

Temporal correlation based general search strategy for multiple reference frame motion estimation

Chun-Su Park · Goo-Rak Kwon · Jae-Young Pyun · Changjae Kim

Received: 8 April 2014 / Revised: 19 October 2014 / Accepted: 25 November 2014 / Published online: 16 December 2014
© Springer-Verlag London 2014

Abstract Multiple reference frame motion estimation (MRFME) is one of the main features of recent video coding standards such as H.264/AVC and high efficiency video coding. This paper presents a fast search strategy for avoiding redundant computations during MRFME. The proposed method reduces the number of pixels to be examined using the precalculated rate-distortion costs of previously searched frames. Experimental results show that the proposed method reduces the complexity of the motion estimation (ME) process by up to 81.43 %.

Keywords Motion estimation · Multiple reference frames · Search strategy

1 Introduction

With the generalization of mobile communication services, user-created video content has been attracting considerable attention because it enables users to share personal experiences with their family and friends [1,2]. The captured videos are compressed in a portable device, and then the

compressed video streams are transmitted over networks with limited bandwidth [3]. Compression schemes which can provide high quality video with low complexity, play a key role for this application.

Block-based motion estimation (ME) exploiting temporal redundancy between adjacent frames is one of the most significant features of state-of-the-art video coding standards such as H.264/AVC and high efficiency video coding (HEVC) [4–7]. For each block in the current frame, the ME process is performed to find the best prediction, called the reference block, which minimizes the matching distortion within the predetermined search space. The motion vector (MV) and the reference index, respectively, are used to specify the spatial and temporal displacement between the current and reference blocks. It is well-known that the ME technique significantly improves video coding efficiency, but it is the most time-consuming part of the video compression process.

Several fast search algorithms have been proposed to accelerate the ME process. This paper focuses on ME algorithms that accelerate the search process without sacrificing coding efficiency. The partial distortion elimination (PDE) algorithm uses the partial sum of the matching distortion to eliminate impossible pixel positions before completing the calculation of the matching distortion [8–10]. The successive elimination algorithm (SEA) utilizes a simple matching criterion to eliminate impossible pixels before more precise distortion calculations are performed. Using an appropriate test, many pixels can be excluded from further consideration in the ME process. In order to further reduce the complexity of SEA, several improved algorithms have been introduced in recent studies [11–15]. Basically, all of these methods are designed for single reference frame ME (SRFME). In multiple reference frame ME (MRFME), the block matching process is conducted using additional reference frames, thereby obtaining a better prediction signal as

C.-S. Park
Department of Digital Contents, Sejong University, Seoul, Korea
e-mail: cspark@sejong.ac.kr

G.-R. Kwon · J.-Y. Pyun
Chosun University, Donggu, Gwangju, Korea
e-mail: grkwon@chosun.ac.kr

J.-Y. Pyun
e-mail: jypyun@chosun.ac.kr

C. Kim (✉)
Department of Civil and Environmental Engineering, Myongji University, Yongin, Korea
e-mail: earth2moon@gmail.com

compared to SRFME. The SRFME algorithms can be directly applied to each reference frame, i.e., all reference frames are searched repeatedly. However, in this case, the complexity of the encoder increases significantly in proportion to the number of available reference frames.

This paper proposes an effective search strategy for accelerating MRFME [16–18]. The proposed method first analyzes the available information from the ME process of the previous frames. Using this information, the proposed method derives a new inequality constraint that will be used for ME of the current frame. In the middle of the MRFME process, the number of pixels to be examined can be reduced by gradually increasing the tightness of the proposed inequality constraint. This paper focuses on the spiral search of the H.264/AVC reference software [19]. However, the proposed strategy can be applied to most conventional full-search-equivalent ME algorithms [20].

