Optimal Mode Decision Method for Interframe Prediction in H.264/AVC

Hongjin Zhu[1], Honghui Fan[1, *], Zhenqiu Shu[1], Congzhe You[1], Xiangjun Chen[1], Qian Yu[1] and Pengzhen Gan[2]

Abstract: Studies show that encoding technologies in H.264/AVC, including prediction and conversion, are essential technologies. However, these technologies are more complicated than the MPEG-4, which is a standard method and widely adopted worldwide. Therefore, the amount of calculation in H.264/AVC is significantly up-regulated compared to that of the MPEG-4. In the present study, it is intended to simplify the computational expenses in the international standard compression coding system H.264/AVC for moving images. Inter prediction refers to the most feasible compression technology, taking up to 60% of the entire encoding. In this regard, prediction error and motion vector information are proposed to simplify the computation of inter predictive coding technology. In the initial frame, motion compensation is performed in all target modes and then basic information is collected and analyzed. After the initial frame, motion compensation is performed only in the middle 8×8 modes, and the basic information amount shifts. In order to evaluate the effectiveness of the proposed method and assess the motion image compression coding, four types of motion images, defined by the international telecommunication union (ITU), are employed. Based on the obtained results, it is concluded that the developed method is capable of simplifying the calculation, while it is slightly affected by the inferior image quality and the amount of information.

Keywords: Video compression, prediction mode decision, H.264/AVC, interframe prediction.

1 Introduction

With the advance of the broadband in the past few years, various video distribution services, including the videophone, video distribution in conferences, and content for mobiles have been developed remarkably [He, Wu, Zeng et al. (2013); Major (2016)]. Studies show that one of the essential technologies for spreading the abovementioned video distribution services is video information compression coding technology (i.e., video coding method) [Joshi and Jain (2020); Nguyen and Tropchenko (2018)]. It is

[1] College of Computer Engineering, Jiangsu University of Technology, Changzhou, 213001, China.

[2] Department of Technology Development, Management System Integrator, Yamagata, 990-2473, Japan.

[*] Corresponding Author: Honghui Fan. Email: fanhonghui@jsut.edu.cn.

Received: 31 May 2020; Accepted: 25 July 2020.

worth noting that in order to transmit a video through the network, constraints of the network and the encoded (compressed) bitstream should be satisfied. Therefore, H.264/AVC standard was established in 2003 by ITU-T and ISO/IEC in the latest international standard compression coding system for moving images [Joshi, Vivekanand, Rajendra et al. (2016); Fan and Zhu (2018)]. In H.264/AVC, the encoding technologies, including the prediction and conversion [Pan, Jin, Lei et al. (2016); Dalal and Juneja (2018)] are the most essential technologies. Studies show that the H.264/AVC standard is significantly more sophisticated than MPEG-4 from different aspects such as the international standard compression encoding method [Yuan and Hu (2019); Molina, Gazzano, Rincon et al. (2018)]. Since the inter prediction technology has the maximum effect on xxx among all essential technologies in the H.264/AVC [Yao, Zhang and Yu (2016)], it shares nearly 60% of the entire coding process. Accordingly, simplifying the computation of the inter prediction coding technology is highly demanded [Xu and Wang (2016); Chen, Zhong and Bao (2019)].

The fundamental technology in the H.264/AVC consists of predictive coding technology. In video signals, neighboring pixels in both spatial (intra-frame) and temporal (inter-frame) directions are tightly correlated with each other [Shang, Zhao, Wang et al. (2019)]. Moreover, the technique, which conducts such prediction and encodes the prediction difference from the true value, is called the predictive coding [Conti, Soares and Nunes (2018)].

H.264/AVC has inter-prediction (inter-frame prediction) as one of the predictive coding techniques that conduct temporal prediction [Choe, Nam and Chu (2018)]. Since a moving image contains at least 30 frames per second, the elapsed time between two consecutive frames is less than 1/30 [sec]=33 [msec], which is a very short time. Therefore, there are many scenarios that the subject exhibits a small moving distance between two consecutive frames. The correlations between screens are considerably high. This is especially more pronounced when the user moves (e.g., in video conferencing and video-phone signals). On the other hand, no difference is identified in the image signal between two consecutive frames of a stationary object (e.g., walls of the conference room in the background of a person).

In the inter-prediction, the conventional MPEG-4 format is normally applied to perform the motion compensation in two modes of 16×16 and 8×8, while the H.264/AVC is capable of performing the motion compensation in seven modes, including 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4. However, the calculation in the latter scheme is inevitably complicated since the mode is ascertained by determining the mode that exhibits the minimum overall information amount of macro error block and the predicted error value [Zhu, Fan, Shu et al. (2019); Fan and Zhu (2018)].

