The International Journal of Engineering and Science (IJES)
|| Volume || 14 || Issue 6 || June || Pages || PP 72-79 || 2025 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Fast Reference Frame Decision Algorithm for Multi-Reference Frame Motion Estimation in H.265

Hung-Ming Hung, Ke-Nung Huang, Chou-Chen Wang
Department of Electronic Engineering, I-Shou University, Kaohsiung, Taiwan
Corresponding Author: Chou-Chen Wang

----------ABSTRACT----------
The H.265 video coding standard promotes the realization of 4K/8K ultrahigh definition (UHD) video applications. To further improve the coding efficiency, H.265 allows motion estimation (ME) performing on multiple reference frame (MRF). Although the MRF can enhance the performance and allow the encoder to search a better reference frame from several previous pictures, the computational complexity of the MRF-based ME (MRF-ME) module dramatically increases. To improve the coding performance of H.265 according to the high spatiotemporal correlation existing in the MRF, we firstly propose neighboring-block-based reference frame decision algorithm (NRFDA) and priority-based reference frame selection algorithm (PRFSA) to reduce the computational complexity of ME-MRF module. The NRFDA utilizes the selected reference frames information among encoded neighboring blocks and the variance of the current block to predict the best reference frame. Therefore, PRFSA define the priority for each reference frame so that ME can perform on the reference frames along the descending order of priority according to the rate distortion cost (RDcost)*rank. Finally, we integrate NRFDA and PRFSA into a fast reference frame decision algorithm (FRFDA) to further speed up the ME-MRF module. Simulation results show that the proposed FRFDA can achieve an average time improving ratio (TIR) about 68.23% when compared to H.265 (HM16.7) under MRF=4. It is clear that the proposed algorithm can efficiently increase the encoding speed of H.265 with insignificant loss of image quality.*

KEYWORDS–*Video coding, H.265, Motion estimation, Ultrahigh definition video.*

Date of Submission: 06-06-2025 Date of acceptance: 17-06-2025

I. INTRODUCTION

Nowadays, H.265 is the most commonly used video formats for recording, compression and distribution of videos. This is because the demand for high resolution video or ultra-high definition (UHD) video has rapidly increased in a number of industries, especially in entertainment, intelligent video surveillance, video conference and live streaming [1-2]. H.265 adopts some new coding structures including coding unit (CU), prediction unit (PU) and transform unit (TU), as shown in Fig. 1 [3]. The CU is the basic unit of region splitting used for inter/intra prediction, which allows recursive subdividing into four equally sized blocks. The CU can be split by coding quadtree structure of 4 level depths, which CU size ranges from largest CU size of 64×64 pixels to the smallest CU size of 8×8 pixels. At each depth level (CU size), H.265 performs motion estimation (ME) and motion compensation (MC) with different size. The PU is the basic unit used for carrying the information related to the prediction processes, and the TU can be split by residual quadtree (RQT) at maximally 3 level depths which vary from 32×32 to 4×4 pixels.

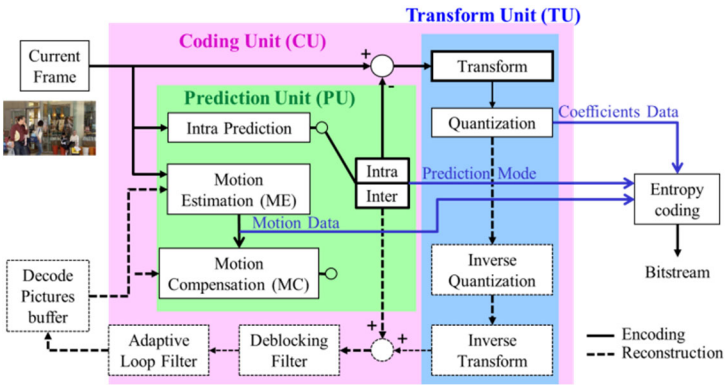

Fig. 1 The block diagram of HEVC encoder.

