A New Motion Vector Composition Algorithm for H.264 Multiple Reference Frame Motion Estimation

Tsz-Kwan Lee, Yui-Lam Chan, Chang-Hong Fu, and Wan-Chi Siu
Centre for Signal Processing, Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
E-mail: {glorylee.eie, enylchan, enchfu, and enwcsiu} @polyu.edu.hk

***Abstract* — Multiple reference frame motion estimation (MRF-ME) is an important tool in H.264 to improve coding efficiency. However, it penalizes the encoder in computational complexity. When the number of reference frames increases, the required computation expands proportionally. Therefore, various motion vector composition algorithms have been proposed to reduce the computational complexity of the encoder. However, it is found that they only perform well in a limited range of reference frames. In this paper, a new composition algorithm is proposed to compose a resultant motion vector from a set of candidate motion vectors. The proposed algorithm is especially suited for temporally remote reference frames in MRF-ME. Compared with existing algorithms, experimental results show that the new algorithm can deliver a remarkable improvement on the rate-distortion performance.**

I. INTRODUCTION

The H.264 standard has currently dominated in the video coding standardization community for the past several years [1]. It has enhanced the compression performance up to 50% compared to the MPEG 4 standard. The gain is introduced by some of its new coding techniques such as the variable block-size motion compensation, quarter-sample accuracy for motion compensation, multiple reference frame motion estimation (MRF-ME), etc. However, these tools also bring the standard with enormous computational complexity, especially when MRF-ME is employed.

MRF-ME is allowed to search multiple reference frames in H.264 in order to achieve more accurate prediction and higher compression efficiency [2]. In MRF-ME, every reference frame is required to undergo full search motion estimation, as shown in Fig. 1. The computational complexity is thus highly increased and proportional to the number of searched reference frames. For example, if the number of searched reference frames is 5, five times of the ME processes for the current frame are required. The more number of reference frames the encoder uses, the more demanding complexity it needs. Therefore, an efficient algorithm for MRF-ME is essential in the H.264 encoder.

To reduce the computational complexity in MRF-ME, some motion vector composition (MV composition) techniques have been introduced [3-6]. Reference [3] has adopted forward dominant vector selection (FDVS) [4], which is considered as one of the best methods in video transcoding, in MRF-ME. Its computational complexity can be greatly reduced by reusing motion vectors (MVs). In [5], the median MV in neighboring blocks has also been suggested in MV composition. The algorithm in [6] has further used a weighted average on MVs after MV composition by FDVS. However, these MV composition techniques do not work well when the reference frame is distant from the current frame since the composition of new MVs may no longer represent the contents of the current macroblock (MB). In this case, the quality of the encoded videos will deteriorate.

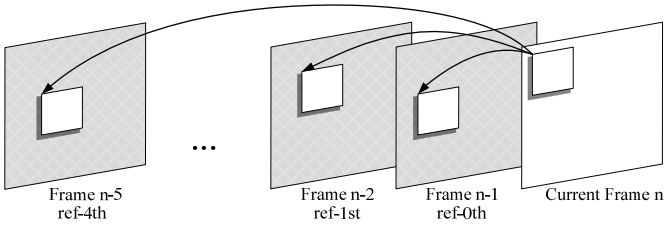

Fig. 1 Motion estimation with multiple reference frames.

In this paper, we propose a more faithful algorithm to compose new MVs when the temporal distance between the reference frame and the current frame is large. The MV composition is based on the relevant area of the current MB. It also tracks several possible candidates related to the current MB and select the best candidate. The organization of this paper is as follows. In Section II, we discuss the impacts on the performance of FDVS when the reference frame is temporally far away from the current frame. Section III describes our proposed algorithm for MRF-ME. Simulation results are presented in Section IV. Finally, some concluding remarks are provided in Section V.