In order to simplify the computation of the mode determination, it is intended to propose a method for determining the prediction mode in H.264/AVC inter prediction with no scene change. In other words, the proposed method is applicable when the statistical properties of video do not vary significantly along the way. In the initial frame, the motion compensation is conducted in all target modes, while different basic information amounts are collected and analyzed. Then, motion compensation is achieved only at 8×8, where the middle of the mode and the basic information amount is shifted. Prediction error values other than 8×8 and motion vector information are estimated. In the proposed method, after the initial frame, the motion compensation is basically performed with 8×8 as the maximal priority. It is expected that the proposed method can significantly simplify the processing computation.

2 Inter predictive coding with motion compensation

In the video compression coding technology (e.g., H.264/AVC), motion estimation is carried out to down-regulate the amount of information in the motion region. The difference is obtained from the motion estimation predicted frame. Such prediction through the motion estimation is called the inter-frame prediction with motion compensation.

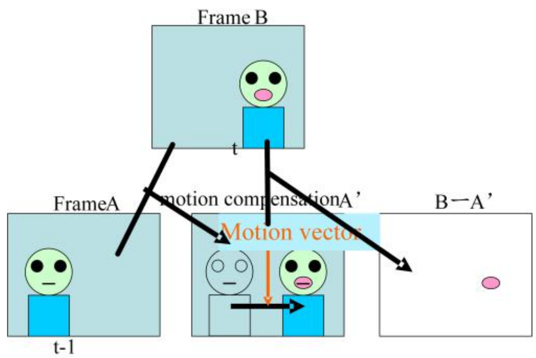


Figure 1: Difference between motion compensation frames

Fig. 1 shows an example of the inter-frame prediction with the motion compensation. It should be indicated that the motion vector refers to the amount that the person corresponding to the moving region moves, while the moving direction and number of pixels in the frames B and A are the same. Moreover, the motion-compensated image (see frame A') is plotted. By calculating differences between frames A' and B, the information in the moving region can be reduced. In the presented motion compensation, one frame in the past is adopted. However, in H.264/AVC, the difference between the frames that have been motion-compensated in the past 2 frames are considered and frames. In other words, a total of 4 frames are considered in H.264/AVC. It should be indicated that in the present study, the experiment is performed with the motion compensation in the last frame.

In the block matching, the motion of the moving region can be accurately estimated by splitting the image into multiple blocks. Studies show that this can enhance the accuracy of the motion estimation. During the block matching, the search is performed by shifting the range of ±16 along the *x*- and *y*-directions from the block at the identical position in the target frame and the reference frame. Therefore, a difference is taken in the search range, which is called the prediction error value. When the prediction error value is the minimum in the search range, the degree of deviation without the search range can be expressed as a vector by splitting it into *x*- and *y*-components and then such the created vector is called the motion vector. During the inter prediction process, the prediction error value and the motion vector are encoded. The search in the block matching exhibits so sophisticated computation that it increases the complexity of the majority of the inter predictions.

Fig. 2 indicates that in the H.264/AVC, when block matching is performed in the search range for estimating the motion, a spiral search is conducted for the outside of the center. The spiral search exploits the fact that a block with a small prediction error value is probably close since the motion range of the motion region between adjacent frames is commonly small.

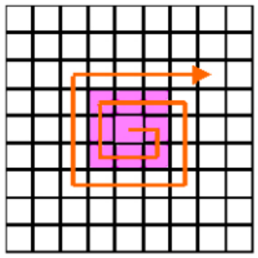


Figure 2: Spiral search

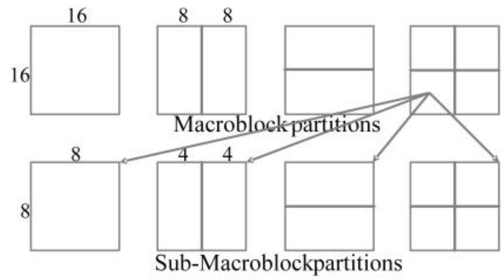


Figure 3: Variable block size in H.264/AVC

In the conventional international standard compression encoding MPEG-4, block matching is carried out in two modes, including 16×16 and 8×8 pixels. Fig. 3 shows that in H.264/AVC, the block can be switched between seven modes in units consisting of 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 pixels.

It should be indicated that the computational complexity of H.264/AVC is 3.5 times higher than that of the MPEG-4. Therefore, it is found that the proposed method simplifies the computation by the variable block size. Fig. 5 presents the result of adapting the block size in H.264/AVC inter prediction to the original image, as shown in Fig. 4. Moreover, it is found that for regions with high motions, the motion compensation with a small block size improves the accuracy of the motion estimation and down-regulates the prediction error value.

Figure 4: Original image

Figure 5: Block size adaptive image

3 Prediction error and motion vector information

For the block matching, in the motion estimation for the motion compensation, a position where a prediction error is minimized is represented by a motion vector. Such a motion vector should be encoded and transmitted. In terms of the inter prediction, the difference between the motion-compensated reference frame and the target frame that should be encoded is harvested and then it is encoded as a prediction error value to reduce the information.