In general, intra-coded CUs have only two PU partition types including 2N×2N and N×N but inter-coded CUs have eight PU types including symmetric blocks (2N×2N,2N×N, N×2N, N×N) and asymmetric blocks (2N×nU, 2N×nD, nL×2N, nR×2N) [1]. The rate distortion costs (RDcost), which include J_{intra} ,J_{inter} and J_{mode},have to be calculated by performing the PUs and TUs to select the optimal partition mode under all partition modes for each CU size. In the PU structure, H.265 adopts ME module to choose the optimal inter prediction mode. In order to improve the accuracy of PU prediction, multiple reference frames (MRF) interframe prediction is performed in the ME module for H.265. Suppose that four reference indexes (RefIdx) of frames are used, the selecting process of inter prediction mode using MRF-based ME (MRF-ME) is shown in Fig. 2. We can summarize the decision process of inter prediction mode. Firstly, H.265 adopts the coding tree unit (CTU), and each CTU allows recursive splitting into four equal CU. And then, the PU performs the inter prediction processes. When pruning the best CTU coding quadtree, the inter prediction module executes 7 different prediction modes to find the best partition mode after MRF-ME procedure.

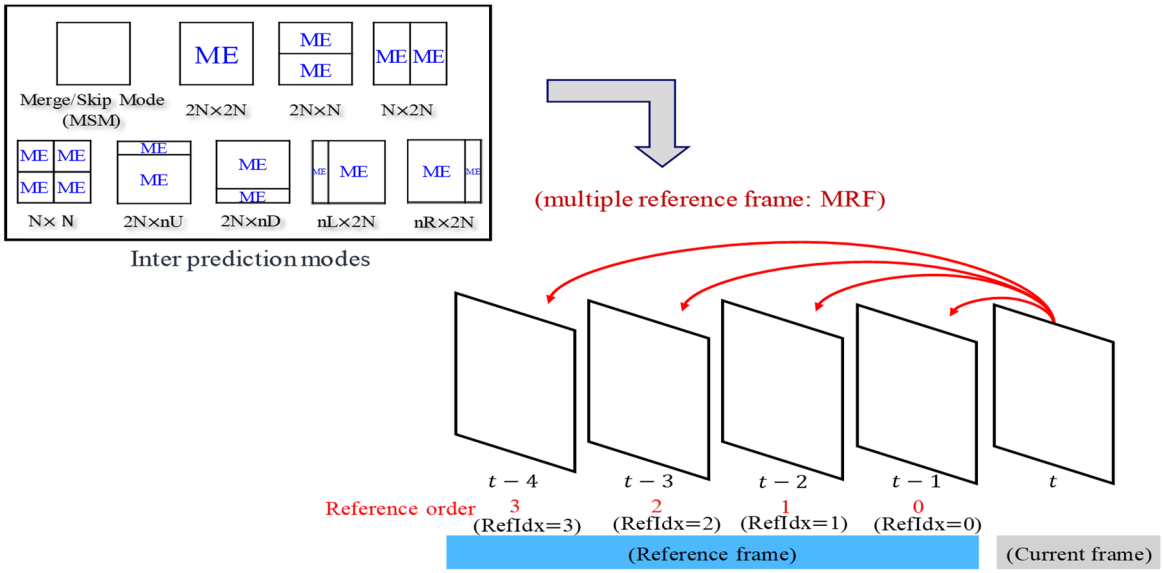

Fig. 2 The selecting process of inter prediction mode using MRF-ME scheme.

Although the MRF-ME module in H.265 can enhance the PU performance and allow the encoder to search a better reference frame from several previous pictures, the computational complexity of the MRF-ME searching process dramatically increases [4-5]. Therefore, a heavily computational complexity becomes a main bottleneck for the real-time applications of H.265 in UHD videos, such as live video broadcasting, mobile video communication and video surveillance. In order to reduce the computational complexity of MRF-ME module in H.265, Yang et. al. proposed a fast reference picture selection algorithm (FRPSA)for H.265encoder [4]. After the statistical analysis in performing MRF-ME searching process, they found that a high correlation exists between the best reference frame and lowest RDcost associated with advanced motion vector prediction (AMVP). Therefore, they use the predefined threshold to determine whether the AMVP-selected reference frame is the best reference frame. However, Yang's method relies on the correlation statistics of AMVP across MRFs and determines the threshold based on the quantization parameter (QP) and PU size. When applied to scenes with significant variations or complex backgrounds, the average number of reference frames tends to increase, which reduces the algorithm's effectiveness in speeding MRF-ME process.