The rest of this paper is organized as follows. Section 2 introduces some preliminaries about rate-distortion (RD) optimization in video processing. Section 3 describes the derivation of a new inequality constraint first and, then, introduces a general search strategy for MRFME based on the inequality constraint. The experimental results of the proposed search strategy are presented in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 Preliminary

The problem of finding the optimal MV and reference index for a given rate constraint can be formulated as finding the best point on the convex hull of all possible RD points. Lagrangian optimization is widely utilized for solving this problem. Let s and $c(m, n)$ be, respectively, the original block and its reconstruction obtained by using the MV m and reference index n . For an inter-coded block, ME is performed to find the optimal MV and reference index by minimizing [21]

$$J(m, n|s, \hat{m}) = D(s, c(m, n)) + \lambda(R(m - \hat{m}) + R(n)) \quad (1)$$

where \hat{m} is the prediction for the MV and λ is the Lagrange multiplier. The distortion $D(s, c(m, n))$ is the difference between s and $c(m, n)$. The rates $R(n)$ and $R(m - \hat{m})$ specify, respectively, the numbers of bits required for encoding the reference index and the difference between the original MV and the predicted one.

The video encoder searches for the best matching position within the predefined search window of size $(2w + 1) \cdot (2w + 1)$ for each reference frame. Thus, given N reference frames, the overall size of the motion search space is $N \cdot (2w + 1) \cdot (2w + 1)$ for each block. Let Ψ be a search window of size $(2w + 1) \cdot (2w + 1)$ and $m_k \in \Psi$, $0 < k \leq (2w + 1) \cdot (2w + 1)$, be an MV indicating a possible pixel position in the spiral order. Then, Ψ is composed of the pixel positions within the

search window as follows

$$\Psi = \{m_1, m_2, \dots, m_W\} \quad (2)$$

where $W = (2w + 1) \cdot (2w + 1)$. Suppose that, in the middle of the ME process, the encoder has examined the sub-window $\Psi_k = \{m_1, m_2, \dots, m_k\}$ in the reference frame with the index n . Then, the encoder can obtain the local minimum RD cost J_k^n of the reference frame as follows

$$J_k^n = \min_{m_i \in \Psi_k} \{J(m_i, n|s, \hat{m})\} \quad (3)$$

where $0 < i \leq k$. Using the above notations, the proposed search strategy is described in the next section.

3 Proposed search strategy

In Sect. 3.1, a new criterion is first derived to decide sequentially whether to skip the precise RD cost calculation of the remaining pixels. Then, in Sect. 3.2, the tightness of the proposed criterion is increased using the precalculated RD costs of the previously searched frames.

3.1 Early skipping criterion

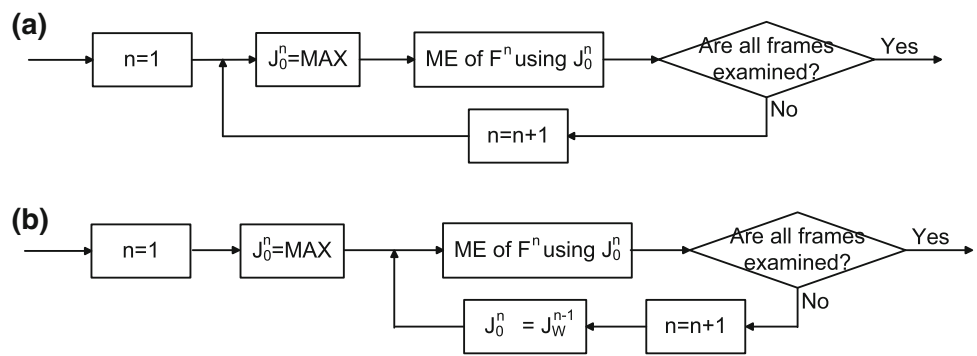
In general, the MV distribution is highly center-biased. Cheung et al. [22] showed that most of the MVs concentrate on the search center, i.e., approximately 81.80% of the MVs are located in the central 5×5 square area. In order to exploit these characteristics, the spiral search adopted in the H.264/AVC reference software starts the search process at the center of the search window and then moves the search position pixel by pixel in the spiral structure around the search center. Note that, in the current implementation, the MV prediction \hat{m} is used as an initial search center. The spiral search produces the best performance for video signals that have the MV distributions gathered at the zero position (0, 0). This is because the probability of finding the minimum matching error at the beginning of the search process will increase. It is well-known that the spiral search achieves better performance than other algorithms utilizing the raster scan order.