Firstly, the average amount of information is calculated for the information amount of prediction error. Fig. 6 presents the results of calculating the average amount of prediction error per frame. The calculation of the average amount of the prediction error E_e can be expressed by calculating probability P_e of the prediction error value per pixel.

$$E_e = -\sum P_e log_2 P_e \tag{1}$$

Where Σ is the sum operator over all pixels per frame (288×352). In Fig. 6, the horizontal and vertical axes represent the block size and the average amount of prediction error per pixel, respectively. It should be indicated that Cheer, Flower, Football, and Mobile refer to four types of sequences that are international standard test moving images specified for experiments. It is found that since the block size is small, the corresponding average amount of information for prediction error in a respective sequence is small. This is because the smaller the block size, the better the accuracy of the motion estimation. Therefore, a large distribution can be conducted with small prediction errors.

The motion vector average information E_v can be mathematically expressed as Eq. (2) adopting probability P_v of the norm $\sqrt{v_x^2 + v_y^2}$ of the motion vector. The calculated information refers to the average amount of information per norm motion vector. Since the block matching is performed in units of macroblocks with 16×16 pixels, a single motion vector is adopted for the block size of 16×16. Moreover, since the block size of 4×4 is capable of splitting the 16×16 pixel unit into 16 parts, 16 motion vectors are detected. Accordingly, the calculated average amount of information is multiplied by 16 for the block size of 4×4.

$$E_v = -\sum P_v \log_2 P_v \tag{2}$$

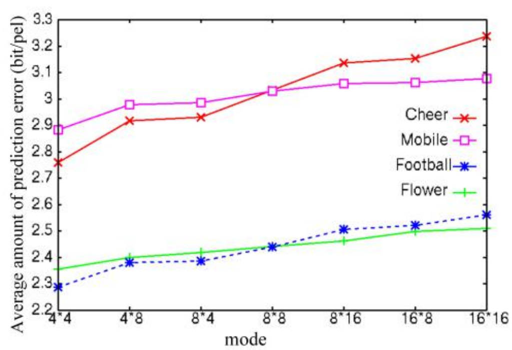


Figure 6: The average amount of prediction error information for different modes

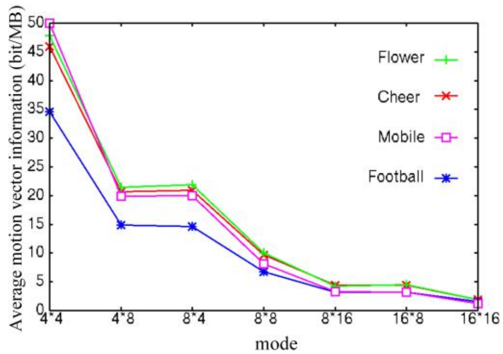


Figure 7: The average amount of motion vector information for different modes

In Figs. 6 and 7, the horizontal and the vertical axes display the block size and the average amount of the motion vector of norm per 16×16 pixels, respectively. It is observed that in the respective sequence, the average amount of motion vector information increases as the block size decreases, which is the opposite of the average amount of information in the prediction error.

It should be indicated that since the block size is small (e.g., 4×4), the amount of the prediction error information is down-regulated, while the motion vector information is up-regulated. In contrast, with a block size of 16×16, the information for prediction errors improves, while the information for motion vectors reduces. It should be indicated that when examining the block size that is capable of reducing information, the sum of the prediction error and the amount of information of the motion vector should be considered. Accordingly, in H.264/AVC, the amount of information, the sum of prediction error and the amount of motion vector information are considered. Accordingly, the cost of the mode decision, and a cost function that calculates the minimum cost is defined.

4 Proposed method for determining the prediction mode of Inter prediction

In the proposed method, a sequence of moving images without scene change is targeted to simplify the computation of the inter prediction. In this method, the identical moving region (e.g., a person) always exists on the screen for each frame of the moving image. When inter prediction is applied for the respective frame in a static region (e.g., the background), a moving image with no scene change has a negligible prediction error value. In a motion region (e.g., a region containing a moving person), the prediction error and the motion vector value are very similar to the block-matched block for the respective frame so that changes in the amount of information are tightly correlated. This can be considered the reason why the amount of change in information is high. In fact, it originates from the short time between successive frames for moving images.

In the present study, an experiment is conducted for three block sizes, including 4×4, 8×8, and 16×16. In the prediction mode decision of the proposed inter prediction, the correlation between the prediction error and the amount of change in the motion vector information for each frame is high. Furthermore, the amount of information changes in the second frame, where the second and third frames act as the reference frame and the target frame, respectively. In this article, a novel method is proposed to estimate the amount of information using the data between frames. When estimating moving vector in the second frame, the estimation is performed based on the initial data change and the obtained 8×8 inter prediction results. The reason for estimating the other modes exploiting the 8×8 inter prediction results is that they are located in the middle of 4×4 and 16×16 networks, so the difference in the amount of information with other modes is treated equally.