II. FDVS MULTIPLE REFERENCE FRAME ENCODING

For MV composition, full search motion estimation between successive frames is carried out to obtain all MVs' information for compositing MVs in MRF-ME. Fig. 2 illustrates the example of using FDVS in MRF-ME. In this example, *ref-2nd* is the 3rd reference frame to the current *Frame n* when the number of searched reference frames, *N*, is equal to 3. Only four MBs are shown in each frame. Assume that MB_n^k represents the kth MB in *Frame n* with the MV $mv_{n\to n-1}^k$ which points to *Frame n-1* in Fig. 2. To have MV

composition between *Frame n* and the target reference frame, *ref-2nd*, it is necessary to find the new MV of MB_n^k to *Frame n-3*, i.e. $mv_{n\to n-3}^k$ in dotted arrow shown in Fig. 2(a). For every MB, FDVS selects one dominant MV carried by a dominant MB which has the largest overlapping segment with the motion-compensated MB of MB_n^k in the previous reference frame. Considering the motion-compensated MB of MB_n^1 overlaps with four MBs, MB_{n-1}^1 , MB_{n-1}^2 , MB_{n-1}^3 , and MB_{n-1}^4 , in *Frame n-1* of Fig. 2(a), MB_{n-1}^2 is chosen as the dominant MB while its MV $mv_{n-1\to n-2}^2$ is selected as the dominant MV. This dominant vector selection process is repeated until the desired reference frame is reached, i.e. *Frame n-3* in this example. $mv_{n\to n-3}^1$ therefore is composed by summing up the selected dominant MVs and can be written as

$$mv_{n\to n-3}^1 = mv_{n\to n-1}^1 + mv_{n-1\to n-2}^2 + mv_{n-2\to n-3}^2 \quad \textbf{(1)}$$

FDVS can provide promising results for MV composition for MRF-ME [3-4]. However, in fast-motion video sequences, the temporal distant reference frame is always used for MRF-ME due to existence of the fast moving objects. FDVS does not work well for this scenario. This phenomenon can be explained as portrayed in Fig. 2(b), which is redrawn from Fig.2 (a). MB_{n-1}^2 is selected to be the dominant MB and the corresponding $mv_{n-1\to n-2}^2$ is used to determine the dominant MB in *Frame n-2*. It is observed that only the shaded area of MB_{n-1}^2 is actually relevant to target MB, MB_n^1 . However, FDVS also utilizes the irrelevant non-shaded area in MB_{n-1}^2 to compute dominant MB in *Frame n-2*. The relevant area of MB_n^1 further diminishes when far away reference frames are used. The cross-hatch shaded area only occupies a very minor portion of MB_{n-2}^2 as shown in Fig. 2(b). It seriously affects the accuracy of the composed MVs since a large irrelevant area to the target MB, is used to decide the dominant MB in *Frame n-3*.

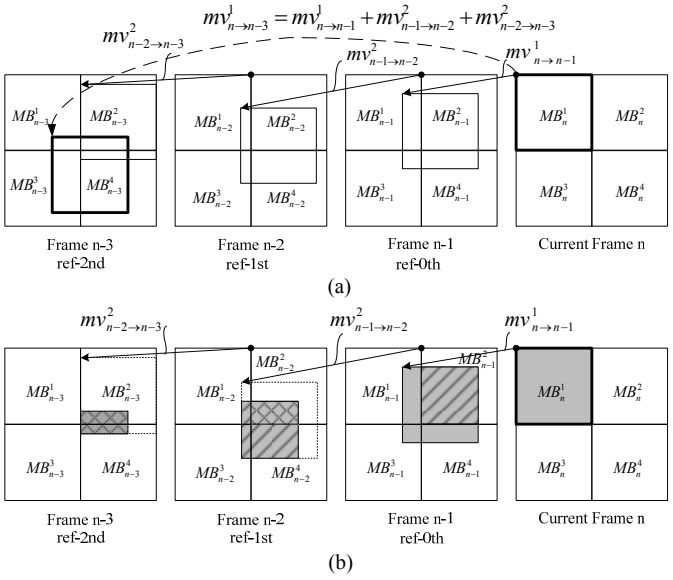

Fig. 2 Working example of FDVS in MRF-ME.