To improve the coding performance of H.265 according to the high spatiotemporal correlation existing in the MRF, we firstly propose two methods including neighboring-block-based reference frame decision algorithm (NRFDA) and priority-based reference frame selection algorithm (PRFSA) to reduce the computational complexity of ME-MRF module. The NRFDA utilizes the selected reference frames information among encoded neighboring blocks and the variance of current block to predict the best reference frame. In addition, PRFSA define the priority for each reference frame so that ME can perform on the reference frames along the descending order of priority according to the RDcost of AMVP (J_{AMVP}) rank. Finally, we combine NRFDA and PRFSA into a fast reference frame decision algorithm (FRFDA) to further speed up the ME-MRF module.

The remainder of this paper is organized as follows. In Section II we briefly review the fast encoding using MRF-ME. Section III elaborates the proposed fast reference frame decision algorithm to speed up ME-MRF selection process in H.265. The experimental results are presented in Section IV.

II. FAST H.265 ENCODING USING MULTIPLE REFERENCE FRAMES

2.1 MRF-ME selecting process in H.265

In the H.265 compression standard, CUs perform inter prediction at different depths, where each CU is further divided into different PUs. The PU structure allows for various MVs of image objects by selecting appropriate prediction modes to perform motion compensation. To improve the accuracy of PU prediction, H.265 permits the ME module to adopt MRF prediction for more precise results.

To select the best reference frame among these MRFs, H.265 must search each reference frame by performing extensive RDcost computation and comparison. This leads to significantly increased computational complexity for the MRF-ME module. The selecting best frame process of ME-MRF module using the temporal correlation is also illustrated in Fig.2. Here, t represents the current frame, while $t-1$ to$t-4$ denote four previous reference frames. According to the reference order, each reference frame is assigned a reference index (RefIdx), and the search proceeds in the order: $t-1 \rightarrow t-2 \rightarrow t-3 \rightarrow t-4$.The ME-MRF process begins with RefIdx = 0, which uses the sum of absolute differences (SAD) to find the best matching block and obtain the most suitable MV. The process is repeated for RefIdx = 1 to 3, and each time generates a corresponding best MV. Finally, the RDcost (J_{inter}) is computed for each reference frame among RefIdx = 0 to 3, and the reference frame with the minimum $J_{inter}$is selected as the best reference frame for MC. Figure 3 shows the complete pruning process to find the best CTU coding tree in each module of H.265 encoder. Therefore, to meet the demands of real-time video transmission applications, it is essential to further reduce the computational complexity of the ME-MRF search process.

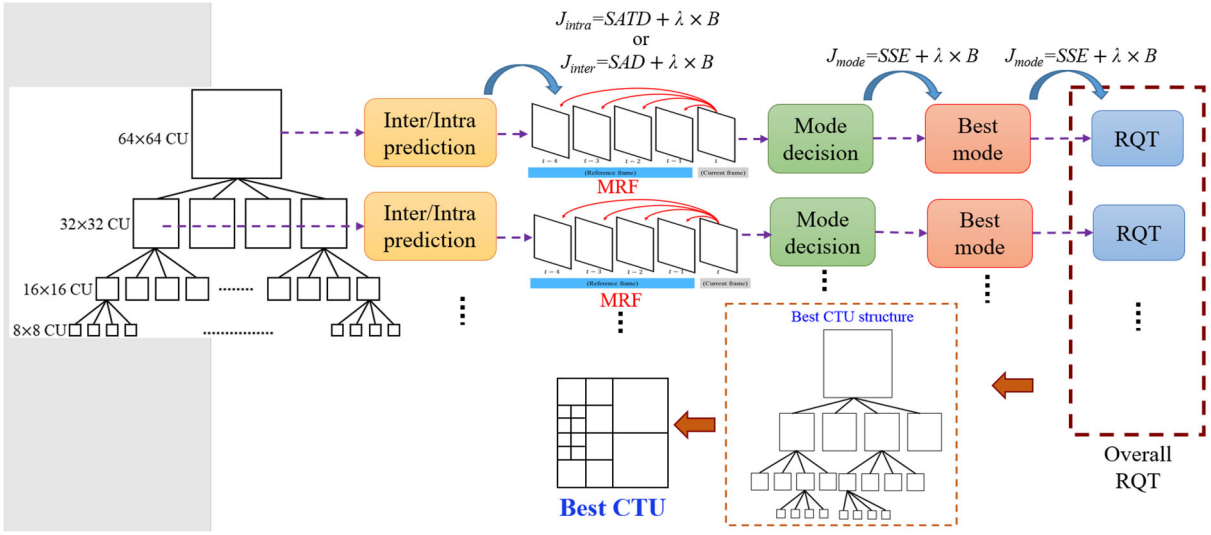

Fig. 3 The pruning process to find the best CTU in each module of H.265 encoder.

