Fast Skip Inter Coding Decision Algorithm for VVC

Xiangkai Hu[1](✉), Yue Li[1](✉), and Wenbin Lin[2](✉)

[1] School of Computer Science, University of South China, Hengyang 421001, China
kait@stu.usc.edu.cn
[2] School of Mathematics and Physics, University of South China, Hengyang 421001, China
lwb@usc.edu.cn

Abstract. Versatile Video Coding (VVC), which mainly introduces the QTMT partition structure and introduces new technologies in inter coding and intra coding, greatly improves the coding efficiency compared with the previous generation video coding standard. However, the improvement in coding efficiency comes at the cost of a huge increase in coding complexity. In order to reduce the computational complexity of VVC, this paper proposes a fast skip coding unit (CU) inter prediction algorithm in the VVC encoder. The algorithm mainly uses the difference between the original pixel value of the CU and its predicted value. After obtaining the difference between the original pixel value and the predicted value and performing a series of calculations, a threshold is set to weigh the loss of coding efficiency and the reduction of coding complexity. Extensive experimental results show that the proposed algorithm reduces the inter coding time by 12% to 60% while BD-Rate improves by less than 1% on average.

Keywords: Versatile Video Coding · QTMT · Inter prediction

1 Introduction

With the development of multimedia technology, new types of videos such as UHD, 4 K, and 360° have emerged. These new types of videos bring a better experience to people's visual perception. At the same time, these new types of videos also have a huge amount of data, which brings great challenges to the storage and transmission of videos. The previous generation of video standards, High Efficient Video Coding (HEVC), is difficult to meet the compression requirements of explosive video data [1]. In order to improve the compression efficiency of video, the Joint Video Experts Group (JVET) released a proposal for a new generation of video coding standards, and then started the standardization process of Versatile Video Coding (VVC), and completed the finalization of the first version of the standard in 2020. Compared with HEVC, VVC has higher coding efficiency and quality, but also greatly increases coding complexity.

G. Zhai et al. (Eds.): IFTC 2022, CCIS 1766, pp. 365–374, 2023.
https://doi.org/10.1007/978-981-99-0856-1_27

On the basis of the QT block partition structure of HEVC, VVC adds a binary tree partition structure and a ternary tree partition structure, which include horizontal binary tree partition, vertical binary tree partition, horizontal ternary tree partition, and vertical ternary tree partition, which are collectively referred to as multi-type tree (MTT), as showed in Fig. 1. VVC has a more advanced and flexible partition structure by using QT plus nested MTT (QTMT) partition structure, for example, sub-blocks can be square or rectangular, so as to achieve better compression coding effect. But at the same time, due to this more flexible partition structure, the coding complexity is also greatly increased.

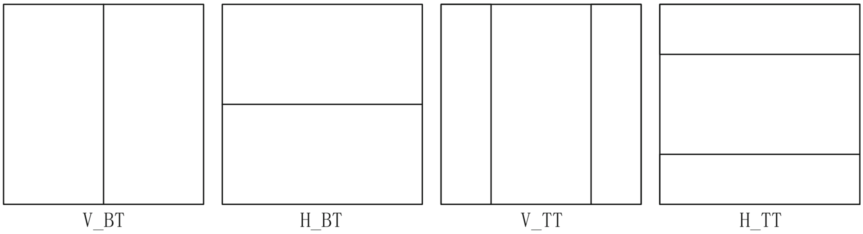
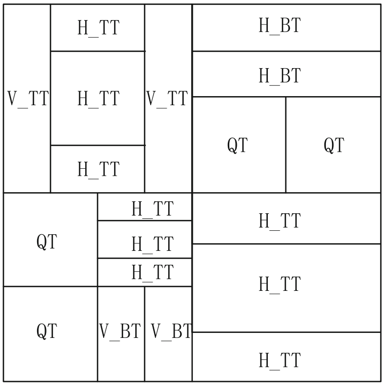


