PAPER

Fast H.264/AVC DIRECT Mode Decision Based on Mode Selection and Predicted Rate-Distortion Cost

Xiaocong JIN[†a)], Jun SUN[†], Yiqing HUANG[††], Jia SU[††], *Nonmembers*, *and* Takeshi IKENAGA[††], *Member*

SUMMARY Different encoding modes for variable block size are available in the H.264/AVC standard in order to offer better coding quality. However, this also introduces huge computation time due to the exhaustive check for all modes. In this paper, a fast spatial DIRECT mode decision method for profiles supporting B frame encoding (main profile, high profile, etc.) in H.264/AVC is proposed. Statistical analysis on multiple video sequences is carried out, and the strong relationship of mode selection and rate-distortion (RD) cost between the current DIRECT macroblock (MB) and the co-located MBs is observed. With the check of mode condition, predicted RD cost threshold and dynamic parameter update model, the complex mode decision process can be terminated at an early stage even for small QP cases. Simulation results demonstrate the proposed method can achieve much better performance than the original exhaustive rate-distortion optimization (RDO) based mode decision algorithm by reducing up to 56.8% of encoding time for IBPBP picture group and up to 67.8% of encoding time for IBBPBBP picture group while incurring only negligible bit increment and quality degradation.

key words: *spatial DIRECT, adaptive, fast mode decision, H.264/AVC*

1. Introduction

The latest H.264/AVC standard [1] outperforms the previous video coding standard [2], [3] in various aspects such as bit rates, video quality, and coding time due to the introduction of new concepts and tools [4]: variable block size (VBS), context adaptive binary arithmetic coding (CABAC), etc. The appearance of small-block encoding ensures that the coding quality is kept at a comparably high level, but on the other hand, this also results in huge computation time in finding the best mode for the encoded macroblock (MB).

In profiles supporting B frame encoding, the DIRECT mode is defined, which can support bi-directional prediction and also transmit the residual information based on the prediction [5]. The DIRECT motion vector (MV) can be generated using temporal or spatial prediction and be classified as temporal DIRECT or spatial DIRECT, respectively [6]. Since this process is based on previously encoded information, any additional motion data will not be encoded. According to block size and availability of residual information, DIRECT mode can be divided into three categories: DIRECT_16x16, SKIP_16x16, and DIRECT_8x8 [1], [6]. This categorization strategy is illustrated in Fig. 1 to give the reader an intuitive understanding. The target of our research is also outlined in the figure: MB level spatial DIRECT (DIRECT_16x16 and SKIP_16x16 using spatial prediction). Besides DIRECT mode, inter modes and intra modes of different block scale are supported [7], [8]. Mode decision is the process to find the best encoding mode for the encoded MB. In this process, the encoder has to loop all the possible modes and carry out motion estimation (ME). This is considered most time-consuming because block matching in the prescribed area is complex [9] and especially so for B frame which supports multiple direction prediction [10]. If DIRECT mode can be determined at an early stage, huge computation time can be saved.

Some research work has been carried out to reduce the complexity. In [11], [12], a set of SKIP conditions is proposed and only the satisfied MB can be determined at an early stage as SKIP MB. This method is supposed to be efficient and accurate enough. Thus, it has already been incorporated into the H.264/AVC standard. But firstly, since the focus is only on SKIP mode, when encoding B frames, the speedup effect is limited by the SKIP_16x16 MB proportion. In small quantization parameter (QP) cases, the proportion of SKIP_16x16 MBs in B frame is generally lower compared to that of DIRECT_16x16 MBs. Secondly, for general QP cases, when encoding different video sequences, the criterion: the transform coefficient are all quantized to zero is too restrictive. In fact, many MBs with different values of coded block pattern (CBP) [13] will also finally fall into the DIRECT mode category. Thus, the method in [11], [12] is considered not effective enough when encoding B frame. Some potential to further speed up the B frame encoding process can be exploited. Reference [14] utilizes the CBP information to decide the SKIP mode early. This method also only targets SKIP mode and it can be seen from the test results that the hit ratio for SKIP mode with an average of 55% is still not high enough. Reference [15] utilizes PSNR prediction generation to determine mean square error (MSE) threshold for fast SKIP and DIRECT mode decision. This method is effective for all kinds of sequences since the threshold set up is based on encoded information. However, the major drawbacks can be clearly identified: the computational load for deciding the threshold is intensive; also, judging from the result, the correctness rate is not high enough. Furthermore, this algorithm targets only temporal DIRECT. Generally, temporal DIRECT provides better performance for sequences with relatively small single direction motion while spatial DIRECT can be utilized to deal with general

Manuscript received November 15, 2010.
Manuscript revised March 23, 2011.
[†]The authors are with Institute of Image Communication and Information Processing, Shanghai Jiaotong University, Shanghai, 200240 China.
[††]The authors are with IPS, Waseda University, Kitakyushu-shi, 808–0135 Japan.
a) E-mail: jxcking@sjtu.edu.cn
DOI: 10.1587/transinf.E94.D.1653

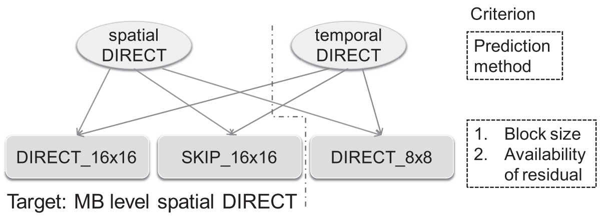


Fig. 1 Classification of DIRECT mode based on different criterion.

motion sequences.

In this paper, we propose an adaptive fast spatial DIRECT mode decision method focusing on DIRECT_16x16 and SKIP_16x16 (MB level DIRECT mode) in order to speed up the encoding process. It is based on the information of mode selection and rate-distortion (RD) cost of neighboring co-located MBs. Also, a novel parameter update model is proposed. Compared to the existing fast algorithm, our method enables large speedup for DIRECT_16x16 MBs and provides comparative encoding time saving for SKIP_16x16 MBs.

The rest of the paper is organized as follows. In Sect. 2, The proposed adaptive spatial DIRECT mode decision scheme is explained in detail. Section 3 summarizes the proposed fast spatial DIRECT mode decision method and explains the overall flowchart. Experimental results are given in Sect. 4. The last section is the conclusion.