The spiral search establishes an inequality constraint in order to accelerate the search process while preserving the RD optimized solution for the MV. The spiral search adopts the following inequality relation as a criterion for skipping the precise RD cost calculation [23]:

$$R(m - \hat{m}) \leq \frac{J}{\lambda}. \quad (4)$$

This leads to the result that, after examining the sub-window Ψ_k , the encoder evaluates the cost function only for the MVs

Fig. 1 MRFME processes of **a** the conventional and **b** the proposed methods



m_j 's which satisfy the following criterion:

$$R(m_j - \hat{m}) \leq \frac{J_k^n}{\lambda} \tag{5}$$

where $k < j \leq W$. Note that, as shown in (4) and (5), the spiral search exploits only the inequality relation between $R(m - \hat{m})$ and J .

It is natural that the complexity reduction is highly dependent on the tightness of the inequality constraint. As the tightness increases, there are smaller number of pixels which the encoder needs to examine in the ME process. In this paper, a new criterion is introduced to increase the tightness of the inequality constraint. As mentioned earlier, when multiple reference frames are used in the ME process, the rate $R(n)$ for encoding the reference index is included in the Lagrangian cost function. Thus, a tighter inequality constraint between the cost and rate can be derived by eliminating the distortion $D(s, c(m))$ from (1) as

$$R(m - \hat{m}) \leq \frac{J}{\lambda} - R(n). \tag{6}$$

Then, after examining the sub-window Ψ_k , the proposed method needs to evaluate the cost function only for the MVs m_j 's which satisfy the following criterion:

$$R(m_j - \hat{m}) \leq \frac{J_k^n}{\lambda} - R(n) \tag{7}$$

where $k < j \leq W$. If this inequality is not satisfied at the search position corresponding to m_j , m_j cannot be the best prediction. Therefore, the RD cost calculation of the search position corresponding to m_j can be safely skipped. Otherwise, the cost function is evaluated at that particular position. If its RD cost is less than J_k^n , J_k^n is updated and the updated J_k^n is used for the remaining pixels.

3.2 Fast ME algorithm for MRFME

From (7), it can be seen that the complexity reduction is dependent on the local minimum RD cost J_k^n . The lower the value of J_k^n is, the smaller the number of pixels that need to be examined is. In the current implementation, the spiral search is applied to each reference frame separately. This means that J_k^n is initialized with the highest possible value when starting

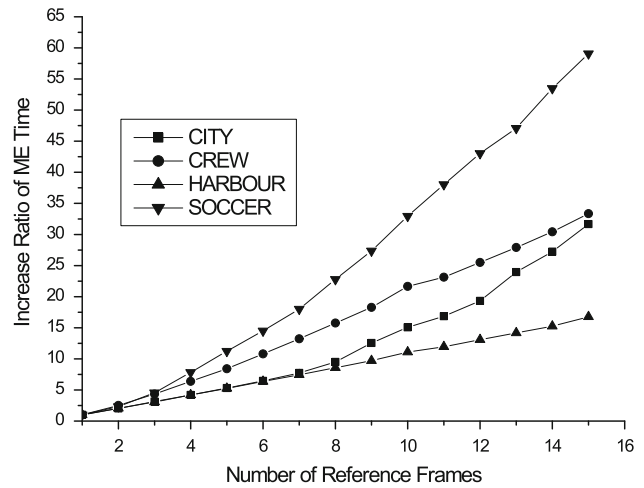


Fig. 2 Increase ratio of ME time to the number of reference frames for QCIF sequences with QP = 22

the search process of a new reference frame (see Fig. 1a). Therefore, as shown in Fig. 2, the computational complexity of the ME process increases drastically in proportion to the number of reference frames.