This study focuses on the fact that the correlation between changes in the amount of information is high, and converting the prediction error and the amount of change in the information amount of the motion vector into data in the initial frame. To this end, an innovative method is proposed to estimate the amount of information using motion vectors, prediction errors in the initial frame, and variations in the amount of information in motion vectors. However, as the first stage of the experiment, 4×4 block size that splits the macroblock into seven modes, the 16×16 block size that does not divide. In fact, the experiment is performed in three modes of 8×8 block size.

In the initial frame, motion compensation is achieved in three modes, including 4×4, 8×8, and 16×16. Then occurrence probability of prediction error value and motion vector value is obtained, respectively. Such occurrence probability is obtained to convert the prediction error value and motion vector value at the decision cost of the prediction mode decision into the information amount.

To this end, the following operations should be carried out: First, probability of prediction error is expressed in the initial frame. It is assumed that e_4 , e_8 and e_{16} represent 4×4, 8×8 and 16×16 prediction errors, respectively. Moreover, $e_{4,8,16}(x,y)$ denote the prediction error value per pixel in each of the three modes corresponding to the luminance value per pixel on the screen. Assuming that the luminance value per pixel in the screen is $I(x,y,t)$, and $\vec{v}_{x4,x8,x16}$ and $\vec{v}_{y4,y8,y16}$ are the *x*- and *y*-components of the motion vector, respectively, the prediction error value can be expressed as:

$$e_{4,8,16} = \left| I(x,y,t) - I(x+\vec{v}_{x4,x8,x16}, y + \right| \vec{v}_{y4,y8,y16}, t-1) \tag{3}$$

On the other hand, when the frequency $h(e_4,e_8,e_{16})$ in the initial frame is calculated for the prediction error value $e_{4,8,16}(x,y)$, prediction error occurrence probability $P_{e_4,e_8,e_{16}}(e_4,e_8,e_{16})$ can be calculated in the form below:

$$P_{e_4,e_8,e_{16}}(e_4,e_8,e_{16}) = \frac{h(e_4,e_8,e_{16})}{N\times N} \tag{4}$$

where *N* and *M* are the number of pixels on the vertical and horizontal axes, respectively. Fig. 8 illustrates the occurrence probability of a prediction error for one Cheer in a video sequence. The horizontal and vertical axes are the prediction error per pixel and occurrence probability, respectively. In is observed that when the maximum prediction error value reaches 255, most of the distributions are concentrated on values below 20.

Assuming a 4×4 norm motion vector $\vec{v}_4$, an 8×8 norm motion vector $\vec{v}_8$, and a 16×16 norm motion vector $\vec{v}_{16}$, the norm motion vector $\vec{v}_{4,8,16}(x,y)$ for each of the three modes can be expressed as Eq. (5).

$$\vec{v}_{4,8,16}(x,y)=\sqrt{\vec{v}^2_{x4,x8,x16}+\vec{v}^2_{y4,y8,y16}} \tag{5}$$

where $\vec{v}_{x4,8,16}$ and $\vec{v}_{y4,8,16}$ are x- and y-components of the motion vector, respectively. Moreover, (x, y) refer to the coordinates in macroblock units. If the frequency (histogram) $h(\vec{v}_4,\vec{v}_8,\vec{v}_{16})$ of the norm motion vector in one frame is calculated, probability $p_{\vec{v}_4,\vec{v}_8,\vec{v}_{16}}(\vec{v}_4,\vec{v}_8,\vec{v}_{16})$ of the norm motion vector can be expressed in the form below:

$$P_{\vec{v}4,\vec{v}8,\vec{v}16}(\vec{v}_4,\vec{v}_8,\vec{v}_{16})=\frac{h(\vec{v}_4,\vec{v}_8,\vec{v}_{16})}{vectors\ of\ 1Frame} \tag{6}$$

Fig. 8 presents the probability distribution of the motion vector for one Cheer in a video sequence. The motion vector length of the norm is $\sqrt{16^2+16^2}=16\sqrt{2}$ when both the x- and y-components are set to 16. It is observed that the distributions approach 0 rapidly.