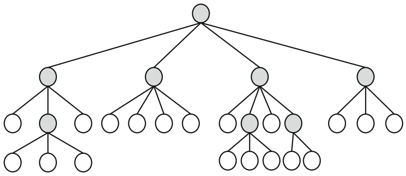
Fig. 1. MTT partition structure

In VVC, the QTMT partition structure determines the optimal partition structure for all blocks through a rate-distortion optimize (RDO) process [2]. It first regards the coding tree unit (CTU) as an undivided coding unit (CU), and after dividing the CTU with QT structure, the quad-tree leaf nodes are further divided by the recursive QT structure or recursive MTT structure. Figure 2 shows an example of dividing a CTU into multiple CUs with the QTMT structure. Note that if a node is divided with MTT structure, its child node can not be divided with QT structure. Due to this flexible QTMT partition structure, the complexity of the VVC Test Model (VTM) is many times that of the HEVC Test Model (HM), which greatly hinders the application of the VVC coding standard to real-time video coding [3,4]. Therefore, simplifying the QTMT structure can greatly reduce the coding complexity of VVC. In addition, VVC also introduces many advanced prediction technologies to improve inter prediction accuracy and VVC compression efficiency, such as Merge mode, affine transformation technology, adaptive motion precision technology, and so on. While these advanced prediction modes improve the accuracy of inter prediction, they also increase the computational complexity of inter prediction. Therefore, within the acceptable coding loss range, implementing some algorithms to conditionally skip the inter prediction mode can also effectively reduce the VVC coding complexity.

At present, the block-based inter prediction method used in mainstream video coding. The principle of this method is to find the reference block with the smallest difference from the current coding block in the adjacent reference reconstructed frames through motion estimation (ME). And take the reconstructed value of the reference block as the predicted value of the current block.



(a) Coding tree unit (CTU) partition structure



(b) Example of the hierarchy of the (a)

Fig. 2. Example of the optimal CTU partition structure

The displacement from the reference block to the current block is called a motion vector (MV). The process of using the reconstructed value of the reference block as a predicted value of the current block is called motion compensation (MC). Since there are generally many moving objects in the video, the prediction accuracy of simply using the value of co-located pixels block in the reference frame as the prediction value of current block is not high, so motion estimation is usually used to obtain the best matching block in the reference frame. Then the motion vector of the reference block and the current block are obtained, and finally the prediction value of the current block is obtained through motion compensation. The basic idea of the inter prediction technology in VVC is to use the temporal correlation between consecutive images of the video, which mainly use the block-based method to perform inter prediction by using the encoded image as the reference image of the current encoded image to remove the temporal redundancy of the image and improve the compression efficiency. The inter prediction in VVC follows and modifies some related technologies such as motion estimation, motion vector prediction and motion compensation in HEVC, and also introduces some new technologies to improve the performance of inter coding. For example, technologies such as extended merge prediction and merge mode with motion vector difference (MVD) are introduced to improve the accuracy of MV prediction. Bi-directional optical flow technology, CU Bi-prediction with CU-level weight and other technologies improve the accuracy of motion compensation. Affine motion compensated prediction technology is introduced to represent non-translational motion. So as to improve such as zoom-in, zoom-out, Motion under various viewing angles such as rotation and various irregular motion inter prediction performance.

This paper is composed of five parts. The Sect. 2 reviews some methods for reducing complexity in inter mode. In Sect. 3, a description of the proposed algorithm for fast skipping inter prediction mode. Then Sect. 4 gives the reference experimental results and test experimental results and their comparison results, and concludes this paper in Sect. 5.

2 Related Work

By inheriting and extending some of the inter prediction technologies in the previous generation video coding standard HEVC and introducing some new technologies, VVC improves the accuracy of its motion estimation and motion compensation, but also greatly increases the inter coding complexity. Based on this, some related methods of reducing the complexity of inter coding studied on HEVC can be applied to VVC, and most of the complexity problems of CU inter coding are related to motion estimation and motion compensation, and also related to the number of reference frames. Due to the complex and diverse motion of objects in the real world, the increase in the number of reference frames can effectively improve the accuracy of the coding results, but the corresponding computation and memory space will also be more complex, resulting in an increase in coding complexity. In order to effectively reduce the complexity of VVC inter coding, many research works have proposed various algorithms for inter coding. These algorithms can be roughly divided into the following three directions: simplifying the search pattern, adjusting the search range, and terminating motion estimation early [5]. The first two methods have been studied for many years, and corresponding experiments have been carried out in previous video coding standard test models. In the previous generation video coding standard HEVC, a diamond search algorithm [6] based on block matching motion estimation was proposed. This algorithm effectively simplifies the search mode on the premise of ensuring the coding quality, and has now become a core component of the motion estimation algorithm in VVC Test Model (VTM) and HEVC Test Model (HM). Reducing the number of searched reference frames is also an effective way to reduce the complexity of motion estimation. [7] proposed a fast reference frame selection method based on content similarity, which is used in the motion estimation of HM and effectively reduces the coding complexity of HM. [8] proposed a method for generating reference frames using a convolutional neural network structure, the basic idea of which is to use the encoded frames to generate virtual reference frames with neural networks instead of the reference frames selected from the reference frame sequence, to achieve the purpose of improving coding efficiency. Early termination of unnecessary motion estimation is also a simple and effective method to reduce coding complexity. Both [5,9] proposed methods to terminate motion estimation early. Among them, [5] skips some CUs' affine motion estimation (AME) by using the statistical features of MTT partition structure and AME to achieve the effect of reducing the spatial and temporal complexity of inter prediction with almost no coding loss. In [9], the texture space structure and temporal correlation of the image are used to terminate the further division of the CU in advance and optimize the affine motion, which saves the inter-coding time by reducing the complexity of motion estimation. Li, Lanlan, et al. [10] proposed a deep neural network structure model to predict motion vectors to reduce the complexity of motion estimation.