In order to reduce the number of pixels to be examined, the tightness of the inequality constraint (7) is further increased using the precalculated RD costs of the previously searched frames. Suppose that there are N reference frames $\{F^1, F^2, \dots, F^n, \dots, F^N\}$ in the reference list, where n specifies the reference index. Analogous to (3), let J_0^n and J_W^n be, respectively, an initial value of J_k^n and the minimum RD cost obtained from the ME process of F^n . Then, from the definitions, the following relation holds:

$$J_0^n \geq J_W^n. \tag{8}$$

Instead of initializing J_0^n with the highest possible value, the proposed algorithm sets the initial value as small as possible using the precalculated RD costs. When starting the search process of F^n , the proposed method sets J_0^n to the lowest value among the precalculated RD costs as

$$J_0^n = \min\{J_W^1, J_W^2, \dots, J_W^{n-1}\}. \tag{9}$$

If the above initialization method is sequentially applied to all reference frames, the following relations hold in the middle of the MRFME process:

$$J_0^n \leq J_W^{n-1} \quad (10)$$

and, from (8),

$$J_W^n \leq J_W^{n-1}. \quad (11)$$

Then, based on these observations, the proposed algorithm shown in (9) can be simplified as

$$J_0^n = J_W^{n-1}. \quad (12)$$

This leads to the result that the proposed method sets J_0^n to J_W^{n-1} and performs ME of F^n using (7). Figure 1b shows the MRFME process adopting the proposed strategy. Note that, after finishing the search process of F^{n-1} , the encoder can obtain the minimum RD cost J_W^{n-1} without performing any additional computations. Therefore, the proposed method does not cause any additional computational overhead. For each block, the overall algorithm proceeds as follows:

- (a) At the beginning of the ME process, the encoder searches for the best matching block within the reference frame F^1 . The minimum RD cost J_W^1 obtained by using F^1 is stored in memory for the following steps.
- (b) Initialize J_0^n with J_W^{n-1} as shown in (12). Then, the proposed method performs the ME process of the reference frame F^n using (7). Note that the resultant J_W^n of F^n is stored in memory for calculating the initial value J_0^{n+1} of the next reference frame F^{n+1} . This procedure is repeated by increasing n by one.
- (c) If n becomes larger than N , the encoder terminates the search process of the current block. The MV and reference index corresponding to the optimal search position are encoded.

It is obvious that the proposed algorithm can be easily integrated into conventional fast search algorithms designed for SRFME [8, 11]. Basically, these algorithms establish inequality constraints using the local minimum RD cost and perform the ME process of each reference frame separately. In the MRFME scenario, the proposed algorithm can keep the local minimum RD cost as small as possible. Therefore, when applying the existing SRFME algorithms to MRFME, the performance of the existing algorithms can be improved significantly by combining them with the proposed search algorithm.

4 Experimental results

In the simulation, JM 18.3 reference software was used and the performance was evaluated by using several video sequences with 1080p format. Only the first frame was encoded in intra mode and ME was performed with integer-pixel accuracy. The quantization parameters (QPs) were set to 22, 26, 30, and 34. The processing time of ME was measured for video sequences with 100 frames. The search range was set to 32×32 and context-based adaptive variable length coding (CAVLC) was used in all simulations. We used sum of absolute differences (SAD) for measuring the distortion between the original and reference blocks. The proposed algorithm was compared to the sorting-based PDE (SPDE) algorithm [8], the column-based SEA (CSEA) [15] algorithm, and the spiral search (SP) of the reference software [19]. Since all algorithms generate the same ME results, only the computational complexity was evaluated through simulations.

4.1 Complexity analysis

The computational complexity of an algorithm is analyzed by measuring the number of examined pixels in the ME process. Figure 3 shows the ratio of the number of examined pixels to the total number of pixels for each reference frame. For the first reference frame with index 1 ($n = 1$), all algorithms can significantly reduce the number of examined pixels in the search process. For the Park Scene sequence, the SPDE, CSEA, and SP algorithms reduce the number of examined pixels by 78.32, 81.23, and 85.58 %, respectively. Note that, when $n = 1$, the number of examined pixels of the proposed algorithm is the same as that of SP.

Figure 3 clearly shows that, in the case of SPDE, CSEA, and SP, the ratio of the examined pixels increases as n increases. The algorithms are applied to each reference frame separately. This results in that the local minimum RD cost used for eliminating impossible pixels increases as n increases. Note that the lower the local minimum RD cost is, the smaller the number of pixels to be examined is. Therefore, in the case of SPDE, CSEA, and SP, the encoder should examine more pixels as n increases.