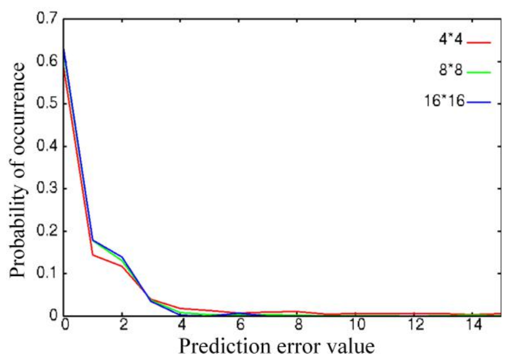


Figure 8: Probability distribution of the motion vector occurrence

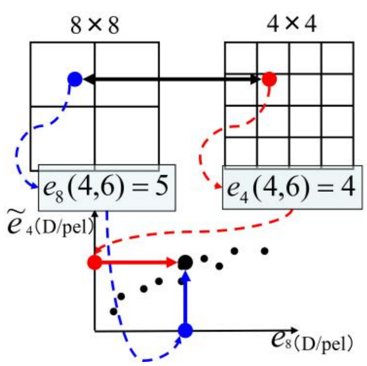


Figure 9: The prediction error data

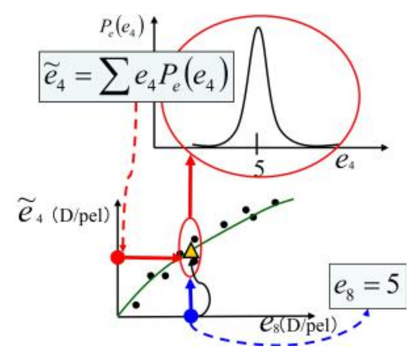


Figure 10: The conversion prediction error

Fig. 9 illustrates the flow of the basic data collection for the prediction error estimation. 4×4 and 8×8 refer to prediction error values per pixel of the identical coordinates at the initial frame when the motion compensation is performed. In this example, it indicates that that 4×4 and 8×8 prediction error values are $e_4(4,6)=4$ and $e_8(4,6)=5$, respectively.

Accordingly, a two-dimensional plane is plotted with $e_4(4,6)=4$ on the *y*-axis and $e_8(4,6)=5$ on the *x*-axis. This data represents the correlation between 8×8 and 4×4 prediction error values, indicating the basic data for prediction error estimation. Since plotting is performed for all pixels in the initial frame, the number of plots is evidently large, while the corresponding distribution is significant.

Fig. 10 illustrates the flow for making the basic data to achieve prediction error estimation into a curve. Assume that there are basic data for prediction error estimation plotted for one frame. For instance, if an 8×8 prediction error value $e_8=5$ is concentrated in the basic data for estimating prediction error, the 4×4 prediction error value e_4 will exhibit innumerable distribution. Consequently, occurrence probability P_e of the distribution e_4 should be found at the value of $e_8=5$ and then prediction error of 4×4 (as the weighted average value) is converted through Eq. (7). Accordingly, occurrence probability P_e of the distribution of e_4 for $e_8=5$ is identified, and the prediction error of 4×4 is converted as the weighted average value $\tilde{e}_4$.

$$\tilde{e}_4 = \sum e_4 \log_2 P_e(e_4) \tag{7}$$

The 4×4 and 16×16 prediction error values are estimated from the basic data of the prediction error estimation, thereby indicating the correlation with 8×8 in the initial frame. It is concluded that the 8×8 prediction error value can be ascertained in the second frame. Fig. 11 illustrates the prediction error estimation for the second and later frames of 4×4. In the second and subsequent frames, motion compensation is conducted at 8×8, and the actual prediction error value is ascertained. Then 8×8 prediction error value $e_8=5$ is detected as shown in Fig. 11. A prediction error of 4×4 is estimated from the basic data for prediction error estimation obtained even in the initial frame. However, if $e_8=5$ is assigned to the *x*-axis, then the *y*-axis is 4×4 with the same coordinates as 8×8. The estimated prediction error value $\tilde{e}_4(4,6)=4$ of 4×4 is achieved. As a result, estimation is performed from the prediction error value of 8×8, and a prediction error value of 4×4 is estimated at the identical coordinates of the prediction error value.

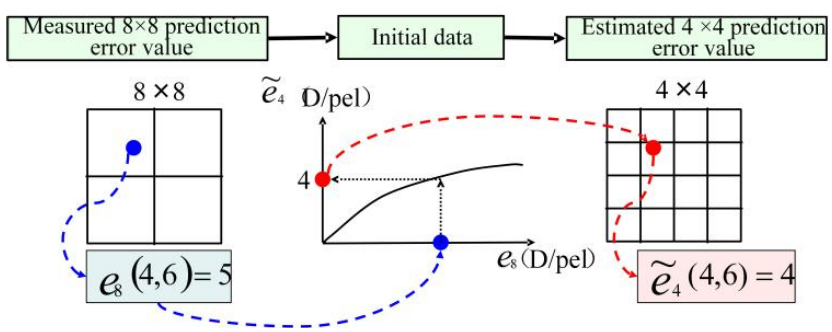


Figure 11: 4×4 prediction error value estimation flow from 8×8 prediction error value

Fig. 12 shows the flow of the basic data collection for the motion vector estimation. The average prediction error e_8 is calculated in 8×8 pixel units of 8×8 block size in the initial frame. Moreover, Eq. (8) presents that the average value e_4 of prediction error is calculated in 8×8 pixel units of 4×4 block size. The value of increment in the motion vector is calculated using Eq. (9) and plotted on the x-axis. The plot indicates the correlation between the increment in prediction error and the increment in motion vector for 8×8 and 4×4.