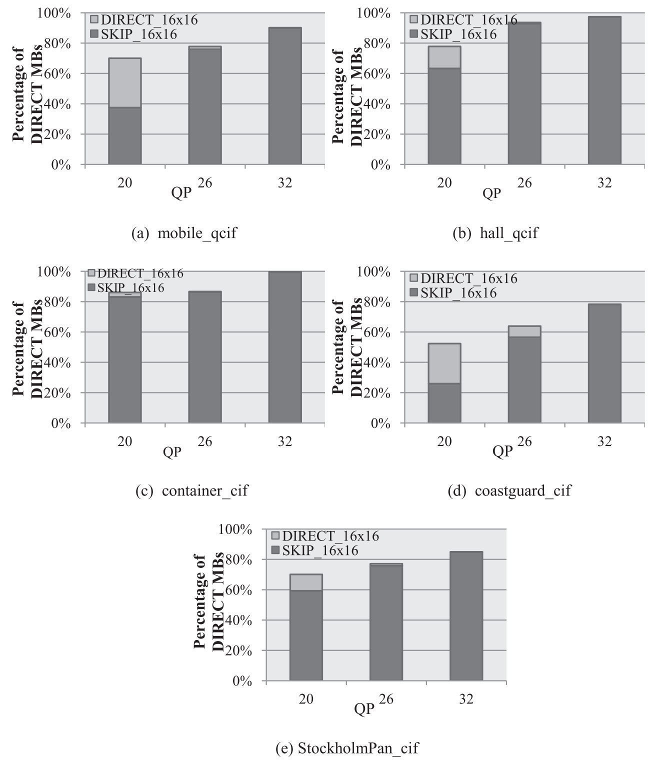
2. Adaptive Fast Spatial DIRECT Mode Decision Algorithm

2.1 Potentiality in Spatial DIRECT Mode Decision

As introduced in the previous part, if a DIRECT MB has no residual information, it can be referred as SKIP_16x16, otherwise it is DIRECT_16x16 [6]. Different from SKIP mode defined in P frame, spatial DIRECT mode in B frame is found to be a common existence among all kinds of sequences regardless of the QP values. Figure 2 shows the ratio of DIRECT MBs (encoded either in DIRECT_16x16 or SKIP_16x16) to all the total encoded B frame MBs using H.264/AVC reference encoder JM 11.0 [16]. The black part corresponds to SKIP_16x16 and the grey portion corresponds to DIRECT_16x16. For "container_cif", the proportion is above 80%. Even for "coastguard_cif", the rate is above 50% at the least, in which DIRECT_16x16 occupies almost the half of the total DIRECT MBs. The data in Fig. 2 indicate that for general sequences, a large proportion of MBs are encoded in spatial DIRECT mode. Thus, we can assume there is a great potential to speed up the encoding process. In the following part, we try to find out some rules about DIRECT mode from two aspects of observations: mode selection relation and RD cost relation.

2.2 Mode Selection Relation

Since redundancy always exists in the original video se-



(a) mobile_qcif

(b) hall_qcif

(c) container_cif

(d) coastguard_cif

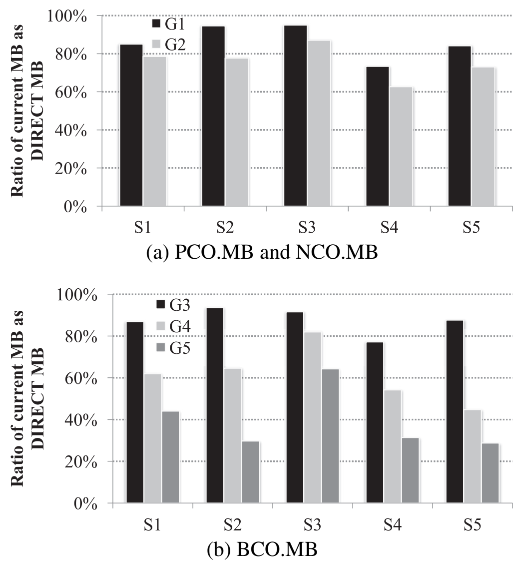
(e) StockholmPan_cif

Fig. 2 Percentage of DIRECT MBs.

quences, it is feasible to utilize either spatial or temporal information to help achieving video compression [17]–[19]. In our paper, mode selection result and RD cost of the co-located MB in the previous or next frame are used as temporal information for the current encoding MB. The relationship of mode selection between B frame and P frame is firstly analyzed in the following part.

Since there is difference between B frame and P frame, we would like to check the relation of mode selection of co-located MB in the previous B frame (BCO.MB), co-located MB in the previous P frame (PCO.MB), co-located MB in the next P frame (NCO.MB), and the current MB. Based on the observation that small-block inter prediction or intra prediction indicates complicated detail and thus such area is of less probability to be encoded in DIRECT mode, we classify the encoding modes of PCO.MB ($MD_{PCO.MB}$) and NCO.MB ($MD_{NCO.MB}$) as two groups: SKIP or INTER_16x16, other modes. The encoding modes for BCO.MB ($MD_{BCO.MB}$) are classified into three groups: DIRECT, INTER_16x16, other modes. In JM 11.0 [16], mode index 0 corresponds to DIRECT mode for B frame and SKIP mode for P frame, respectively; mode index 1 means INTER_16x16 mode. Figure 3 gives the average result after encoding 199 frames of various video sequences in group of picture (GOP) IBPBP with QP belonging to the set {20, 24, 28, 32}. It shows the possibility of deciding the current MB in B frame as DIRECT MB (either SKIP_16x16 or DIRECT_16x16) when both PCO.MB and NCO.MB, or BCO.MB is of the classified mode-set which is defined previously.

Figure 3 (a) shows that when both PCO.MB and



G1:when ($MD_{PCO.MB} \leq 1$) && ($MD_{NCO.MB} \leq 1$)
G2:when ($MD_{PCO.MB} > 1$) || ($MD_{NCO.MB} > 1$)
G3:when $MD_{BCO.MB} = 0$ G4:when $MD_{BCO.MB} = 1$
G5:when $MD_{BCO.MB} > 1$

S1:mobile_qcif S2:hall_qcif S3:container_cif
S4:coastguard_cif S5:StockholmPan_cif

Fig. 3 Analysis of mode selection relation.

NCO.MB are of SKIP or INTER_16x16, the possibility of the current MB's optimal mode to be DIRECT mode is much higher than other cases. Especially for the video sequences with slow motion such as container, the rate can be up to 95%. Besides, SKIP or INTER_16x16 is the dominant encoding mode for these sequences. Similar results can also be found in Fig. 3 (b): when BCO.MB is adopted as DIRECT mode, the current MB to be DIRECT MB is of the highest possibility. So we get a conclusion that high correlation exists between the DIRECT MB and the co-located MB(s) in both neighboring P and B frames. Based on the analysis of mode selection, the following criterion is proposed in our fast mode decision algorithm:

$$(MD_{PCO.MB} \leq 1 \&\& MD_{NCO.MB} \leq 1) || \\ (MD_{BCO.MB} = 0). \tag{1}$$

It means that the current MB can be judged as DIRECT MB early if Eq. (1) is satisfied.

To verify the effectiveness of the mode constraint in Eq. (1), the coverage percentage (CP) is defined as Eq. (2) and the analysis result is shown in Fig. 4.

$$CP = \frac{\#\ DIRECT\ MB_{satisfying\ Eq.\,(1)}}{\#\ TOTAL\ DIRECT\ MB} \tag{2}$$

For sequences consisting of general speed movement, CP is able to be kept above 80%. This proves the mode constraint can be applied to most of the DIRECT MBs.