The proposed algorithm can keep the local minimum RD cost as small as possible by exploiting the precalculated RD costs of the previously searched frames. Figure 3 presents that the proposed algorithm reduces the ratio of the number of examined pixels as n increases. Therefore, it can be expected from Fig. 3 that the performance improvement of the proposed algorithm will increase sharply as n increases. The processing time comparison is presented in the following subsection.

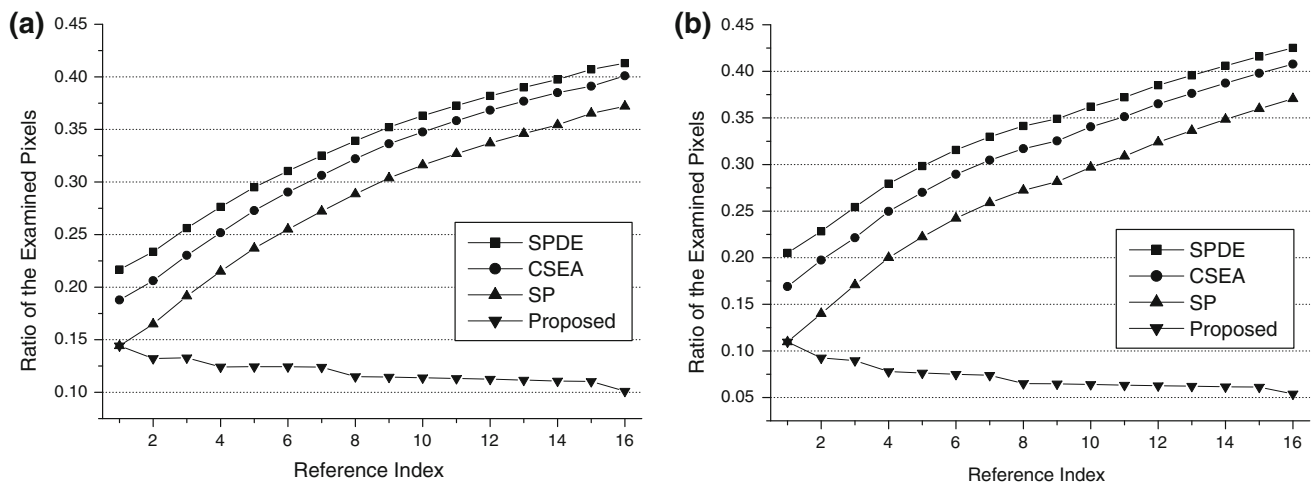


Fig. 3 Ratio of the examined pixels with varying reference indexes. **a** Park Scene. **b** Rush hour

Table 1 Processing time comparison of the conventional and proposed algorithms for $N = 4$

Sequences	QP	Bitrate (kbps)	PSNR (dB)	Time saving		
				ΔT_{SPDE} (%)	ΔT_{CSEA} (%)	ΔT_{SP} (%)
Park scene	22	37,768.20	40.44	24.62	22.70	13.88
	26	17,480.83	37.84	35.05	32.12	16.57
	30	8,292.62	35.27	43.21	39.24	18.77
	34	3,845.52	32.91	50.59	46.68	20.77
Pedestrian area	22	15,092.32	42.10	42.31	37.44	30.83
	26	7,533.54	40.59	51.17	44.75	36.94
	30	4,438.59	38.93	58.08	50.67	41.55
	34	2,703.09	37.08	63.95	55.14	45.46
Rush hour	22	12,093.92	42.65	35.10	28.87	18.27
	26	6,008.27	41.73	48.62	41.83	25.41
	30	3,464.76	40.41	58.74	51.82	31.24
	34	2,015.47	38.72	66.39	58.81	35.62
Sunflower	22	8,622.47	43.70	51.25	45.34	38.63
	26	4,385.60	42.59	63.94	56.89	49.93
	30	2,512.49	40.81	71.17	63.90	56.45
	34	1,434.38	38.76	75.65	67.76	59.66
Average				52.49	46.50	33.75

4.2 Processing time comparison

The processing time of several ME algorithms was measured with variation of the number of reference frames (N) and, in the simulation, N was set to 4 and 8. The measured results are summarized in Tables 1 and 2 where ΔT_{SPDE} , ΔT_{CSEA} , and ΔT_{SP} indicate the time savings of the proposed algorithm as compared to SPDE, CSEA, and SP, respectively. Note that, since all ME algorithms accelerate the ME process without sacrificing the coding efficiency, their bitrates and PSNRs are exactly the same as each other.