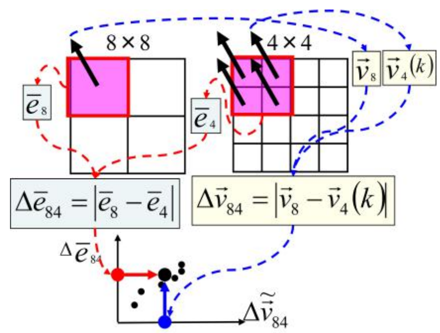


Figure 12: Flow of motion vector estimation basic data acquisition

$$\bar{e}_{84}(k) = \left|\bar{e}_8(k) - \sum_{l=0} \bar{e}_4(k,l)\right| (0 \le k \le 3, 0 \le l \le 3) \tag{8}$$

$$\bar{e}_{84}(k) = |\vec{v}_8(k) - \vec{v}_4(k,l)| (0 \le k \le 3, 0 \le l \le 3) \tag{9}$$

The distribution of the basic data for the motion vector estimation is converted by plotting. Fig. 13 illustrates the flow for converting basic data for the motion vector estimation.

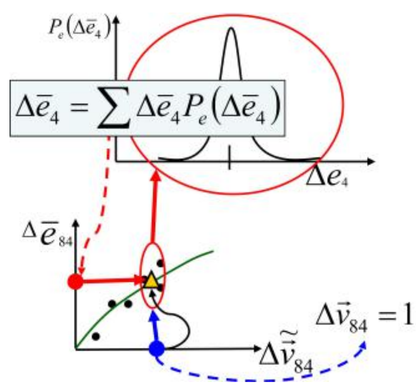


Figure 13: Conversion motion vector change data calculation flow

Fig. 13 shows that there is data expressing the increase in prediction error as well as the increase in motion vector for one frame. The 4×4 motion vector consists of 8×8 and 4×4 motion vector estimation basic data representing the correlations between the prediction error increase and the motion vector increase in the initial frame; the 8×8 estimation is performed from 8 prediction error values, motion vectors, and 4×4 estimated prediction error values. Fig. 14 elucidates the flow of the motion vector estimation from the second frame of 4×4.

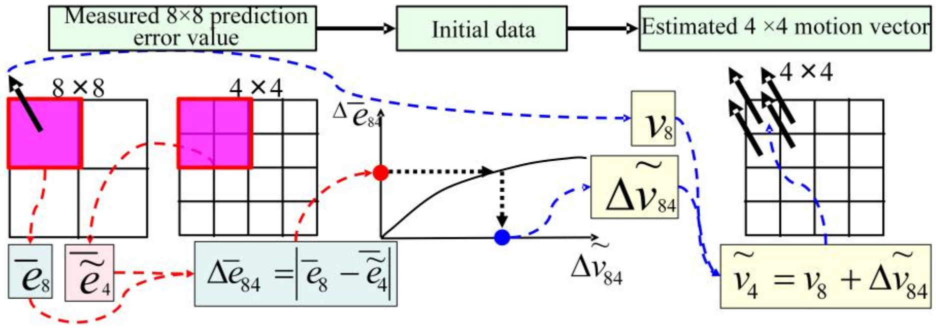


Figure 14: Motion vector estimation flow

Eq. (10) presents that the value of the norm motion vector is the minimum when there is only one component. On the other hand, Eq. (11) denotes that the value of the norm motion vector is the maximum. In order to split the norm motion vector into its components, it is assumed that the value of the norm motion vector is the highest in Eq. (11), and the components in the x and y directions are approximated by Eq. (12).

$$\left|\vec{V}\right|_{min} = min(v_x, v_y) \tag{10}$$

$$\left|\vec{V}\right|_{max} = \sqrt{v_x^2 + v_y^2} \tag{11}$$

$$v_x = v_y = \frac{\vec{V}}{\sqrt{2}} \tag{12}$$

Accordingly, if the amount of the information of the norm motion vector is $infor(V)$, the sum of the information amounts of the components in the x and y directions are defined by Eq. (13) as the following:

$$infor(\vec{V}) \Rightarrow infor(\frac{\vec{V}}{\sqrt{2}}) \times 2 \tag{13}$$

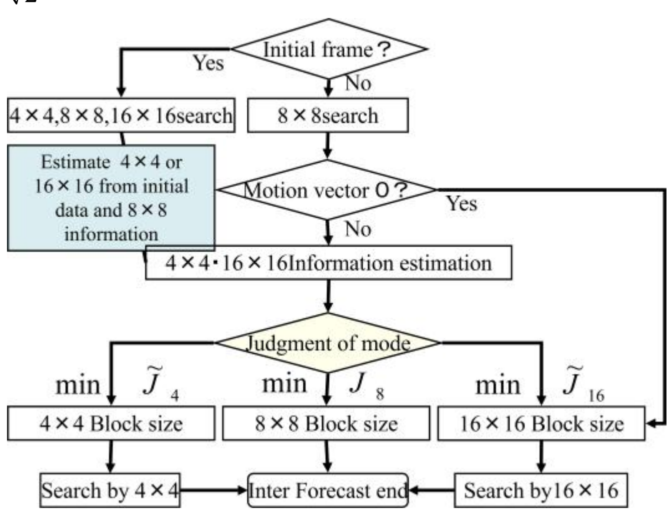


Figure 15: Flowchart of the proposed method

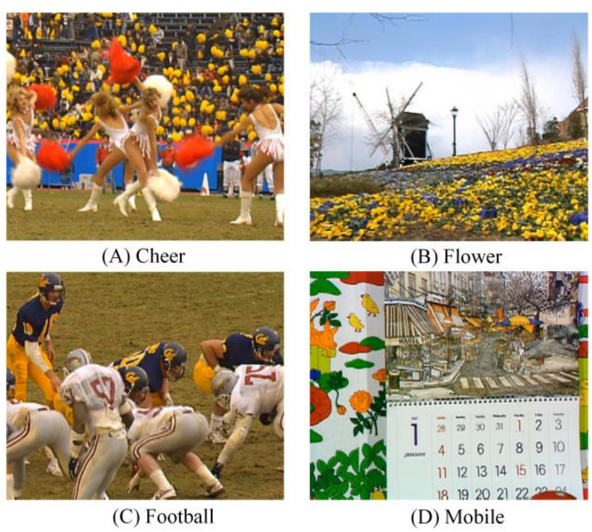
Fig. 15 illustrates the flowchart of the proposed method. It should be indicated that the flow of the proposed method is illustrated in the flow of three steps, named the initial frame, the second and subsequent frames, as well as the prediction mode decision. However, the flowchart loops macroblock units of 16×16 pixels and performs one frame. Moreover, in the present study, the mode decision is tested in only three modes (4×4, 8×8, 16×16).

Tab. 1 shows a comparison of the computational complexity between the proposed method and the three modes. Comparing the computational complexity of the proposed method, it is 13=0.33 [times] when the prediction mode is ascertained only for the 8×8 mode, thereby requiring the minimal amount of computation in one frame. 23=0.66 [times] when the prediction mode of only 4×4 or 16×16 mode with the maximum amount of calculation is ascertained. If the prediction mode is adaptively ascertained with 20% for 4×4, 20% for 8×8 and 60% for 16×16, 23×0.2+13×0.2+23×0.6=0.60 [times].

Table 1: Calculation amount of motion compensation

Model	Computation amount compared to 3 mode (times)
8×8 only	0.33
4×4 or 16×16 only	0.66
4×4(20%) 16×16(60%) 8×8(20%)	0.60

The four types of moving images employed in the experiment include the international standard moving images to assess the moving image compression coding defined by the international telecommunication union (ITU). The international standard video acts as one of the international standard formats of the video called common intermediate format (CIF) defined by ITU. CIF is a moving image of 288×352 pixels and 30 frames per sec.



(A) Cheer (B) Flower

(C) Football (D) Mobile

Figure 16: Four types of standard moving images

The mentioned four types of standard moving images are given in Figs. 16(A) to (D). Two of the four types of standard moving images refer to moving images connected in one scene change. The amount of change in information has a tight correlation with frames. Moreover, Cheer in Fig. 16(A) and Mobile in Fig. 16(D) have sophisticated moving areas, for example in the background. The experiment is performed on the frame shown in Fig. 16(B) and football is shown in Fig. 16(C), where the moving area is relatively simple and the background area is nearly stationary.

Table 2: Mode decision rate (2nd frame)

Model		Mode decision rate in each block size [%] (number of mode decisions)		
		4×4	8×8	16×16
Cheer	7 model	58%(223)	1%(3)	3%(12)
	3 model	77%(305)	18%(70)	5%(21)
	Proposed	56%(223)	2%(9)	42%(164)
Flower	7 model	5%(223)	5%(20)	39%(155)
	3 model	7%(223)	28%(122)	65%(255)
	Proposed	50%(223)	30%(122)	20%(78)
Football	7 model	26%(223)	2%(8)	9%(34)
	3 model	42%(223)	43%(173)	15%(56)
	Proposed	47%(223)	1%(4)	52%(205)
Mobile	7 model	14%(223)	5%(20)	23%(91)
	3 model	26%(223)	25%(99)	49%(196)
	Proposed	48%(223)	10%(39)	32%(168)

Since the image size of one frame is 288×352 pixels, there are 18×22=396 blocks per macroblock. Therefore, the respective mode is determined by performing the motion compensation for 396 times. In order to harvest and delve into the generation probability and estimation data for the initial frame, Tab. 2 lists the 396 mode decision rates in the second frame as 7 modes, 3 modes, as well as the proposed method. A range of mode decision results is achieved for the respective sequence. There are sequences with different results compared with the three modes. For instance, in terms of the mode decision rate for football, 8×8 is 43% in 3 modes, while the proposed method takes up only 1%, and the mode decision is shifted to 16×16. Furthermore, figures and tables should be inserted in the text of the manuscript.

