times(when the search range is set as 7 pixels).

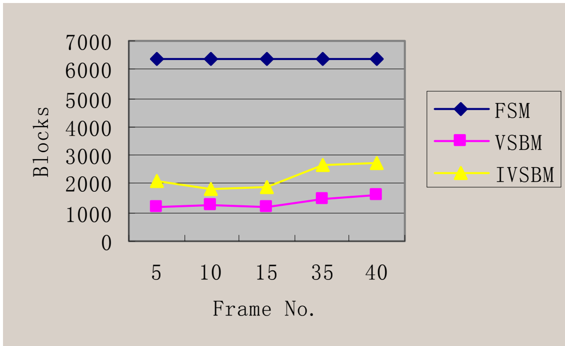


Fig.4 The number of blocks relationship. The test sequence is basketball.

4. Conclusions

In this paper, we proposed an improved VSBM algorithm. For each block, we only keep those motion vectors with minimum prediction error, and we do not need a threshold to determine which vector should be contained in the candidate set. In a sense, this method is equivalent to using the minimum matching error of each block as the threshold. To speed up the motion estimation procedure, we used a mode prediction method for macro-blocks, which expedites the process without introducing any loss of prediction precision.

5. References

[1] Zhibo Chen, Peng Zhou, Yun He, “Fast Integer Pel and Fractional Pel Motion estimation in for JVT”, JVT-F017r1.doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 6th meeting, Awaji, Island, JP, 5-13 December, 2002.

[2] Yu-Kuang Tu, Jar-Ferr Yang, Yi-Nung Shen, Ming-Ting Sun, “Fast Variable-size Block Motion Estimation Using Merging Procedure With an Adaptive Threshold”, Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on , Volume: 2 , 6-9 July 2003 Pages:II 789-92 vol.2

[3] M. Chan, Y. Yu, and A. Constantinides, “Variable size block matching motion compensation with applications to video coding,” Proc. Inst. Elect. Eng., pt. I, vol. 137, no. 4, pp. 205–212, Aug. 1990.

[4] I. Rhee, G. Marlin, S . Muthukrishnan, and R.Packwood, “Quadtree-Structured Variable-Size Block-Matching Motion Estimation with Minimal Error,”IEEE Trans. Circuits Syst. Video Technol., vol.lO, pp.42-50, Feb. 2000.

[5] Sullivan, G.J., Baker, R.L., “Rate-distortion optimized motion compensation for video compression using fixed or variable size blocks”, Global Telecommunications Conference, 1991. GLOBECOM '91. Countdown to the New Millennium. Featuring a Mini-Theme on: Personal Communications Services, pp.85-90, vol.1,2-5 Dec 1991.

[6] Yan Yang; Hemami, S.S., “Rate-distortion-based combined motion estimation and segmentation”, ICIP 98. Proceedings. 1998 International Conference on, 4-7 Oct. 1998.

[7] Ze Wang, Na Lin, Guangxing Wang, “Constrained Gowing Based Variable Size Block Motion Estimation”, Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004 International Conference on , Volume: 2 , 27-29 Pages:910-913, June 2004.