On the other hand, simply applying the mode constraint is not enough, since not all of the MBs satisfying Eq. (1) will

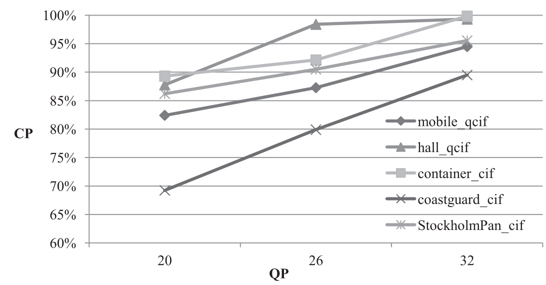


Fig. 4 Coverage percentage of mode constraint Eq. (1).

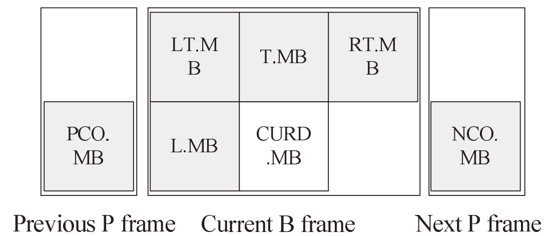


Fig. 5 Candidate MBs.

be DIRECT MBs. Thus, RD cost constraint in the following section is also treated as one factor of the criteria.

2.3 Rate-Distortion Cost Relation

In the previous section, it is demonstrated that complexity reduction can be achieved based on mode selection information. However, such adoption is still not enough for accurate early DIRECT mode decision. To balance the quality and complexity, another criterion called RD cost information is further introduced. In high complexity mode of H.264/AVC, the encoder computes RD cost based on an exhaustive Lagrangian optimization framework [20]. During the encoding process, for each mode, RD cost is calculated by

$$J = SSD(s, c, MODE|QP) \\ + \lambda_{MODE} \times R(s, c, MODE|QP), \tag{3}$$

where QP is the quantization parameter, $MODE$ corresponds to the encoding mode of the MB and SSD is the sum of the squared difference between the original video signals (s) and the reconstructed video signals (c). λ_{MODE} is the Lagrangian multiplier. R represents rate after quantization.

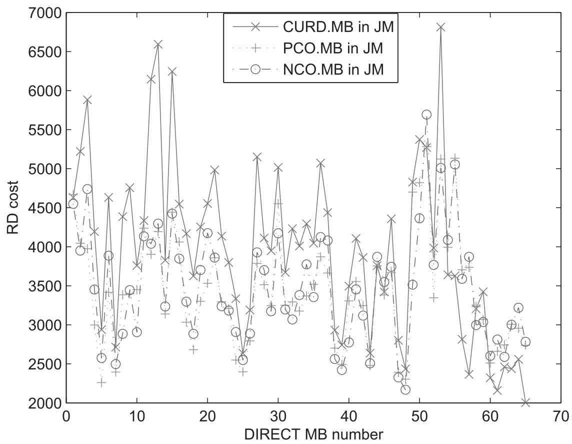
For evaluating the relation of RD cost, six MBs are chosen as candidate MBs and shown in Fig. 5: left MB (L.MB), left-top MB (LT.MB), top MB (T.MB), right-top MB (RT.MB) in the current B frame, PCO.MB and NCO.MB. To evaluate the RD cost relation between the current DIRECT MB (CURD.MB) and the six candidate MBs, correlation coefficient is utilized [21]:

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}, \tag{4}$$

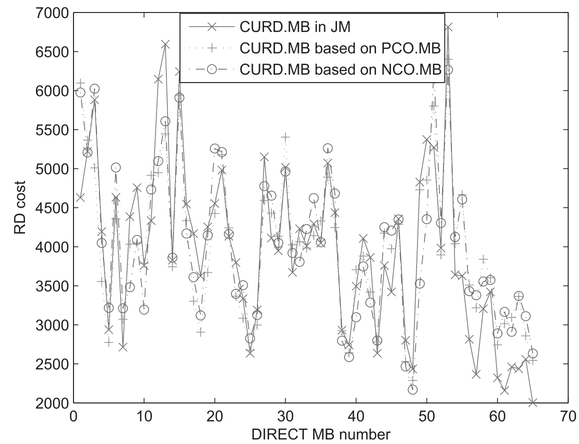
in which X is replaced with the RD cost of CURD.MB under all QP settings and Y is replaced with the RD cost of one candidate MB under all QP settings, correspondingly. Since

Table 1 Correlation coefficient of RD cost between CURD.MB and candidate MBs.

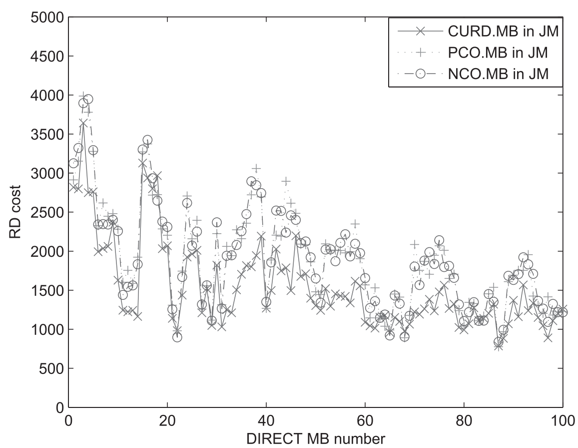
	L.MB	LT.MB	T.MB	RT.MB	PCO.MB	NCO.MB
mobile_qcif	0.875	0.792	0.828	0.833	0.956	0.960
hall_qcif	0.641	0.511	0.671	0.563	0.964	0.968
container_cif	0.889	0.703	0.728	0.688	0.982	0.984
coastguard_cif	0.868	0.728	0.753	0.705	0.925	0.929
StockholmPan_cif	0.888	0.816	0.864	0.804	0.977	0.980