Table 1 presents the processing time comparison for $N = 4$. In this case, the average ME time of the proposed algorithm is 52.49 and 46.50% less than those of SPDE and CSEA, respectively. Further, if $N = 8$, Table 2 shows that the amount of the ME time reduction increases to 59.62 and 54.65% as compared to SPDE and CSEA, respectively. As compared to the SP algorithm, the average ME time saving of the proposed algorithm is 33.75% for $N = 4$ and 47.30% for $N = 8$. The results clearly show that the performance of the proposed algorithm is improved as N increases. This is because the proposed algorithm gradually increases the tight-

Table 2 Processing time comparison of the conventional and proposed algorithms for $N = 8$

Sequences	QP	Bitrate (kbps)	PSNR (dB)	Time saving		
				ΔT_{SPDE} (%)	ΔT_{CSEA} (%)	ΔT_{SP} (%)
Park scene	22	36,097.83	40.44	32.03	29.02	25.08
	26	16,740.68	37.86	42.03	38.44	28.94
	30	8,057.44	35.29	49.96	45.82	31.86
	34	3,787.65	32.93	56.18	50.53	34.79
Pedestrian area	22	14,571.09	42.12	50.66	45.59	43.02
	26	7,310.54	40.60	58.51	52.05	49.24
	30	4,337.30	38.94	65.42	56.62	53.78
	34	2,663.08	37.10	70.57	61.47	58.55
Rush hour	22	11,831.08	42.65	42.43	38.08	30.33
	26	5,892.65	41.74	56.05	51.64	40.33
	30	3,409.85	40.44	66.00	61.31	46.63
	34	1,999.66	38.77	72.81	68.02	53.06
Sunflower	22	8,532.24	43.72	60.19	56.60	53.86
	26	4,348.18	42.61	71.75	67.95	64.44
	30	2,505.30	40.84	77.91	73.82	69.57
	34	1,426.43	38.81	81.43	77.41	73.25
Average				59.62	54.65	47.30

ness of the inequality constraint (7) by using the precalculated RD costs of the previously searched frames. The proposed algorithm can reduce the ME time by up to 81.43% for the Sunflower sequence with $N = 8$ as compared to the SPDE algorithm. The experimental results reveal that the proposed algorithm consistently outperforms the SPDE, CSEA, and SP algorithms.

It can also be observed from Tables 1 and 2 that the amount of the ME time reduction becomes larger as the QP increases. For example, when the QP is set to 22 and $N = 8$, the ME time reduction of the proposed method is 46.33% as compared to the SPDE algorithm. Further, the average ME time savings increase to 57.08, 64.82, and 70.25% for the QPs 26, 30, and 34, respectively. As compared to the SP algorithm, the average ME time savings of the proposed method are 38.07, 45.74, 50.46, and 54.91% for the QPs 22, 26, 30, and 34, respectively. Therefore, the proposed method tends to achieve better performance at low bitrates.

5 Conclusion

A new MRFME algorithm with reduced computational complexity was presented in this paper. At first, a new criterion was derived to decide sequentially whether to skip the precise RD cost calculation of the remaining pixels. Then, this paper proposed an algorithm that can gradually increase the tightness of the proposed criterion utilizing the precalculated

RD costs of the previously searched frames. The experimental results clearly demonstrated that the proposed method achieves a substantially higher speedup than the SPDE and SP algorithms in the MRFME scenario.