III. Proposed Vector Selection Algorithm

Based on the above observations of the FDVS process, two rules are set for our proposed algorithm in MRF-ME. First, only the area relates to the target MB should be contributed for dominant MB selection. Second, the area relevant to the target MB should be kept as large as possible during MV composition. Fig. 3 shows an improvement mechanism for FDVS. When MB_{n-1}^2 is chosen as the dominant MB in the first step of FDVS, only the shaded area with slash lines in Fig. 3, which is the relevant region to the target MB_n^1 is used to decide the next dominant MB in *Frame n-2*. Note that MB_{n-2}^4 is selected which contrasts to the selection of original FDVS where MB_{n-2}^2 is picked. For further MV composition step to the target reference *Frame n-3*, only the cross-hatch shaded area in Fig. 3 is used to determine the next dominant area in *Frame n-3*. This mechanism ensures only relevant area of MB_n^1 is employed in MV composition. From Fig. 3, the resultant MV $mv_{n\to n-3}^1$ is different from the result obtained by using FDVS in (1), and can be formed as

$$mv_{n\to n-3}^1 = mv_{n\to n-1}^1 + mv_{n-1\to n-2}^2 + mv_{n-2\to n-3}^4 \quad \textbf{(2)}$$

To maximize the relevant area used in MV composition, other non-dominant areas in the reference frames, but relevant to MB_n^1, can also be utilized to enhance the usage of relevant area in MB_n^1. In *Frame n-1* of Fig. 4(a), if the largest overlapping segment with the motion-compensated MB of MB_n^1 is not dominant enough, its size is very close to the second largest one. Since only the relevant region to MB_n^1 is employed in MV composition, the relevant area may diminish in temporally remote reference frames. In the example shown in Fig. 4(a), the cross-hatch shaded area in *Frame n-2* for selecting the next dominant MB becomes very small so it decreases the reliability of the resultant MV. To fully utilize the relevant area in MB_n^1, the proposed algorithm also considers the homogeneity of MVs, which is essential to enlarge the relevant area for MV composition. We reuse the example in Fig. 4(a), but $mv_{n-1\to n-2}^2$ is now equal to $mv_{n-1\to n-2}^4$ as shown in Fig. 4(b). In this case, the shaded area overlapped with MB_{n-1}^2 and MB_{n-1}^4 could be combined, and this merging

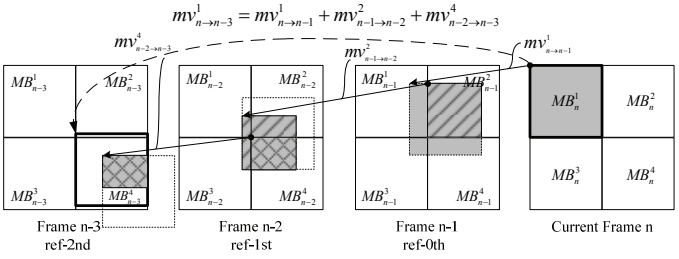

Fig. 3 Improvement mechanism for the FDVS process in which only the relevant area to the target MB is used.

area is for deciding the next dominant MB in *Frame n-2*. The selected MB in *Frame n-2* is MB^4_{n-2} where the area relevant to MB^1_n is larger and is more reliable to determine the dominant MB in *Frame n-3*, compared to the case of Fig. 4 (a).

This merging process is appropriate for areas with homogeneous motion and it is particularly true for MBs in the background and inside the moving objects. At the object boundary of a video object, no homogeneous motion field exists. We suggest using more than one candidate MB in order to expand the area relevant to the target MB in MV composition. In the following, we propose to use multiple-candidate MBs for each reference frame. Assume that C^i_{n-1} is the ith candidate in *Frame n-1* sorted by the area of the overlapping segment. In Fig. 4(c), two candidate MBs are used to compose the MV for each step. In *Frame n-1*, C^1_{n-1} and C^2_{n-1} are the largest and second largest overlapping segments with the motion-compensated MB of MB^1_n, respectively. Therefore, both MB^2_{n-1} and MB^4_{n-1} are used to determine the next dominant MBs in *Frame n-2* because both of the shaded areas in MB^2_{n-1} and MB^4_{n-1} are relevant to MB^1_n.