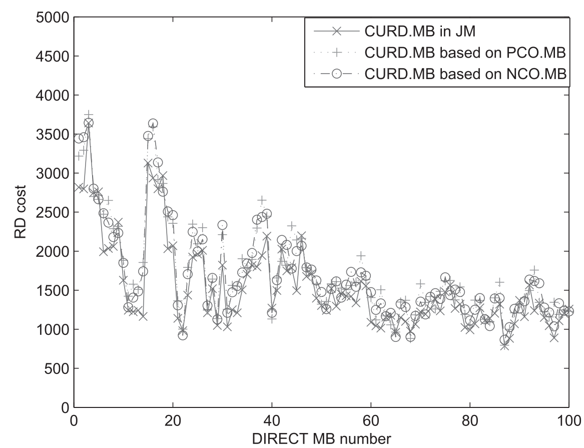
(a) mobile_qcif, 179th frame



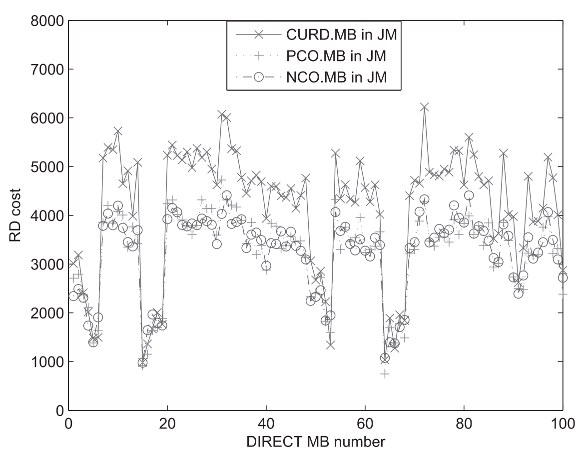
(b) mobile_qcif, 179th frame



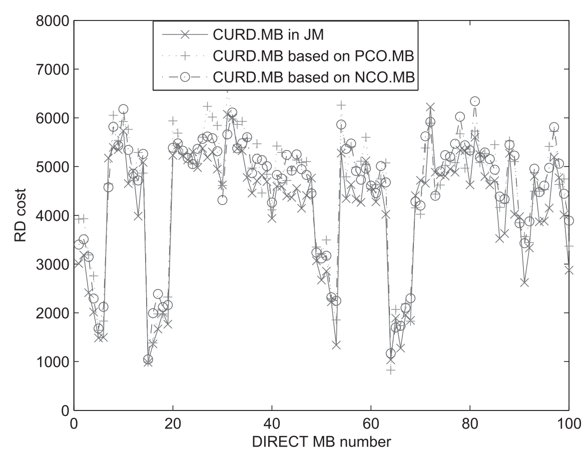
(c) container_cif, 179th frame



(d) container_cif, 179th frame



(e) StockholmPan_720p, 179th frame



(f) StockholmPan_720p, 179th frame

Fig. 6 RD cost comparison curve. (a)(c)(e) Situation of CURD.MB, PCO.MB and NCO.MB in JM. (b)(d)(f) Situation of CURD.MB in JM, CURD.MB based on PCO.MB and CURD.MB based on NCO.MB.

through this analysis, we try to find out a common relation among all QPs, the data are summarized based on different sequences rather than further separating them on different QP cases. *cov* is the covariance and σ means standard deviation. Table 1 gives the analysis result after encoding 199 frames in GOP IBPBP with QP belonging to the set {20, 24, 28, 32}.

By analyzing data in Table 1, it is shown that for all the sequences, the relation coefficients of PCO.MB and NCO.MB are almost the same and both of them are much larger than that of other candidates. Since they are close to 1, which means linear dependency between RD cost of CURD.MB and PCO.MB or NCO.MB is strong, approximately linear relationship is deduced:

$$\begin{cases} J_{CURD.MB} = \alpha \times J_{PCO.MB} \\ J_{CURD.MB} = \beta \times J_{NCO.MB} \end{cases} \tag{5}$$

This relationship is further utilized in our proposal: predict the RD cost of the current B frame MB in DIRECT mode with the existing RD cost of PCO.MB and NCO.MB.

Figure 6 (a)(c)(e) show the RD cost comparison of the CURD.MB, PCO.MB and NCO.MB (all generated from original JM [16]) based on encoding the sequences "mobile_qcif" "container_cif" "StockholmPan_720p" for 199 frames in GOP unit IBPBP at QP 20. The data were selected at the 179th frame (B frame). Altogether, there are 65, 347 and 2072 DIRECT MBs in this frame of the three sequences, respectively. For detail analysis, we pick up 65, 100, 100 DIRECT MBs for qcif, cif and 720p sequences as shown in Fig. 6 (a)(c)(e). By analyzing the curves, we can see that there is great difference between the RD cost of CURD.MB and PCO.MB or NCO.MB, but trends of the three sets of value are almost the same which proves the correctness of linear relation assumption in Eq. (5). Another result we can get from Fig. 6 (a)(c)(e) is that, the α and β vary among different sequences. For example, α : RD cost ratio of CURD.MB as to PCO.MB is larger than 1 in most cases for the selected frame of "mobile" sequence, but for "container" sequence, α should be smaller than 1. So it is necessary to utilize an adaptive scheme to improve the performance.

From our observation, the multipliers α and β in Eq. (5) are not only related with QP, but also affected by image features such as degree of motion and image detail. Therefore, a dynamic parameter update model is proposed based on information of PCO.MB and NCO.MB, as shown in Fig. 7.

The pseudocode shown involves two steps when the current MB's optimal mode is DIRECT: calculation of current RD cost rate and rate parameter update. Here, θ and λ are RD cost ratios to be updated when the current MB in B frame is decided as DIRECT MB. i in Fig. 7 symbolizes the MB location in the current frame. $n[i]$ is the variable to record how many DIRECT MBs exist in position i for the total encoded frames. The calculations of α and β are actually based on average values. Once the current MB is decided as DIRECT MB, the multipliers α and β in Eq. (5) should be averaged again according to Fig. 7 for the new

if (current MB is DIRECT)

{

//Get the current RD cost rate

$\theta = \dfrac{J_{\text{CURD.MB}}}{J_{\text{PCO.MB}}}, \lambda = \dfrac{J_{\text{CURD.MB}}}{J_{\text{NCO.MB}}};$

//Update rate parameter

$\alpha[i] = \dfrac{\theta + \alpha[i] \times n[i]}{n[i]+1} \times 1.1,$

$\beta[i] = \dfrac{\lambda + \beta[i] \times n[i]}{n[i]+1} \times 1.1;\}$

Fig. 7 Pseudocode of proposed dynamic parameter update model.

entries of θ and λ; $n[i]$ should also be incremented by 1. So, the proposed model fully considered image features for accurate decision process. The constant 1.1 in Fig. 7 was determined experimentally with concern to the trade-off of encoding complexity and quality.

Figure 6 (b)(d)(f) show three curves which are RD cost of CURD.MB (generated from original JM [16]), predicted value based on PCO.MB, and NCO.MB using Eq. (5) and equations in Fig. 7. In most cases, this adaptive method can achieve a good prediction of the RD cost of CURD.MB.

All in all, from the description above, mode constraint and RD cost constraint can be used to make early judgment of DIRECT mode. In the proposed overall algorithm, we fully utilize these two kinds of information for fast accurate DIRECT mode decision.

3. Overall Scheme of The Proposed Algorithm

Based on the observation in Sect. 2, the proposed early DIRECT mode decision method for current MB is summarized in Eq. (6)(7). It means that at an early stage, the co-located MBs' mode selection result and the RD cost of current MB in DIRECT mode will be checked. The current MB can be pre-determined as DIRECT MB when both the criterion in Eq. (6)(7) are satisfied.

$$\begin{aligned} Mode: &(MD_{PCO.MB} \leq 1 \&\& MD_{NCO.MB} \leq 1) \| \\ &(MD_{BCO.MB} = 0) \end{aligned} \tag{6}$$

$$\begin{aligned} RDcost: &(J_{CURRENT\_MB} < \alpha \times J_{PCO.MB}) \| \\ &(J_{CURRENT\_MB} < \beta \times J_{NCO.MB}) \end{aligned} \tag{7}$$

The whole flowchart of the algorithm is given in Fig. 8. Specifically, in case of B frame encoding, instead of generating motion vector of DIRECT mode in a latter stage, we move this part to the first stage and calculate the RD cost of DIRECT mode. In the following step, constraints for early decision of DIRECT mode for the current MB will be checked according to Eq. (6)(7).