References

1. Park, I., Capson, D.W.: Improved motion estimation time using a combination of dynamic reference frame selection and residue-based mode decision. *Signal Image Video Process.* **6**, 25–39 (2012)
2. Pan, X., Ye, Y., Wang, J.: Fractional directional derivative and identification of blur parameters of motion-blurred image. *Signal Image Video Process.* **8**, 1–12 (2013)
3. Park, C.S.: Spatial and interlayer hybrid intra-prediction for H.264/SVC video. *Opt. Eng.* **52**(7), 071503 (2013)
4. Sullivan, G., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012)
5. Draft ITU-T Recommendation and final draft international standard of join video specification, ITU-T Rec. H.264/ISO/IEC 14496–10 AVC (2003)
6. Park, C.S.: Level-set-based motion estimation algorithm for multiple reference frame motion estimation. *J. Vis. Commun. Image Represent.* **24**(8), 1269–1275 (2013)
7. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 560–576 (2003)
8. Choi, C., Jeong, J.: New sorting-based partial distortion elimination algorithm for fast optimal motion estimation. *IEEE Trans. Consum. Electron.* **55**(4), 2335–2340 (2009)
9. Gersho, A., Gray, R.M.: *Vector quantization and signal compression*. Kluwer, Boston (1991)

10. Kim, J.N., Ryu, T.K.: A fast PDE algorithm using adaptive matching scan order for real-time video coding. *Lect. Notes Comput. Sci.* **6111**, 336–343 (2010)
11. Li, W., Salari, E.: Successive elimination algorithm for motion estimation. *IEEE Trans. Image Process.* **4**(1), 105–107 (1995)
12. Gao, X.Q., Duanmu, C.J., Zou, C.R.: A multilevel successive elimination algorithm for block matching motion estimation. *IEEE Trans. Image Process.* **9**, 501–504 (2000)
13. Zhu, C., Qi, W.-S., Ser, W.: Predictive fine granularity successive elimination for fast optimal block matching motion estimation. *IEEE Trans. Image Process.* **14**(2), 213–221 (2005)
14. Liu, S.W., Wei, S.D., Lai, S.H.: Fast optimal motion estimation based on gradient-based adaptive multilevel successive elimination. *IEEE Trans. Circuits Syst. Video Technol.* **18**(2), 263–267 (2008)
15. Paramkusam, A.V., Reddy, V.S.K.: The efficient optimal and suboptimal motion estimation algorithms. *Signal Image Video Process.* 1–6 (2013). doi:[10.1007/s11760-013-0562-y](https://doi.org/10.1007/s11760-013-0562-y)
16. Su, Y., Sun, M.-T.: Fast multiple reference frame motion estimation for H. 264/AVC. *IEEE Trans. Circuits Syst. Video Technol.* **16**(3), 447–452 (2006)
17. Jun, D., Park, H.: An efficient priority-based reference frame selection method for fast motion estimation in H. 264/AVC. *IEEE Trans. Circuits Syst. Video Technol.* **20**(8), 1156–1161 (2010)
18. Ho, H., Klepko, R., Ninh, N., Wang, D.: A high performance hardware architecture for multi-frame hierarchical motion estimation. *IEEE Trans. Consum. Electron.* **57**(2), 794–801 (2011)
19. Lim, K.P., Sullivan, G., Wiegand, T.: Text description of joint model reference encoding methods and decoding concealment methods. Joint Video Team, Doc. JVT-N046 (2005)
20. Coban, M.Z., Mersereau, R.M.: A fast exhaustive search algorithm for rate-constrained motion estimation. *IEEE Trans. Image Process.* **7**(5), 769–773 (1998)
21. Park, C.S., Jung, S.W., Choi, K.S., Ko, S.J.: Fast encoding algorithm avoiding repetition of motion estimation in scalable video coding. *Electron. Lett.* **46**(4), 280–282 (2010)
22. Cheung, C.H., Po, L.M.: A novel cross-diamond search algorithm for fast block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **12**(12), 1168–1177 (2002)
23. Park, C.S.: A general search strategy for multiple reference frame motion estimation. In: 2013 IEEE International Conference on Consumer Electronics (ICCE), pp. 328–329 (2013)