A Novel Fast Inter-Prediction Mode Decision for H.264/AVC

Yi Guo*, Houqiang Li*, Shibao Pei*, Changwen Chen**
*Department of Electric Engineering and Information Science
MOE-Microsoft Key Laboratory of Multimedia Computing and Communication
University of Science and Technology of China, Hefei, 230027
**Department of Electrical and Computer Engineering
Florida Institute of Technology, Melbourne, FL 32901
(e-mail: guoyi@mail.ustc.edu.cn*, lihq@ustc.edu.cn*, psb@ustc.edu*, cchen@fit.edu**)

ABSTRACT

Video coding in wireless environment requires lower computational complexity and lower energy consumption than that used in storage oriented or network oriented application. Although H.264/AVC standard provides considerable higher compression efficiency as compared to the previous standards, its complexity is significantly increased at the same time. In a H.264/AVC encoder, the most time-consuming components are variable block sizes motion estimation and mode decision using rate-distortion optimization (RDO). In this paper, we propose a novel fast inter-prediction mode decision by exploiting high correlation between rate-distortion costs (RD cost) of macroblocks in the current inter frame and their co-located macroblocks in the previous inter frame. Using this new algorithm, we can reduce a number of inter mode candidates and skip motion estimation for these modes. In addition, our algorithm can also decrease a number of tested intra modes. Simulation results show that our approach can save 20% to 50% encoding time, with a negligible PSNR loss less than 0.1 dB and bit rate increase no more than 2% for almost all the test sequences.

Keywords: video coding, H.264/AVC standard, mobile device, inter prediction, Rate Distortion Optimization (RDO), mode decision, RD cost, encoder complexity

1. INTRODUCTION

Video communication over wireless links is a challenging task. There are two fundamental constraints in video coding for mobile wireless transmission: low bit rate transmission and limited computational capacity [1]. Low bit rate transmission requires video signals to be highly compressed by complex coding algorithm, whereas limited computational capacity demands that the coding algorithms should be simple so as to be implemented on mobile devices. Nowadays, H.264/AVC [2] offers a significant improvement of coding efficiency compared to previous ITU-T and ISO/IEC standards. However, much more complexity is brought into its scheme due to newly adopted algorithms such as variable block sizes in intra prediction modes and inter prediction modes, quarter-pixel motion estimation and motion compensation, deblocking filter in the reconstruction loop, context-adaptive variable-length coding (CAVLC), context-based adaptive binary arithmetic coding (CABAC), and rate-distortion optimization (RDO) in encoder, etc. Among these techniques, full search of all modes in mode decision process by using RDO contributes a lot to the computation complexity.

H.264/AVC supports both intra prediction mode and inter prediction mode for each macroblock. Inter prediction mode is used in variable block sizes motion estimation and motion compensation to eliminate the temporal redundancy coming from high correlation between consecutive frames. Variable block sizes motion estimation and motion compensation employs a tree-structured hierarchical macroblock partition method which has eight types of inter prediction mode (Skip, Inter16×16, Inter16×8, Inter8×16, Inter8×8, Inter8×4, Inter4×8, Inter4×4), as is shown in Figure 1. Skip mode represents the case where the block size is 16×6 but no motion and residual information are coded, so it has the lowest computational complexity. Intra prediction mode is used to remove the spatial redundancy by referring to neighboring samples of previously coded blocks which are to the left of and/or above the block. For the luma samples, the size of intra prediction mode is either 4×4 (Intra4×4) or 16×16 (Intra16×16). There are a total of nine optional

Multimedia on Mobile Devices II, edited by Reiner Creutzburg, Jarmo H. Takala, Chang Wen Chen,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6074, 607402, © 2006 SPIE-IS&T · 0277-786X/06/$15

prediction modes for Intra4×4, four prediction modes for Intra16×16, and four prediction modes for the chroma components.

RDO is used to achieve the much better performance in terms of minimizing compressed video data bits and maximizing coding quality [7]. For each inter mode, prior to mode decision, RDO motion estimation must be used to find the optimal motion vectors, which can achieve a good compromise between the distortion and the rate to code the motion vectors. After variable block sizes motion estimation, RDO mode decision is conducted to select the best one in all inter and intra modes which can minimize RD cost. Although this full search process can provide excellent coding efficient, it also results in a heavy burden of computation complexity.