After mode decision of the current MB, if the best mode is DIRECT mode, the proposed dynamic parameter update model as discussed in Fig. 7 will be executed for preparation of next B frame encoding. Specifically: DIRECT MB counter $n[i]$ in Fig. 7 should be incremented by

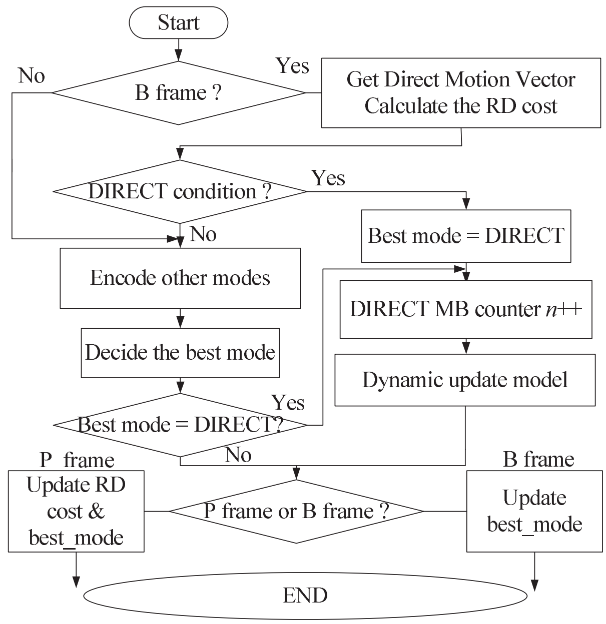


Fig. 8 Flowchart of proposed fast spatial DIRECT mode decision algorithm.

1, rate factors α and β should also be updated and *best_mode* will be stored.

As for P frame, the encoder provides a buffer to store the encoding mode as well as the RD cost for the MBs after rate-distortion optimization (RDO) based mode decision.

4. Experimental Results and Discussion

4.1 Test Conditions

In order to verify the performance of the proposed fast spatial DIRECT mode decision algorithm, we implemented our algorithm in JM 11.0 software [16] based on the high complexity mode (RDOptimization = 1). One alternative method is used for comparison: fast SKIP mode decision algorithm [11], [12] which was already incorporated into the JM software as fast high complexity mode (RDOptimization = 2). For this algorithm, we turn off the fast mode decision for P frame. So only the B frame speed up effect is taken into consideration. Analysis is performed with encoding frames = 199 for IBPBP picture group and 208 for IBBPBBP picture group; only spatial DIRECT mode is considered; QP belongs to the set {20, 24, 28, 32}; the number of reference frame is 2; MV search range is set as ±16 for QCIF, ±32 for CIF and ±64 for 720p sequences.

4.2 Performance Evaluation and Comparisons

The encoding quality is evaluated with ΔBits and ΔPSNR, which are the differences in bit rate and PSNR between the fast mode decision method and the referenced one (high complexity mode). The result with "–", for example in Table 2, indicates PSNR degradation or bit rate saving. The

Table 2 TS (%), ΔPSNR (dB) and ΔBits (%) in GOP IBP.

Sequence	QP	method	TS	ΔPSNR	ΔBits
mobile (qcif)	20	[11][12]	7.1	−0.00	+0.25
		our	32.0	−0.02	+0.16
	24	[11][12]	14.5	−0.00	+0.37
		our	34.3	−0.02	+0.77
	28	[11][12]	28.2	−0.01	+0.06
		our	39.8	−0.06	−0.20
	32	[11][12]	39.8	−0.02	−0.56
		our	45.4	−0.03	−1.04
hall (qcif)	20	[11][12]	5.5	−0.00	−0.08
		our	42.0	−0.02	+0.08
	24	[11][12]	28.7	−0.00	−0.14
		our	45.9	−0.01	+0.51
	28	[11][12]	43.3	−0.00	−0.06
		our	47.8	−0.01	+0.90
	32	[11][12]	47.1	−0.00	−0.18
		our	49.5	−0.01	+0.55
container (cif)	20	[11][12]	40.0	−0.02	−1.07
		our	52.5	−0.02	−1.10
	24	[11][12]	48.3	−0.04	−2.05
		our	54.0	−0.04	−2.13
	28	[11][12]	54.1	−0.01	−1.15
		our	55.6	−0.01	−1.16
	32	[11][12]	56.5	−0.00	−0.18
		our	56.8	−0.00	−0.22
coastguard (cif)	20	[11][12]	4.9	−0.00	+0.20
		our	30.9	−0.02	+0.51
	24	[11][12]	13.5	−0.01	+0.13
		our	32.3	−0.03	+0.32
	28	[11][12]	25.0	−0.02	+0.35
		our	34.6	−0.03	+0.28
	32	[11][12]	35.5	−0.05	+0.59
		our	38.5	−0.04	−0.17
stefan (cif)	20	[11][12]	13.4	−0.00	+0.15
		our	25.1	−0.01	+1.37
	24	[11][12]	17.5	−0.00	+0.24
		our	26.4	−0.03	+0.44
	28	[11][12]	24.3	−0.02	+0.38
		our	28.5	−0.06	+0.85
	32	[11][12]	31.6	−0.04	+0.62
		our	33.2	−0.06	+1.10
harbour (720p)	20	[11][12]	1.4	−0.00	+0.32
		our	32.1	−0.02	+0.64
	24	[11][12]	5.9	−0.00	+0.39
		our	28.7	−0.02	+0.56
	28	[11][12]	18.1	−0.02	+0.30
		our	28.4	−0.05	−0.56
	32	[11][12]	34.7	−0.07	+0.09
		our	35.0	−0.05	−0.71
mobcal (720p)	20	[11][12]	3.9	−0.00	+0.16
		our	44.0	−0.02	+0.39
	24	[11][12]	12.1	−0.00	+0.38
		our	42.3	−0.02	−0.17
	28	[11][12]	33.7	−0.02	+0.63
		our	46.2	−0.03	−0.88
	32	[11][12]	45.5	−0.05	+0.56
		our	49.3	−0.03	−0.85
StockholmPan (720p)	20	[11][12]	1.6	−0.00	+0.21
		our	41.6	−0.06	−0.22
	24	[11][12]	14.4	−0.00	+0.18
		our	45.6	−0.02	−0.75
	28	[11][12]	40.1	−0.01	+0.54
		our	49.5	−0.01	+0.25
	32	[11][12]	47.9	−0.04	+1.28
		our	50.9	−0.01	−0.16

meaning of "+" can be deduced by analogy. The reduction of computation complexity is measured as:

$$TS = \frac{TIME_{reference} - TIME_{proposal}}{TIME_{reference}} \times 100\% \qquad (8)$$

where TS is the total encoding time saving as a percentage between the fast mode decision method and the reference: high complexity mode.