In order to further improve the coding efficiency, H.265 standard employs AMVP algorithm to find the best MV predictor for the current PU. The AMVP algorithm produces initial MVs (IMVs) for all the reference frames for current PU. An IMV is chosen from the available MVs of spatially neighboring or temporally collocated coded PUs of the current PU [3]. In other words, every reference frame in the MRF will produce one IMV. Figure 4 shows the encoding process of MRF-ME module using four reference frames to find the best reference frame. We can simply describe the working procedure in MRF-ME module as follows. Firstly, the RD costs (J_{AMVP}) associated with those IMVs from AMVP are evaluated and one best IMV with minimum J_{AMVP} is chosen for every reference frame (denoted as J_{AMVP}_ref _m: m=0~3). And then, the MRF-ME performs the ME to search the minimum RD cost in every reference frame using the corresponding IMV and decides the best inter prediction mode which denoted as J_{inter}_ref_m: m=0~3. Finally, the MRF-ME selects the best reference frame according to the lowest RD cost (J_{inter}_ref_min) among those four reference frames.

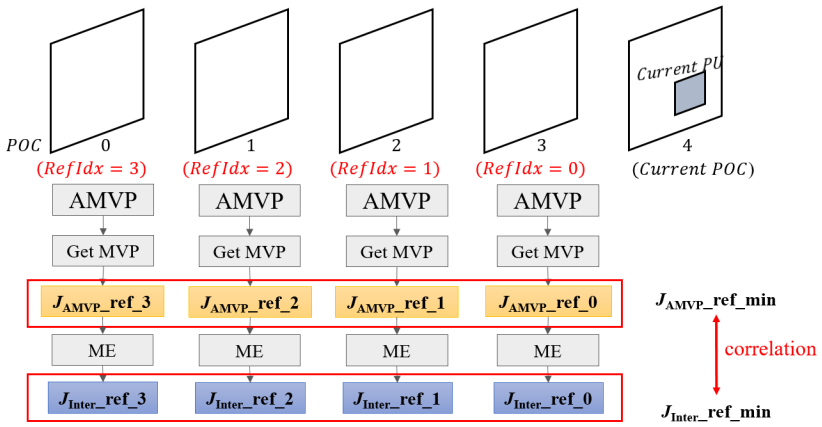

Fig. 4 The selectiong process of MRF-ME module on H.265 using four reference frames.

2.2 Fast reference picture selection algorithm (FRPSA)

Although the MRF-ME module improves the compression performance, the computational complexity increases in proportion to the number of reference frames. The ME is the most time-consuming computations in H.265 when performing exhaustive MV searching within the entire search range for all the reference frames. Since the main target resolution of H.265 is 4K/8K UHD videos, this leads to a big obstacle for real-time applications. To reduce the computational complexity of MRF-ME module in H.265, FRPSA proposed by Yang et al. found that there is a high correlation existing between the best reference frame (J_{inter}_ref_min) and lowest RD cost associated with AMVP(J_{AMVP}_ref_min) among four reference frames [4]. Their simulation results reveal that the corresponding frame with J_{AMVP}_ref_min is much more likely to be the optimal choice than the other reference pictures are. Since there is an average hit rate higher than 64% for J_{AMVP}_ref_min=J_{inter}_ref_min, the ME on the other reference frames can be avoided. This high probability indicates a strong correlation between J_{AMVP}_ref_min and J_{inter}_ref_min, which FRPSA takes these characteristics to make early decisions on selecting the optimal reference frame.