From the top diagram of Fig. 4(c), four candidates (C^2_{n-2}, C^3_{n-2}, C^4_{n-2}, and C^5_{n-2}) due to the motion-compensated segment of C^1_{n-1} are considered for the next step. In addition, one candidate C^1_{n-2} contributed from the motion-compensated segment of C^2_{n-1} is regarded as the possible candidate in the next step, as depicted in the bottom diagram of Fig. 4(c). Since two candidates are used for each step, from Fig. 4(c), C^1_{n-2} and C^2_{n-2} are chosen as the largest and second largest overlapping segments with their corresponding MBs, MB^4_{n-1} and MB^2_{n-1}, respectively. The top diagram of Fig. 4(c) shows the same procedure of MV composition as illustrated in Fig. 4(a). Furthermore, the bottom diagram gives an alterative path to compose the new MV, which uses the second largest candidate MB besides the largest candidate MB in the *Frame n-1*. From Fig. 4(c), we observe that the cross-hatch shaded area in *Frame n-2* of the bottom diagram, which is relevant to MB^1_n and is used to decide the dominant MB in *Frame n-3*, is larger than that of the top diagram. In other words, even though C^1_{n-1} represents the largest overlapping segment in the first reference frame, *Frame n-1*, it cannot guarantee that it is still the largest overlapping segment in the next reference frame, *Frame n-2*. The use of multiple-candidate MBs for each reference frame can increase the possibility of keeping the MBs with large relevant area to the target MB during MV composition. Since only three frames are referenced in this working example, two candidates are sufficiently enough for each reference frame. When more reference frames are adopted in ME, a larger number of possible candidates is necessary to be kept. Note that the number of candidates can be selected by the user according to the number of reference frames and the desired video quality.

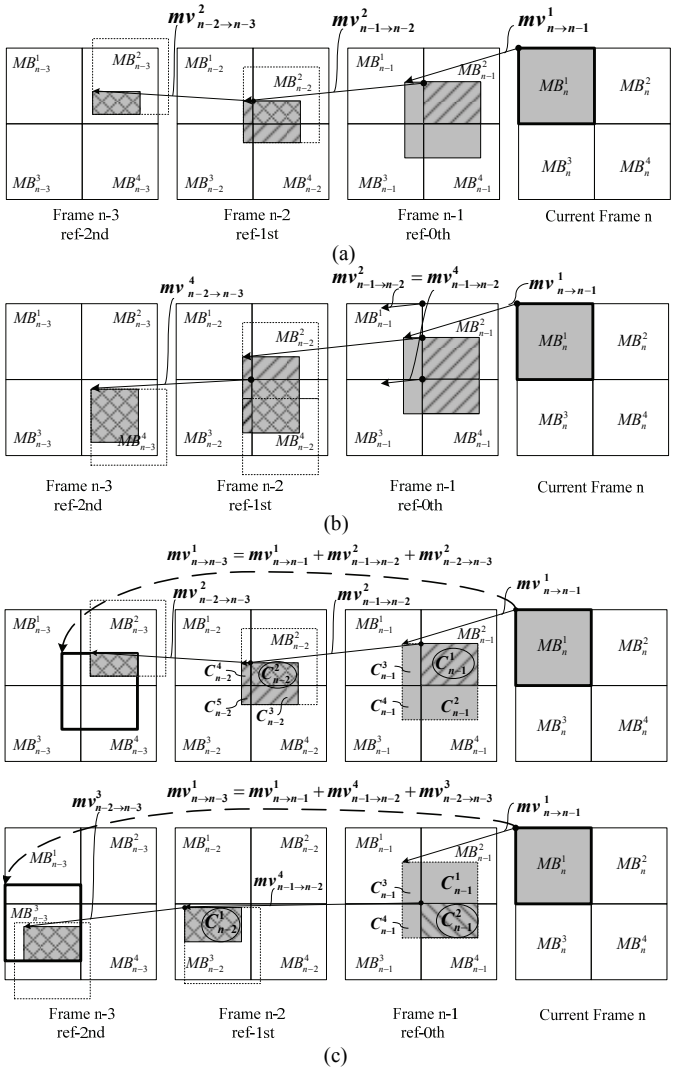

Fig. 4 (a) Scenario in which the largest overlapping segment is not dominant enough, (b) merging process with neighboring MBs of same motion vectors, and (c) multiple-candidate MB selection.