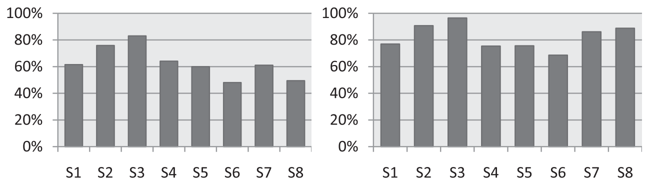
Firstly, the algorithm performance for the IBPBP picture group is tested. The encoding results and comparisons are shown in Table 2. "mobile_qcif" is a sequence containing slow zooming and a complex horizontal, vertical movement. It can be observed that when QP is 20, the proposed fast spatial DIRECT mode decision scheme can achieve 32% computation time saving. The PSNR degradation is only 0.02 dB. When QP increases, the proposed algorithm can save more bits than the compared fast SKIP mode decision method [11], [12] while keeping the same level of time reduction. Similar result can be found with "coastguard_cif" which contains fast horizontal movement, camera panning and complex spatial detail. As for "container_cif", even when QP is 20, more than 50% encoding time is reduced based on our algorithm. Performance gets even better with increase of QP. We can also find that the bit rate for all the QP cases is slightly reduced which proves the prediction correctness is at a high level. "stefan_cif" is a fast motion sequence containing fast horizontal camera movement and complicated background information. In this case, comparably larger bit rate increment can be observed. This implies that our proposal still has difficulties in precisely encoding fast motion sequences. In time saving aspect, our method still outperforms the referenced method in all QP settings. Since the proposed method is based on the temporal information and the relation has been proved to exist in both qcif and cif sequences, we further expect DIRECT MBs in high resolution sequences also satisfy the proposed DIRECT condition. Three 720p high resolution sequences are also tested. It is shown that for QP larger than 24, our algorithm can achieve larger computation time saving as well as better encoding quality. So, the proposed algorithm is suitable for encoding high resolution sequences.

Secondly, Table 3 shows the encoding result for IBBPBBP picture group. Averagely, the time saving for video sequences at different QPs is larger than the result using IBPBP picture group coding due to the fact that, in IBBPBBP picture group, the proportion of B frame in the total encoding frames has increased to 2/3. Also, since the number of B frames is increased, the distance between B frame and the reference frame (P frame) also increases, which results in larger prediction error when utilizing fast mode decision scheme. By comparing the results shown in Table 2 and Table 3, it can be demonstrated that the proposed method works well with both IBPBP and IBBPBBP picture groups. Averagely, compared with [11], [12], about 20.6% improvement in time saving is achieved among qcif, cif and 720p sequences.

Thirdly, it should also be mentioned that although our

Table 3 TS (%), ΔPSNR (dB) and ΔBits (%) in GOP IBBP.

Sequence	QP	method	TS	ΔPSNR	ΔBits
mobile (qcif)	20	[11][12]	2.9	−0.00	+0.31
		our	29.1	−0.02	+1.31
	24	[11][12]	7.7	−0.01	+0.42
		our	31.8	−0.05	+1.15
	28	[11][12]	22.0	−0.03	+0.27
		our	39.5	−0.10	+0.81
	32	[11][12]	33.9	−0.07	+0.68
		our	45.5	−0.12	−0.00
hall (qcif)	20	[11][12]	6.5	−0.01	−0.34
		our	50.2	−0.04	−0.18
	24	[11][12]	34.6	−0.01	−0.22
		our	55.8	−0.02	+1.80
	28	[11][12]	53.3	−0.01	−0.14
		our	57.7	−0.04	+1.08
	32	[11][12]	59.4	−0.01	−0.20
		our	61.5	−0.04	+0.24
container (cif)	20	[11][12]	44.0	−0.02	+0.41
		our	57.1	−0.05	−0.03
	24	[11][12]	55.9	−0.05	−0.23
		our	60.9	−0.07	−0.81
	28	[11][12]	63.5	−0.04	−0.46
		our	65.2	−0.05	−1.63
	32	[11][12]	67.6	−0.04	−0.72
		our	67.8	−0.04	−1.55
coastguard (cif)	20	[11][12]	4.1	−0.00	+0.28
		our	34.9	−0.04	+0.95
	24	[11][12]	11.8	−0.01	+0.32
		our	36.0	−0.05	+0.49
	28	[11][12]	24.5	−0.02	+0.44
		our	39.1	−0.05	+0.06
	32	[11][12]	38.5	−0.05	+1.28
		our	46.5	−0.05	+0.20
stefan (cif)	20	[11][12]	14.5	−0.01	+0.16
		our	25.5	−0.03	+1.53
	24	[11][12]	19.0	−0.01	+0.31
		our	27.9	−0.03	+2.17
	28	[11][12]	25.9	−0.02	+0.53
		our	30.5	−0.08	+1.37
	32	[11][12]	34.9	−0.05	+0.73
		our	37.4	−0.09	+1.62
harbour (720p)	20	[11][12]	1.1	−0.00	+0.37
		our	34.7	−0.03	+1.01
	24	[11][12]	4.6	−0.00	+0.54
		our	30.6	−0.02	+1.24
	28	[11][12]	16.9	−0.02	+0.71
		our	29.9	−0.07	−0.52
	32	[11][12]	37.3	−0.08	+0.56
		our	39.0	−0.08	−0.46
mobcal (720p)	20	[11][12]	4.4	−0.00	+0.30
		our	52.8	−0.02	+1.41
	24	[11][12]	13.9	−0.00	+0.91
		our	49.8	−0.04	+1.04
	28	[11][12]	39.2	−0.04	+1.53
		our	54.7	−0.07	−1.14
	32	[11][12]	54.6	−0.09	+0.77
		our	57.8	−0.07	−1.91
StockholmPan (720p)	20	[11][12]	1.6	−0.00	+0.31
		our	50.1	−0.08	+0.02
	24	[11][12]	17.1	−0.00	+0.43
		our	54.4	−0.04	−0.61
	28	[11][12]	47.2	−0.02	+0.77
		our	59.8	−0.02	+0.24
	32	[11][12]	59.0	−0.06	+0.96
		our	62.4	−0.03	+0.96

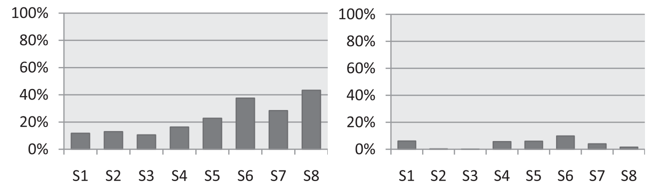


(a) QP=20 (b) QP=32

S1: mobile_qcif S2: hall_qcif S3: container_cif S4: coastguard_cif

S5: stefan_cif S6: harbour_720p S7: mobcal_720p S8: StockholmPan_720p

Fig. 9 Hit ratio using the proposed algorithm.