As shown in Fig.4, a predefined threshold is set by FRPSA, and *RefIdx* = 0 represents the reference frame temporally closest to the current frame, and the indices increase accordingly. FRPSA first calculates the minimum J_{AMVP}_ref_min value among the four reference frames. Suppose the smallest value occurs at J_{AMVP}_ref_2, and then it checks whether this value is below the predefined threshold. If it is, the ME module is executed only on *RefIdx* = 2 to obtain the corresponding J_{inter}_ref_2. If the minimum J_{AMVP}_ref_min value does not fall below the threshold, the ME module must be executed for all reference frames to calculate each J_{inter}_ref_min, after which the best reference frame is selected based on the lowest cost. However, since FRPSA does not only consider the correlation of optimal reference frames among neighboring blocks, but also explore the priority order of reference frames. Therefore, the overall coding performance and speed are consequently reduced.

III. PROPOSED METHOD

To further improve the performance of FRPSA, we utilize the correlation of RefIdx from neighboring blocks to design an early decision condition. In addition, we also propose a priority-based algorithm to select reference frame in advance. Furthermore, we combine both approaches into a unified fast reference frame decision algorithm to increase the encoding efficiency and speed of the reference frame selection process.

3.1 Neighboring-block-based reference frame decision algorithm (NRFDA)

In the PU structure, both the merge/skip mode (MSM) and the AMVP modes are utilized to predict characteristics of varying scene and object motion within the video content. Because H.265 performs ME on MRF structure to find the best reference frame (RefIdx_best) in each PU as shown in Fig. 2, this will lead to reduce the encoding performance of PU module. Natural video sequences have strongly spatial correlations, especially in the homogeneous regions. The best reference frame of a current CU is the same as the RefIdx_best of its spatially adjacent PUs due to the high correlation between neighboring CUs. Therefore, we analyze and calculate the spatial correlation values of RefIdx_best from the spatial neighboring blocks of the current CU as shown in Fig. 5.To employ the spatial correlation of RefIdx_best, we took a statistical analysis of the probability with same RefIdx_best for four spatial neighboring PUs in the current CU. Figure 6 shows the four neighboring best reference of performed PU for current CU (X), including left (A: ref_A), above left(B: ref_B), upper (D: ref_D) and right upper (E: ref_E), respectively. Figure 6 shows the statistical probability of occurrence with the same

RefIdx_best among four spatial neighboring Pus of the current CU, (i.e. ref_A=ref_B=ref_D=ref_E). From Fig. 7, we can find that the probability of occurrence with the same RefIdx_best in different depths exceeds 60% on average.

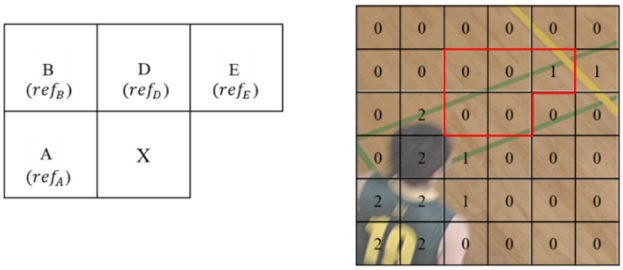

Fig. 5 The correlation of RefIdx_best in neighboring blocks.

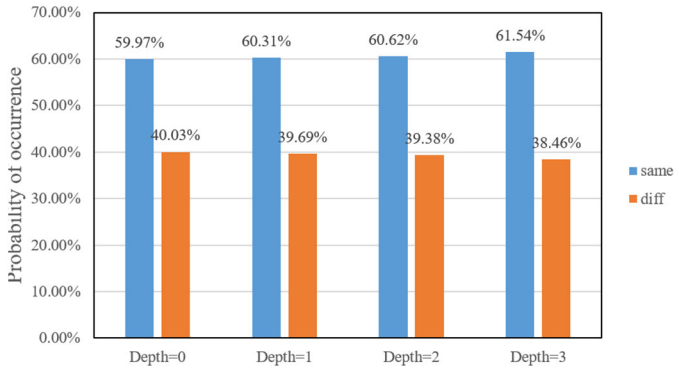

Fig. 6 The probability of occurrence with the same RefIdx_best in different depths.