In order to reduce the computational complexity of VVC inter coding, a fast skip inter prediction decision algorithm is proposed in this paper. Unlike the previously discussed ways to reduce the computational complexity of inter-coding, this algorithm decides whether to skip inter-prediction for the current CU. When the CU's inter prediction is skipped, various inter prediction techniques including motion estimation, motion compensation, etc. will also be skipped at the same time, so the inter coding time can be greatly saved. The corresponding experimental reference software and test software in this paper have chosen VTM16.0 version. Compared with previous versions, many new methods are integrated in VTM16.0 to reduce the complexity of inter coding. However, the inter-coding complexity of the CU in VTM can still be further reduced by using the fast decision algorithm discussed in the next section. The main contributions and innovations of this paper are as follows:

1) A fast decision algorithm for skipping CU inter coding is proposed for VVC inter coding. To the best of our knowledge, our work is the first algorithm that attempts to skip the entire inter encoding mode of CU.
2) Compared with the work based on machine learning that needs to extract multiple features, the proposed algorithm only needs two features and can be directly obtained from CU. Because feature acquisition is more convenient and feature dimensions are reduced a lot, the decision algorithm is easier to implement in VVC and integrate into different video coding standards.
3) The proposed algorithm of fast skip inter mode is integrated into VTM, and the experimental results show that the inter coding time can be greatly reduced while the rate distortion performance can be neglected.

3 Proposed Method

In this part, a fast skip CU inter prediction decision algorithm is proposed to determine whether to skip the inter mode of the current CU, in order to achieve the goal of reducing the computational complexity of inter coding in VTM. For this algorithm, we mainly use the luminance difference between the original luminance value and the predicted luminance value of the CU as a feature to perform a series of computations. Then an algorithm model is constructed by using the obtained results as one of the decision conditions. Algorithm 1 shows the fast skip CU inter prediction decision algorithm in pseudo code. The feature selection and its calculation method are discussed in Sect. 3.1, and the decision algorithm is established in Sect. 3.2 and the overall flow chart of the algorithm is described.

Algorithm 1: Fast skip inter prediction decision algorithm

Input: μ, threshold, bestCS->cu[0]->skip
Output: skip inter prediction or not
  if μ>=threshold **then**
    return not skip inter prediction ;
  else if bestCS->cu[0]->skip == TRUE **then**
    return skip inter prediction ;
  else
    return not skip inter prediction ;
  end if
final ;

3.1 Feature Calculation

In order to make the coding loss of the fast skip CU inter prediction decision algorithm within an acceptable range, it is very critical to select appropriate CU features for calculation. In VVC, the optimal partition structure of a CU is related to the rate distortion (RD) cost. The CU with the smallest RD cost will be selected as the optimal partition structure. Therefore, if the current CU is the optimal partition structure, the minimum RD cost value can be obtained by calculation for its inter coding. However, when the CU has a high probability of not being the optimal partition structure, its inter coding is likely not to obtain the minimum RD and the inter coding should be skipped. Therefore, the selected features should be able to well reflect the optimal partition structure of CU. Since VVC uses a more flexible QTMT partition structure, the selected features should be more relevant to this new partition structure. Different from machine learning-based methods that need to use multiple feature information, and also different from previous work on fast inter coding that directly uses image texture or directly uses original CU luminance values, we try to select more efficient and more relevant features from CU. In previous work, the luminance value of a CU is usually selected as a feature to predict whether the partition of a CU is the best partition structure. Therefore, we decide to select the features related to the CU luminance value, and finally we choose the absolute difference between the CU's original luminance value and its predicted luminance value as the feature of the decision algorithm for calculation. At the same time, because the RD cost value calculated by the coding of the CU is related to the QP initial value of the experimental configuration parameter, the QP is also calculated as a feature of the decision algorithm. We then construct a temporary CU called Predicted CU with the same width and height as the original CU and its per-pixel value calculated from formula (3). Since the mean value can well reflect the concentration of a set of data, we use the mean value (μ) of the predicted CU as an effective decision condition for the fast skip inter prediction decision algorithm. It is calculated as:

$$Res_{abs\_i} = |pixel_{org\_i} - pixel_{pred\_i}| \tag{1}$$

$$QP_{level} = QP/RESIDUAL\_QP\_LEVEL \tag{2}$$

$$Pre\_CU_{pixel\_i} = Res_{abs\_i}\ QP_{level} \tag{3}$$

$$\mu = \frac{\sum_0^{h-1} \sum_0^{w-1} Pre\_CU_{pixel\_i}}{h \times w} \tag{4}$$

where $pixel_{org\_i}$ represents the original luminance value of the i-th pixel of the current CU, and $pixel_{pred_i}$ represents the predicted luminance value of the i-th pixel of the current CU. QP is the initial parameter value configured before VTM encoding. h and w represent the height and width of the CU respectively. In order to assign different levels to the initial QP parameter values, we introduced a parameter $Residual\_QP\_Para$ and set its value as 9 in the experiment. Finally, μ represents the average pixel value of the predicted CU, which is used to determine whether the decision algorithm will skip the inter coding.

3.2 Established Decision Algorithm Model

According to formula (1–4), the average μ of the predicted CU can be calculated. When μ value is large, it indicates that the current encoded CU may be the smallest RD cost value, and then the current CU inter coding mode performs. Therefore, we set a threshold value of 10 to define μ. When μ is greater than or equal to the threshold, we continue to execute the current CU inter encoding. In order to ensure that the coding loss is within the acceptable range, the algorithm should be more careful to decide whether to skip the inter prediction. Therefore, we also select the skip attribute of the first CU in the best coding structure as another condition for building the decision algorithm model. When μ is less than the threshold, if Skip is True, the current CU inter encoding mode is skipped; otherwise, the current CU inter coding mode is continued. The overall flow chart of the algorithm is shown in Fig. 3. Because the proposed decision algorithm is used to determine whether to skip the inter mode, the algorithm is executed when the CU encoding mode is ETM_INTER_ME.

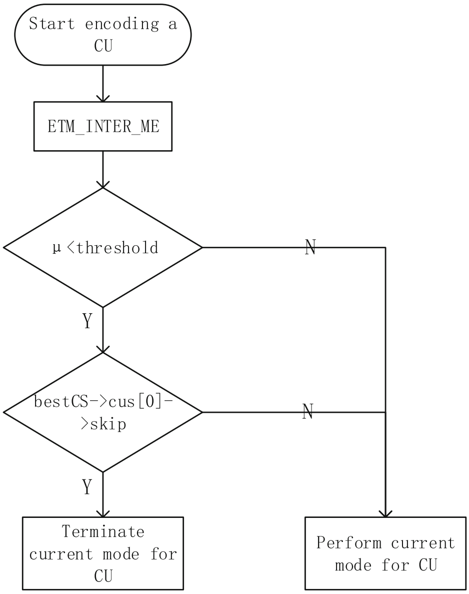


Fig. 3. Flowchart of the proposed method.

4 Experiments and Results

4.1 Experimental Configurations

In order to verify the effectiveness of the proposed decision algorithm for fast skip CU inter prediction, the experiment was integrated into VTM16.0 and performed under Random Access(RA) configuration. A total of 20 standard test video sequences from Class A to Class E were used for experimental testing. Note that the resolution of the video sequence file varies in different class. The encoded standard test video sequences all use four Quantization Parameters (QP), whose QP values are set to 22, 27, 32, and 37. All the experimental results were obtained by encoding a 30-frame video sequence, and the widely used BD-Rate [11] and encoding time saving rate were used to measure the performance of the proposed algorithm. The encoding time saving rate is defined as:

$$TS = \frac{T_{ref} - T_{test}}{T_{ref}} \times 100\% \tag{5}$$

where T_{ref} and T_{test} represent the total inter coding time of the four QP values of the original VTM16.0 and the total inter coding time of the four QP values of the VTM16.0 using the proposed algorithm respectively.