(a) QP=20 (b) QP=32

S1: mobile_qcif S2: hall_qcif S3: container_cif S4: coastguard_cif

S5: stefan_cif S6: harbour_720p S7: mobcal_720p S8: StockholmPan_720p

Fig. 10 Wrong prediction rate (WPR) using the proposed algorithm.

method can be combined with the method in [11], [12], limited improvement can be expected. Because the method in [11], [12] targets only SKIP mode while our method is proposed for MB level DIRECT mode, i.e., DIRECT_16x16 and SKIP_16x16. By analyzing the data in Table 2 and Table 3, it can be found that with QP increases, the complexity reduction of the proposed method in [11], [12] becomes greater while the performance of our method remains almost at the same level. In fact, based on Fig. 2, we found that most DIRECT_16x16 MBs in low QP case will turn to SKIP_16x16 MBs in large QP case and the number of MB level DIRECT mode, which is the sum of the aforementioned two cases, will not change too much when QP increases. Thus, even when QP is becoming larger, we can expect that the complexity reduction of [11], [12] and our method will become almost the same. So there is no need to combine these two methods in terms of B frame encoding.

Lastly, the hit ratio which is defined in Eq. (9) is given, as shown in Fig. 9. It demonstrates the efficiency of our early DIRECT MB detection scheme.

$$hit\_ratio = \frac{\#DIRECT\ MBs_{true\_positive}}{\#DIRECT\ MBs_{JM}} \times 100\% \qquad (9)$$

In the equation above, $\#DIRECT\ MBs_{JM}$ indicates the number of DIRECT MBs according to original JM; $\#DIRECT\ MBs_{true\_positive}$ means the number of correctly predicted DIRECT MBs using our proposal. According to this definition of hit ratio, we evaluate the percentage of correctly early decided DIRECT MBs, which will directly lead to computation time saving without any quality loss.

From Fig. 9, it can be observed that when QP is equal to 20, the hit ratio varies among different sequences and can be 49.3% at the least. The hit ratio will increase when QP increases. In the case when QP is 32, the hit ratio can be kept above 70%. In case of "container_cif", the ratio reaches 96.5%, which means the proposed algorithm almost achieves perfect early detection of all candidate DIRECT MBs. For other sequences, the ratio can be kept at a reasonably high level.

In our proposal, in accordance with the flowchart shown in Fig. 8, even if an MB whose best mode is DIRECT according to original JM, fails to satisfy the sets of criteria defined in Eq. (6)(7), it will still be decided as DIRECT MB in our proposal after the exhaustive loop of all possible modes. So the false negative case will not lead to quality loss either.

The only possible case that will result in quality loss by employing our scheme is the false positive case, which is when an MB whose best mode is not DIRECT according to original JM satisfies the criteria defined in Eq. (6)(7). As a result, these kinds of MBs are wrongly treated as DIRECT MBs in our proposal. Thus, we further introduce a metric to evaluate the wrong prediction rate (WPR):

$$WPR = \frac{\#DIRECT\ MBs_{false\_positive}}{\#DIRECT\ MBs_{positive}} \times 100\% \qquad (10)$$

In the equation above, $\#DIRECT\ MBs_{positive}$ is the number of MBs determined as DIRECT MBs at an early stage using our method, i.e., number of MBs satisfying Eq. (6)(7). $\#DIRECT\ MBs_{false\_positive}$ indicates the number of MBs that are wrongly judged as DIRECT MBs in our method. By evaluating WPR, we can know the percentage of wrongly predicted DIRECT MBs among the total number of early decided MBs using the proposed method, which is the only case that will bring about quality degradation.

Figure 10 illustrates the WPR value of different sequences under various QP settings. It can be observed from (a) that when QP is small, WPR can be kept below 20% for most small resolution sequences (qcif or cif). However, when resolution gets larger to 720p, WPR is comparably high, which means a large proportion of early DIRECT MBs' optimal mode should not be DIRECT according to original JM. This will surely bring down the encoding performance in terms of bit rate or PSNR. Nevertheless, by checking the performance of our proposal in large resolution sequences such as "StockholmPan_720p" when QP is 20 in Table 2, it can be found the performance degradation is rather limited. For example, the bit rate for this sequence under QP 20 is even reduced. The PSNR drop is also below 0.1 dB. In fact, although our method wrongly judges many DIRECT MBs in this case, the sub-optimal modes of these MBs are most likely to be DIRECT. Thus, even the optimal mode for these MBs are not chosen, it will not cause too much effect on encoding quality since sub-optimal mode can also result in low coding effort (RD cost is also low). This is also the corollary of our DIRECT criteria in Eq. (6)(7). Equation (6) requires the co-located MBs in neighboring frames are smooth blocks, so the current MB has a tendency to be smooth. With further regulation on RD cost using Eq. (7), even the MB is wrongly judged as DI-

Table 4 Small QP test based on TS (%) for GOP IBPBP.

sequence	method	TS	
		QP = 12	QP = 16
mobile_qcif	[11][12]	1.9	2.1
	our	33.2	32.0
hall_qcif	[11][12]	1.7	1.9
	our	44.8	39.3
container_cif	[11][12]	1.4	8.1
	our	47.1	49.3
coastguard_cif	[11][12]	1.3	1.3
	our	33.6	31.2
stefan_cif	[11][12]	5.7	6.6
	our	23.5	23.4
harbour_720p	[11][12]	0.0	0.1
	our	40.6	36.2
mobcal_720p	[11][12]	0.3	0.3
	our	50.2	47.7
StockholmPan_720p	[11][12]	0.3	0.3
	our	42.1	39.9

RECT MB, the additional cost is quite limited. When QP becomes larger, from Fig. 10 (b), we can see *WPR* is kept below 10% for every sequence. This is also conformable to the test result in Table 2.

4.3 Small QP Case Encoding

In the previous subsection, we showed the effectiveness of the proposed algorithm when encoding in general QP settings. Based on further experiments, it is shown that when QP is small, the proportion of SKIP_16x16 MBs is also becoming smaller, thus, our algorithm can be more effective than the existing work. Table 4 shows the time saving result comparison between our algorithm and [11], [12]. It can be seen that in small QP cases, our algorithm can still achieve huge time reduction without much influence of QP value change while the performance of [11], [12] deteriorates in these cases since the number of SKIP_16x16 greatly drops. In addition, only negligible quality loss is observed in the experiment using our scheme.

In conclusion, our algorithm can achieve complexity reduction for both small QP case and large QP situation. Also, this algorithm can be suitable for implementation when encoding video sequences of different resolutions: from low resolution (qcif) to high resolution (720p). About 25.1% to 56.8% of time reduction for IBPBP case and 25.5% to 67.8% of time reduction for IBBPBBP case can be achieved. In addition, our algorithm can be combined with other inter mode or intra mode fast scheme to achieve even more time saving gain.