Although the statistics show a high probability with the same reference frame among neighboring blocks, there remains approximately 40% with different reference frame. This mismatch implies differing RefIdx values which often indicating the presence of objects or significant motion within the block. In order to judge the object of rapid motion or object boundaries, we adopt the variance of the block as a measure for decision rule [6]. The mathematical expression is as follows:

$$\sigma_x^2 = \frac{1}{m \times n} \sum_{x=0}^{m} \sum_{y=0}^{n} (p(x,\ y) - \bar{m})^2 \tag{1}$$

where m and n denote the width and height of the current CU mode, $p(x,\ y)$is the pixel value at coordinate (x, y), and $\bar{m}$ represents the average pixel value of the PU block. If the variance (σ_x^2) is less than a predefined threshold (Thr_{NRFDA}), the block is considered to be neither on an object boundary nor rapid motion object. Conversely, if the σ_x^2exceeds Thr_{NRFDA}, the block is likely located on an object boundary or within a region of fast motion.

To maintain encoding efficiency, we set the predefined threshold which considers the average number of reference pictures (*AvgRefPic*). While different depths exhibit slightly varying threshold curves, the differences are minor. Figure 7shows the relationship curve between the overall *AvgRefPic* and Thr_{NRFDA} for depths ranging from 0 to 3.

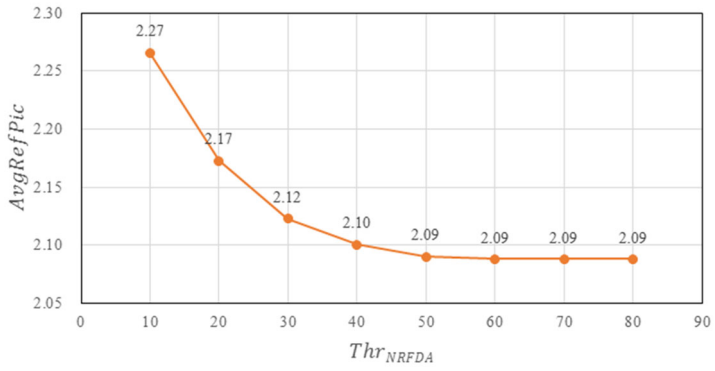

Fig. 7 The relationship curve between AvgRefPic and Thr_{NRFDA}.

3.2 Priority-based reference frame selection algorithm (PRFSA)

In the earlier discussion of FRPSA, we introduced the observed correlation between the J_{AMVP}_ref_min and J_{inter}_ref_min. Upon further analysis, we found a strong positive correlation between the entire sets of J_{AMVP} and J_{Inter} values after sorting. To illustrate this relationship, we consider an example with four reference frames. The relationship between J_{AMVP} obtained from AMVP process and J_{Inter} obtained from the ME module is shown in Table I. After performing MRF-ME module, if we sort the J_{AMVP} values in ascending order, and the corresponding RefIdx and J_{Inter} values also follow the same order. The sorted results are tabulated in Table II, which clearly show that as J_{AMVP} increases, J_{Inter} also tends to increase. This indicates that there is a consistent increasing trend between J_{AMVP} and J_{Inter}, revealing a positive correlation between the two metrics.

Through extensive statistical analysis across different coding unit depths, we found that the probability of a positive correlation between J_{AMVP} and J_{Inter} is approximately 59% on average. Based on these results, we propose the use of a predefined threshold (Thr_{PRFSA}) as a decision criterion for quickly selecting the reference frame.

TABLE I. Unsorted J_{AMVP} and J_{Inter}.

RefIdx	0	1	2	3
J_{AMVP}	2,500	2,000	100	1,500
J_{Inter}	5,000	3,000	1,000	2,500

TABLE II. Sorted J_{AMVP} values in ascending order.

RefIdx	2	3	1	0
JAMVP	100	1,500	2,000	2,500
JInter	1,000	2,500	3,000	5,000

Based on the above observations and statistical analysis, we first compute the J_{AMVP} values for each reference frame. These J_{AMVP} values are then used as the basis for sorting the reference frames, which determines the priority of reference frames to be processed by the ME module. For each sorted reference frame, if the corresponding J_{AMVP} is smaller than the predefined threshold (Thr_{PRFSA}), it is selected as the optimal reference frame. Otherwise, the algorithm proceeds to the next reference picture in the sorted list and repeats the process.

To maintain encoding efficiency, the threshold is set based on the average number of reference frames (*AvgRefPic*). Although the threshold varies slightly across different CU depths, the overall trend remains consistent. Figure 8 illustrates the relationship curve between *AvgRefPic* and Thr_{PRFSA} for Depth = 0. Furthermore, our experimental results show that similar correlation curves can be observed for other depth levels (Depth = 1 to 3) as well.