IV. Simulation Results

The proposed multiple-candidate vector selection (MCVS) algorithm has been integrated into the H.264/AVC JM9.2 codec [7] for performance evaluation in MRF-ME. In all simulations, MVs between consecutive frames by full search motion estimation with a search range of -16 to +16 pixels were obtained. The JM9.2 codec, integrated with different MV composition algorithms, was employed to reuse the previously saved MVs. Six test sequences, including "Container", "Foreman", "Salesman", "Tempete", and "Mobile" in CIF, and "Stefan" in SIF, at 30 frames/s were used. The testing conditions are listed as follows.

- 100 frames were encoded with IPPPPP…structure.
- Only inter 16x16 mode was enabled.
- No rate control was adopted.
- Quantization parameters QP 20, 24, 28 and 32 were used.
- The number of the reference frames was fixed as 5.

For comparison, the full-search motion estimation algorithm (FS), the forward dominant vector selection algorithm (FDVS) [3-4], and the proposed MCVS were adopted for MRF-ME. The number of candidate MBs selected for each stage is 4. The rate-distortion (R-D) coding performance comparisons were conducted for the following four cases:

(1) One reference frame in JM9.2, FS ref1;
(2) Five reference frames in JM9.2, FS ref5;
(3) Five reference frames for FDVS, FDVS ref5;
(4) Five reference frames, proposed MCVS ref5.

Fig. 5 and Fig. 6 show the R-D curves by using different algorithms as listed in the above four cases for MRF-ME for "Mobile" and "Stefan", respectively. The proposed MCVS outperforms FDVS, especially in the high bit-rates scenario. It is because MCVS considers only the area related to the target MB, and tries to keep it as large as possible in every MV composition step. It ensures that the resultant MV is highly correlated to the contents of target MB in current frame, which cannot be achieved by FDVS.

In TABLE I, ΔPSNR and ΔBits represent a PSNR change and a percentage change in total bit-rate respectively when compared to FS at high bit-rate scenario (QP20). The positive values mean increments whereas negative values mean decrements. It is observed that the performance of FDVS gets worse compared to FS in MRF-ME. MCVS outperforms FDVS in terms of both PSNR and total generated bits. Besides, its PSNR reduction is within 0.1dB compared to that of FS while it is 0.2 dB in FDVS. From these statistics, we can conclude that the proposed MCVS can provide outstanding performance.

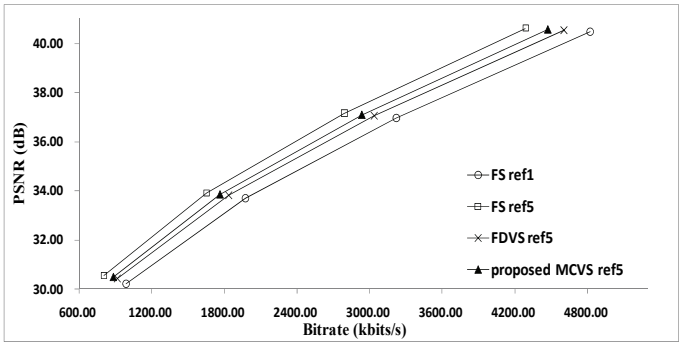

Fig. 5 R-D Comparison, "Mobile" (CIF)