5. Conclusions

A fast spatial DIRECT mode decision algorithm is contributed in this paper. It targets both DIRECT_16x16 and SKIP_16x16 in B frame. Statistical analysis is carried out based on mode selection and RD cost between the current DIRECT MB and the co-located MBs. In an early stage of the mode decision process, by checking the mode condition and predicted RD cost threshold, all the remaining motion-compensation, intra prediction processes can be skipped if the criteria are satisfied. Moreover, a dynamic parameter update model is proposed to balance quality and time saving. Based on this method, the scheme can achieve significant time savings as compared to the original full mode algorithm (up to 56.8% of encoding time for IBPBP picture group and up to 67.8% of encoding time for IBBPBBP picture group) and existing fast scheme adopted in JM software (additionally, up to 40.1% time saving for IBPBP picture group and up to 49.5% time saving for IBBPBBP picture group under general QP settings).

Acknowledgements

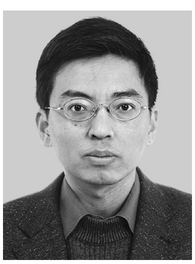
This work was supported by KAKENHI (23300018).

References

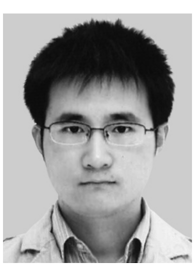
[1] Advanced Video Coding For Generic Audiovisual Services, ISO/IEC 14496-10:2005(E) ITU-T Rec. H.264 (E), March 2005.
[2] "Information technology-generic coding of moving pictures and associated audio information: Video, iso/iec 13818-2 and ITU-T Rec.H.262," 1996.
[3] "Video coding for low bit rate communication, ITU-T Rec.H.263," 1998.
[4] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.560–576, July 2003.
[5] M. Flierl and B. Girod, "Generalized B pictures and the draft H.264/AVC video-compression standard," IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.587–597, July 2003.
[6] A.M. Tourapis, F. Wu, and S.P. Li, "Direct mode coding for bipredictive slices in the H.264 standard," IEEE Trans. Circuits Syst. Video Technol., vol.15, no.1, pp.119–126, Jan. 2005
[7] I.E.G. Richardson (2003, April 30), "H.264/MPEG-4 Part 10 white paper: inter prediction," available at http://www.vcodex.com/files/h264_interpred.pdf
[8] I.E.G. Richardson (2003, April 30), "H.264/MPEG-4 Part 10 white paper: intra prediction," available at http://www.vcodex.com/files/h264_intrapred.pdf
[9] Y.W. Huang, B.Y. Hsieh, S.Y.M. Shao, Y. Chien, and L.G. Chen, "Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol.16, no.4, pp.507–522, April 2006.
[10] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," IEEE Circuits and Systems Magazine, vol.4, no.1, pp.7–28, First Quarter, 2004.
[11] J. Lee, I. Choi, W. Choi, and B. Jeon, "Fast mode decision for B slice," JVT-K021, Antalya, Turkey, Dec. 2003.
[12] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," IEEE Trans. Circuits Syst. Video Technol., vol.16, no.12, pp.1557–1561, Dec. 2006.
[13] I.E.G. Richardson (2003, April 30), "H.264/MPEG-4 Part 10 white paper: variable length coding," available at http://www.vcodex.com/files/h264_vlc.pdf
[14] B.Y. Chen and S.H. Yang, "Using H.264 coded block patterns for fast inter-mode selection," Proc. ICME 2008, Hannover, Germany, June 2008.
[15] M. Nieto, L. Salgado, and J. Cabrera, "Fast mode decision on H.264/AVC main profile encoding based on PSNR predictions," Proc. ICIP 2006, Atlanta, USA, Oct. 2006.

[16] Joint Video Team Reference Software, Version 11.0, available at http://iphome.hhi.de/suehring/tml/download/old_jm
[17] D. Wu, F. Pan, K.P. Lim, S. Wu, Z.G. Li, X. Lin, S. Rahardja, and C.C. Ko, "Fast intermode decision in H.264/AVC video coding," IEEE Trans. Circuits Syst. Video Technol., vol.15, no.6, pp.953–958, July 2005.
[18] F. Pan, X. Lin, S. Rahardja, K. Lim, Z. Li, D. Wu, and S. Wu, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," IEEE Trans. Circuits Syst. Video Technol., vol.15, no.7, pp.813–822, July 2005.
[19] J.L. Lee, D.S. Jun, and H.W. Park, "An efficient and fast mode decision method for inter slice of H.264/AVC," Proc. ICIP 2009, Cairo, Egypt, Nov. 2009.
[20] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.74–90, Nov. 1998.
[21] Y.C. Lin, T. Fink, and E. Bellers, "Fast mode decision for H.264 based on rate-distortion cost estimation," Proc. ICASSP 2007, Hawaii, USA, April 2007.

Xiaocong Jin received the B.E. degree of electronic engineering from Shanghai Jiaotong University, China in 2008 and M.E. degree in Information, Production and Systems Engineering from Waseda University, Japan in 2010. He is currently pursuing master degree in the Institute of Image communication and Information Processing of Shanghai Jiaotong University. His research interest includes video coding and related hardware architecture.



Jun Sun is presently a professor in the Institute of Image communication and Information Processing of Shanghai Jiaotong University, China. He is also a member of The HDTV Technical Executive Experts Group of China. His research interest includes image communication, multimedia technology, high definition TV broadcasting, video coding and its standards.



Yiqing Huang received a B.Eng. degree in communication and information engineering from Shanghai University in China. He received M.S. and Ph.D. degrees, both in system LSI (large-scale-integration) design, from Waseda University in Japan. From 2010, Doctor Huang joined RICOH Research and Development Center in Yokohama, Japan. His research interest includes image processing, video compression, television conference system, and integrated circuit design.

Jia Su received the B.E. degree in Telecommunications Engineering school of Xidian University, China, in 2006; and received M.E. degree both in Graduate School of Information, Production and Systems, Waseda University and School of Microelectronics in Xidian University in 2008 and 2009, respectively. She is currently working towards her Ph.D. degree at Waseda University. Her research interests include video compression and computer vision.

Takeshi Ikenaga received his B.E. and M.E. degrees in electrical engineering and Ph.D. degree in information & computer science from Waseda University, Tokyo, Japan, in 1988, 1990, and 2002, respectively. He joined LSI Laboratories, Nippon Telegraph and Telephone Corporation (NTT) in 1990, where he had been undertaking research on the design and test methodologies for high performance ASICs, a real-time MPEG2 encoder chip set, and a highly parallel LSI & system design for image-understanding processing. He is presently a professor in the Graduate School of Information, Production and Systems and the School of Fundamental Science and Engineering, Waseda University. His current interests are application SoCs for image and video processing, which covers video compression (e.g. H.264/AVC, H.264/SVC, H.265/HEVC), video filter (e.g. super resolution, noise reduction), video recognition (e.g. feature point detection, object tracking) and video communication (e.g. UWB, LDPC, public key encryption). He also has interests in application-oriented many-core processor design. Dr. Ikenaga is a member of IEEE, IPSJ and IIEEJ.