5 Conclusion and discussion

In the present study, a simplified method of the computation is proposed to determine the prediction mode for the inter prediction in the international standard compression coding method H.264/AVC for moving pictures. In the proposed method, motion compensation is initially achieved in three modes of 4×4, 8×8, and 16×16 in the initial frame, and 8×8 and 4×4, 16×16 modes, which are intermediate between the three modes. The correlations between each amount of information are obtained. Moreover, the basic data for the prediction error estimation and motion vector estimation are calculated. Then, after the initial frame, the motion compensation is conducted with only 8×8, 16×16 and 4×4 prediction error values. Moreover, the motion vector information is estimated from the basic data for prediction error estimation and motion vector estimation. Finally, a mode decision method is developed by comparing 8×8 information with 4×4 and 16×16 estimated information.

It is concluded that the mode decision rate is characteristic for the respective sequence compared with only three modes. Although a decision rate is achieved similar to 3 modes and different results, the calculation time is reduced by 30% in the total time of the initial frame and the next frame. The total time from the initial frame to the 30th frame decreases by 60%. Moreover, the amount of information is up-regulated by only 1% in the second frame alone, and the image quality degradation is only about 0.1 [dB] in PSNR.

Funding Statement: This work was supported by QingLan Project of Jiangsu Province and National Science Fund of China (Nos. 61806088, 61902160), and was supported by Changzhou Science and Technology Support Plan (No. CE20185044).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Chen, X.; Zhong, H.; Bao, Z. (2019): A GLCM-feature-based approach for reversible image transformation. *Computers, Materials & Continua*, vol. 59, no.1, pp. 239-255.

Choe, G.; Nam, C.; Chu, C. (2018): An effective temporal error concealment in H. 264 video sequences based on scene change detection-PCA model. *Multimedia Tools and Applications*, vol. 77, no. 24, pp. 31953-31967.

Conti, C.; Soares, L. D.; Nunes, P. (2018): Light field coding with field-of-view scalability and exemplar-based interlayer prediction. *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 2905-2920.

Dalal, M.; Juneja, M. (2018): H. 264/AVC video steganography techniques: an overview. *International Journal of Computer Science and Engineering*, vol. 6, no. 5, pp. 297-303.

Fan, H.; Zhu, H. (2018): Separation of vehicle detection area using Fourier descriptor under Internet of things monitoring. *IEEE Access*, vol. 6, pp. 47600-47609.

Fan, H.; Zhu, H. (2018): Preservation of image edge feature based on snowfall model smoothing filter. *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 67.

He, J.; Wu, D.; Zeng, Y.; Hei, X.; Wen, Y. (2013): Toward optimal deployment of cloud-assisted video distribution services. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 10, pp. 1717-1728.

Joshi, Amit M.; Vivekanand, M.; Rajendra M. P. (2016): FPGA prototyping of video watermarking for ownership verification based on H. 264/AVC. *Multimedia Tools and*

Applications, vol. 75, no. 6, pp. 3121-3144.

Joshi, V.; Jain, S. (2020): Tampering detection and localization in digital video using temporal difference between adjacent frames of actual and reconstructed video clip. *International Journal of Information Technology*, vol. 12, no. 1, pp. 273-282.

Major, A. (2016): Powering premium content: an analysis of ooyala's online video distribution services. *The Projector*, vol.16, no. 1, pp. 4-22.

Molina, R.; Gazzano, J. D.; Rincon, F.; Gil-Costa, V.; Barba, J. et al. (2018): Heterogeneous soc-based acceleration of MPEG-7 compliance image retrieval process. *Journal of Real-Time Image Processing*, vol. 15, no. 1, pp. 161-172.

Nguyen, V. T.; Tropchenko, A. A. (2018): Improving efficiency of entropy coding method in modern video compression standards. *Trudy SPIIRAN*, vol. 58, pp. 111-127.

Pan, Z.; Jin, P.; Lei, J.; Zhang, Y.; Sun, X. et al. (2016): Fast reference frame selection based on content similarity for low complexity HEVC encoder. *Journal of Visual Communication and Image Representation*, vol. 40, pp. 516-524.

Shang, X.; Zhao, H.; Wang, G.; Zhao, X.; Zuo, Y. (2019): A novel objective quality assessment method for transcoded videos from H.264/AVC to H.265/HEVC utilizing probability theory. *IEEE Transactions on Broadcasting*, vol. 65, no. 4, pp. 777-781.

Xu, D.; Wang R. (2016): Two-dimensional reversible data hiding-based approach for intra-frame error c.oncealment in H. 264/AVC. *Signal Processing: Image Communication*, vol. 47, pp. 369-379.

Yao, Y.; Zhang, W.; Yu, N. (2016): Inter-frame distortion drift analysis for reversible data hiding in encrypted H. 264/AVC video bitstreams. *Signal Processing*, vol. 128, pp. 531-545.

Yuan, S.; Hu, J. (2019): Research on image compression technology based on huffman coding. *Journal of Visual Communication and Image Representation*, vol. 59, pp. 33-38.

Zhu, H.; Fan, H.; Shu, Z.; Yu, Q.; Zhao, X. et al. (2019): Edge detection with chroma components of video frame based on local autocorrelation. *IEEE Access*, vol. 7, pp. 48543-48550.