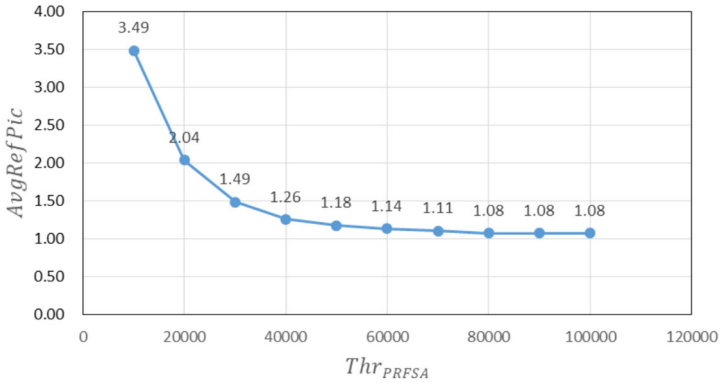

Fig. 8 The relationship curve between *AvgRefPic* and Thr_{PRFSA} for Depth = 0.

3.3 Fast reference frame decision algorithm (FRFDA)

To accelerate the H.265 encoding process, we propose a fast reference frame decision algorithm (FRFDA) by integrating NRFDA and PRFSA, aiming to significantly reduce the computational complexity of the ME-MRF module. The proposed method primarily exploits the spatial correlation between neighboring blocks to reduce the number of reference frames. In addition, it also utilizes the positive correlation between the RD cost obtained from the AMVP process which derived from the ME-MRF module. By sorting the RD costs of AMVP (J_{AMVP}), FRFDA prioritizes the evaluation of reference frames, thereby decreasing the number of ME operations required. The overall flowchart of the proposed FRFDA algorithm is illustrated in Fig. 9.

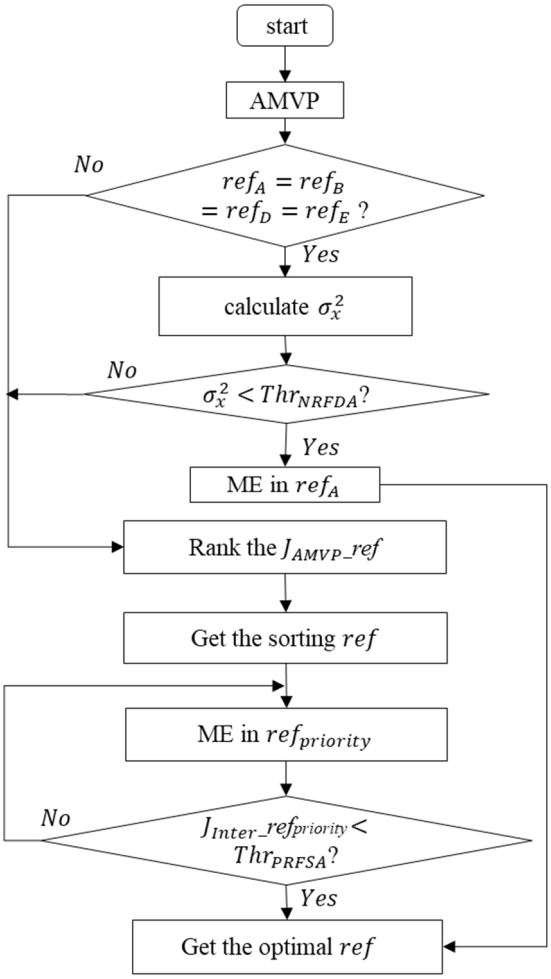

Fig. 9 The overall flowchart of the proposed FRFDA algorithm.