4.2 Performance Comparison

Tested in Table 1 shows the comparison between the proposed decision algorithm and VTM16.0. Tested represents the fast skip inter prediction decision algorithm proposed in this paper. The results show that the fast skip inter coding decision algorithm can save inter coding time from 12.44% to 60.67%, the average saving inter coding time is 35.17%, while the BD-rate only increases 0.98%. Experimental results show that the proposed algorithm can effectively save CU inter coding time while the increase of BD-rate is almost negligible. In order to further prove the effectiveness of the proposed decision algorithm, the fast skipCU affine motion part of the proposed method [5] was integrated into VTM16.0 for testing, and Park [5] showed its experimental results. It can be seen that although the fast skip affine motion method does not increase the BD-Rate, the time saved by inter coding is only from 6.72% to 8.98%, with an average of 7.74%. This further proves that the proposed method can greatly reduce the complexity of inter coding.

Table 1. Performance of proposed method compared with VTM 16.0 (UNIT: %)

Class	*Sequence*	Park [5]		Tested	
		BD-rate	TS	BD-rate	TS
A	BasketballPass	0.00	8.44	0.64	37.06
	BlowingBubbles	0.00	8.00	1.17	33.32
	BQSquare	0.00	7.85	1.28	49.12
	RaceHorses	0.00	7.17	1.41	25.09
B	BasketballDrill	0.00	8.45	1.04	25.68
	BQMall	0.00	7.82	0.87	36.61
	PartyScene	0.00	6.72	0.56	23.62
	RaceHorsesC	0.00	6.98	1.25	22.67
C	ChinaSpeed	0.00	7.41	0.98	22.35
D	FourPeople	0.00	8.92	0.66	52.44
	Johnny	0.00	8.47	2.49	59.60
	KristenAndSara	0.00	8.29	1.48	51.00
	SlideEditing	0.00	6.91	0.34	60.67
	SlideShow	0.00	7.69	-0.05	38.52
E	BasketballDrive	0.00	7.77	0.95	25.38
	BQTerrace	0.00	6.78	1.64	44.46
	Cactus	0.00	7.80	0.59	31.98
	Kimono	0.00	8.03	0.48	21.09
	ParkScene	0.00	7.13	0.79	44.03
	RitualDance	0.00	8.19	0.99	12.44
	Average	**0.00**	**7.74**	**0.98**	**35.17**

5 Conclusions

In the above work, this paper proposed a fast algorithm to skip CU inter prediction for VVC. In order to reduce the coding complexity of VVC, the algorithm uses two conditions to decide whether to skip the inter prediction mode of the current coding block. The experimental results show that the fast skip inter prediction algorithm can effectively reduce the VVC coding complexity and save the inter coding time, while the BD-Rate only increases slightly.

In the following work, the fast skip inter prediction algorithm can be combined with other modes such as intra mode to further reduce the VTM encoding time; It can also be combined with deep learning and other methods to quickly determine the CU partition structure, so as to further reduce the coding complexity and reduce the coding time while ensuring a small coding loss.

References

1. Yun, S., et al.: An efficient low-complexity block partition scheme for VVC intra coding. J. Real-Time Image Process. **19**(1), 161–172 (2022)
2. Tang, N., et al.: Fast CTU Partition Decision Algorithm for VVC Intra and Inter Coding. In: IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), vol. 2019, pp. 361–364 (2019). https://doi.org/10.1109/APCCAS47518.2019.8953076
3. Tissier, A., et al.: Complexity reduction opportunities in the future VVC intra encoder. In: 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP). IEEE (2019)
4. Pan, Z., et al.: Low complexity versatile video coding for traffic surveillance system. Int. J. Sens. Netw. **30**(2), 116–125 (2019)
5. Park, S.-H., Kang, J.-W.: Fast affine motion estimation for versatile video coding (VVC) encoding. IEEE Access **7**, 158075–158084 (2019)
6. Zhu, S., Ma, K.-K.: A new diamond search algorithm for fast block-matching motion estimation. IEEE Trans. Image Process. **9**(2), 287–290 (2000)
7. Pan, Z., et al.: Fast reference frame selection based on content similarity for low complexity HEVC encoder. J. Vis. Commun. Image Representation **40**, 516–524 (2016)
8. Lee, J.-K., et al.: Deep video prediction network-based inter-frame coding in HEVC. IEEE Access **8**, 95906–95917 (2020)
9. Guan, X., Sun, X.: VVC fast ME algorithm based on spatial texture features and time correlation. In: 2021 International Conference on Digital Society and Intelligent Systems (DSInS). IEEE (2021)
10. Li, L., et al.: Deep motion vector prediction for versatile video coding. In: 2021 7th International Conference on Computer and Communications (ICCC). IEEE (2021)
11. Gisle, B.: Calculation of average PSNR differences between RD-curves. In: VCEG-M33 (2001)