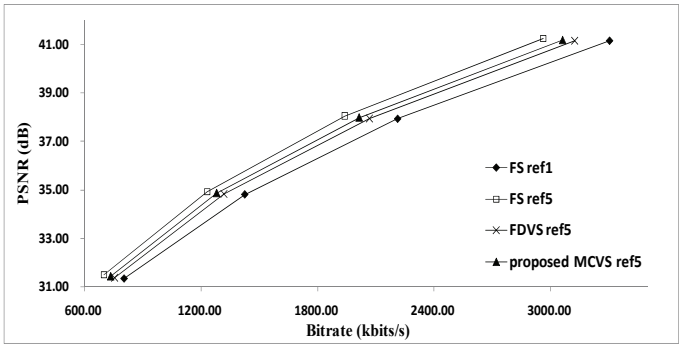

Fig. 6 R-D Comparison, "Stefan" (SIF)

TABLE I
PERFORMANCE COMPARISON USING DIFFERENT ALGORITHMS AT NUMBER OF REFERENCE FRAMES = 5 FOR VARIOUS SEQUENCES AT QP20.

Sequences	Full Search		FDVS		MCVS	
	Bits (kbps)	PSNR (dB)	ΔBits (kbps)	ΔPSNR (dB)	ΔBits (kbps)	ΔPSNR (dB)
Container	962.2	41.6	991.4 (+3.0%)	41.5 (-0.1)	980.8 (+1.9%)	41.5 (-0.1)
Foreman	1511.6	41.7	1642.7 (+8.7%)	41.6 (-0.1)	1603.9 (+6.1%)	41.6 (-0.1)
Salesman	813.8	41.1	905.5 (+11.3%)	40.9 (-0.2)	886.7 (+8.9%)	41.0 (-0.1)
Tempete	3228.1	41.0	3457.3 (+7.1%)	40.9 (-0.1)	3392.5 (+5.1%)	40.9 (-0.1)
Mobile	4287.9	40.6	4607.1 (+7.4%)	40.6 (0.0)	4475.2 (+4.4%)	40.6 (0.0)
Stefan	2959.5	41.2	3124.0 (+5.6%)	41.2 (0.0)	3061.5 (+3.4%)	41.2 (0.0)

V. CONCLUSIONS

In this paper, we have proposed a novel MV composition algorithm for MRF-ME. Our proposed multiple-candidate vector selection (MCVS) algorithm can entirely make use of the relevant area to the target MB, and it is beneficial to perform ME to a reference frame with a large temporal distance. Its performance verified experimentally in terms of both quality and bit rate is remarkably better than that of FDVS. Besides, the proposed MCVS is adaptive in nature, and the number of candidate MBs can be adjusted according to the number of reference frames.

ACKNOWLEDGMENT

The work described in this paper is partially supported by the Centre for Signal Processing, Department of EIE, PolyU and a grant from the Research Grants Council of the HKSAR, China (PolyU 5120/07E). Tsz-Kwan Lee acknowledges the research studentships provided by the University.

REFERENCES

[1] A. Luthra, G. J. Sullivan, and T. Wiegand, "Introduction to the special issue on the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 557–559, July 2003.

[2] A. Tamhankar and K. R. Rao, "An overview of H.264/MPEG-4 Part 10," in *Proc. EURASIP Conf. Focused on Video/Image Processing and Multimedia Communications*, vol. 1, pp. 1–51, Jul. 2003.

[3] M. J. Chen, Y. Y. Chiang, H. J. Li, and M. C. Chi, "Efficient multi-frame motion estimation algorithms for MPEG-4 AVC/JVT/H.264," in *Proc. IEEE ISCAS*, Vancouver, BC, Canada , pp. 737–740, May 2004.

[4] J. Youn and M. T. Sun, "A fast motion vector composition method for temporal transcoding," in *Proc. IEEE ISCAS*, Orlando, FL, vol. 4, pp. 243–246, June 1999.

[5] S. E. Kim, J. K. Han, and J. G. Kim, "An efficient scheme for motion estimation using multireference frames in H.264/AVC," *IEEE Trans. on Multimedia*, vol. 8, no. 3, pp. 457-466, June 2006.

[6] Y. Su and M. T. Sun, "Fast multiple reference frame motion estimation for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 447–452, March 2006.

[7] Reference Software JM9.2 from http://iphome.hhi.de/suehring/tml/download/