IV. EXPERIMENTAL RESULTS

To fairly evaluate the performance of the proposed FRFDA in comparison with FRPSA [4], both fast encoding algorithms were implemented and tested on the HM16.7 software platform [7]. A series of standard test video sequences were used for experiments and simulations. The encoding configuration is summarized as follows:

(1) Scenario: Low Delay (LD)
(2) QP = 22、27、32、37
(3) To be encoded frames: 48 frames
(4) Reference frames: 4 frames
(5) Standard test sequences: Traffic, PeopleOnStreet, ParkScene, Kimonol, BasketballDrive, BQMall, RaceHorses, BloowingBubbles

To evaluate the acceleration performance of each fast ME-MRF module, we adopt the time improving ratio (TIR) as the measurement metric. The mathematical definition is as follows:

$$TIR = \frac{AvgRefPic_{HM16.7} - AvgRefPic_{method}}{AvgRefPic_{HM16.7}} \quad (2)$$

where $AvgRefPic_{method}$ denotes the average number of reference pictures used in FRPSA or FRFDA, and $AvgRefPic_{HM16.7}$ represents the number of reference pictures used in HM16.7. Since Thr_{NRFDA} and Thr_{PRFSA} have significant impacts on the acceleration performance and the decoded video quality for ME-MRF module [8]. From experimental results show that the reduction in average reference pictures **(**AvgRefPic**)**starts to alleviate when $Thr_{NRFDA} \geq 50$ for *Depth*=0~3, and the decoded video quality remains nearly unchanged. Therefore, Thr_{NRFDA} is set to 50. Similarly, based on experimental observations, the values of Thr_{PRFSA} for different depths are set as follows: (Depth, Thr_{PRFSA}) = (0, 80000), (1, 20000), (2, 5000), and (3, 1250).

Tables III and IV present the average number of reference pictures and the TIR for FRPSA and FRFDA, respectively. From the tables, it can be observed that when MRF = 4, the proposed FRFDA reduces the average number of reference pictures by 2.73 compared to HM16.7, with an average TIR of approximately 68.23%. When compared to FRPSA, FRFDA reduces the number of reference pictures by approximately 0.99 on average and improves TIR by around 24.6%.

Moreover, the thresholds Thr_{NRFDA} and Thr_{PRFSA} used in this work are configurable. When the thresholds are set to zero, the results are identical to those obtained with HM16.7. Conversely, if Thr_{NRFDA}和Thr_{PRFSA} are adjusted to match the thresholds used in FRPSA, the proposed FRFDA achieves a similar video quality to FRPSA, while still providing better TIR performance.

TABLE III. Comparison of average number of reference pictures at MRF = 4.

Sequence	Average reference picture		
	HM16.7	FRPSA	FRFDA
Traffic	4	1.97	1.43
PeopleOnStreet	4	2.30	1.32
Kimonol	4	1.96	1.27
ParkScene	4	2.12	1.33
BasketballDrill	4	1.88	1.37
BQMall	4	2.04	1.07
ParkScene	4	2.85	1.13
RaceHorses	4	2.36	1.17
BQSquare	4	2.44	1.24
BloowingBubbles	4	2.65	1.37
Average	**4**	**2.26**	**1.27**

TABLE IV. Comparison of time improvement ratio at MRF = 4.

Sequence	TIR (%)	
	FRPSA	FRFDA
Traffic	50.75	64.25
PeopleOnStreet	42.63	67.32
Kimonol	51.28	68.25
ParkScene	47.00	66.75
BasketballDrill	52.96	65.75
BQMall	49.08	73.25
ParkScene	28.69	71.75
RaceHorses	41.06	70.55
BQSquare	39.08	69.32
BloowingBubbles	33.84	65.75
Average	**43.63**	**68.23**

REFERENCES

[1]. B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," in *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 31, no. 10, pp. 3736-3764, Oct. 2021.

[2]. G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," in *IEEE Tra*ns. Circuits System Video Technology, vol. 22, no. 12, pp. 1649- 1668, Dec. 2012.

[3]. High efficiency video coding, document ITU-T Rec. H.265, Oct. 2014.

[4]. S. H. Yang and K. S. Huang, "H.265 fast reference picture selection," Electronics letters, vol. 51, no. 25, pp. 2109–2111, Dec. 2015.

[5]. S. Wang and S. Ma, "Fast multi-reference frame motion estimation for high efficiency video coding," IEEE International Conference onImage Processing (ICIP), pp. 2005-2009, 15-18 Sept. 2013

[6]. A. N. Netravali and B. G. Haskell, Digital Pictures, 2nd ed., Plenum Press: New York, 1995, pp. 120-123.

[7]. Reference software HM16.7, https://hevc.hhi. fraunhofer.de/svn/svn_HEVCSoft are/branches/

[8]. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in Proc. 13th VCEG Meeting, pp. 1–5, Austin, TX, USA, 2001