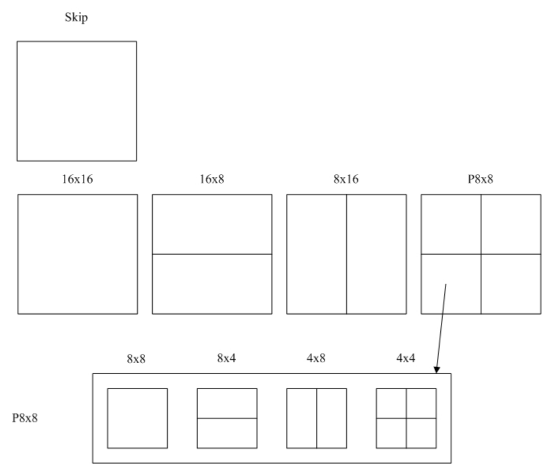

Figure 1 The Inter prediction modes for macroblock

In order to reduce the computation complexity of the mode decision process, many methods have been proposed [3]-[8], each of which uses its own criterion to predict the best mode. The method described in [3] is based on the simulation results that about 70% of the macroblocks choose Inter16×16 mode as their final mode, so the authors propose to use accumulated sum absolute difference (SAD) of 16×16 to calculate a threshold and by which to judge whether should early terminate the search of Inter16×8 and Inter8×16 modes. The fast inter mode decision algorithm proposed in [4] depends only on the absolute differences between successive frames. It is based on the viewpoint that the absolute frame difference contains lots of motion information, for example, large difference values often appear on the edges or boundaries of moving objects while small values usually exist in homogeneous areas. So the mean absolute frame difference (MAFD) of the current frame and mean absolute difference (MAD) of the current macroblock are used to determine whether the current macroblock belongs to homogeneous regions or not. Therefore, if the difference values in a macroblock are small, it is most likely that this macroblock belongs to a homogeneous region and inter prediction modes with larger block sizes should be more proper for it. Otherwise, motion in the macroblock may be complex and using smaller block sizes can achieve better rate distortion performance. In [5], the best mode for the current macroblock is obtained by analyzing the modes for a maximum of four macroblocks in the current and previous frames. It is based on the assumption that the frames in a video sequence are spatially and temporally related. The method proposed in [6] begins with the calculation of the cost of three modes, Inter16×16, Inter8×8, and Inter4×4, and checks if the cost tends to monotonically increase or decrease with the block size. If there is a monotonic tendency, only the modes between the best two modes are tested. Otherwise, all other modes are tested. For example, if the cost is indeed monotonic and Inter16×16 and Inter8×8 are found to be better than Inter4×4, the method then tests Inter16×8 and Inter8×16 modes in the first group and omits Inter8×4 and Inter4×8 modes in the other group. The method described in [7] is based on the observation that homogeneous regions tend to move together and hence should not be split into smaller blocks. The homogeneity of a block is determined by using the amplitude of the edge vector computed by the Sobel operator. A framework brought up in [8] is based on the mode distribution and the mode similarity of neighboring macroblocks. It requires the computation of the two thresholds, the first one for Skip mode and the second one for mode similarity test. The other information used in the method, such as the modes of neighboring macroblocks, is already available in the encoding processes. Therefore, it requires very little computation time to predict the best mode.

In our study, we find that the RD curves of two consecutive frames with no significant scene changes may approximate to each other, and there is high correlation between the RD costs of macroblocks in the current inter frame and those of their co-located macroblocks in the previous inter frame. Therefore, in this paper, we propose a novel inter-prediction mode decision by exploiting such high correlation. Our algorithm reduces a number of potential inter modes, meanwhile, eliminates motion estimation for these modes. In addition, our algorithm can also decrease a number of tested intra modes.

The paper is organized as follows. Section 2 gives an overview of RDO in H.264/AVC. Section 3 describes our proposed method. The simulation result is presented in Section 4, and finally concluding remarks are given in Section 5.

2. OVERVIEW OF RDO IN H.264/AVC

The H.264/AVC is a new generation video coding standard jointly developed by ITU-T and ISO/IEC. It is expected that H.264/AVC will be adopted in various applications ranging from handheld devices to HDTV in the near future. H.264/AVC has achieved a significant improvement in rate-distortion performance compared to existing standards by using various techniques as stated in the above section.

In H.264/AVC, partitions with luma block sizes of 16×16, 16×8, 8×16 and 8×8 are supported. In case Inter8×8 mode is chosen, each of the 8×8 submacroblocks can be further partitioned into blocks of 8×8, 8×4, 4×8, 4×4 luma samples [9]. In addition to inter modes, Intra4×4 and Intra16×16 are also provided to remove the spatial redundancy. For large areas with no change or constant motion, Skip mode can be very useful. In order to choose the best mode which can achieve much better coding performance in terms of minimizing compressed video data bits and maximizing coding quality, RDO is adopted in H.264/AVC as one of the essential parts of the whole encoder.

2.1 RDO

The overall video codec performance can be increased by using various efficient coding options. The rate distortion optimization task is to choose the most efficient coded representation (segmentation, prediction modes, motion vectors, quantization parameters (QP), reference frames, etc.) in the RD sense.

The coding options, which are assigned to the coding process, have different RD characteristics, and the goal of an encoder is to optimize its overall fidelity: Minimize distortion D, subject to a R_c constraint, on the number of bits used R. This constrained problem reads as follows

$$\min\{D\}, \text{ subject to } \; R < R_c \tag{1}$$

The optimization task in equation (1) can be elegantly solved using Lagrangian optimization where a distortion term is weighted against a rate term. The Lagrangian formulation of the minimization problem is given by

$$\min\{J\}, \text{ where } \; J = D + \lambda R \tag{2}$$

where the Lagrangian RD functional J is minimized for a particular value of the Lagrange multiplier λ . Each solution to equation (2) for a given value of the Lagrange multiplier λ corresponds to an optimal solution to equation (1) for a particular value of R_c

This technique has gained importance due to its effectiveness, conceptual simplicity, and its ability to effectively evaluate a large number of possible coding choices in an optimized fashion [10].

H.264/AVC uses RDO to choose the best segmentation, prediction modes, reference frames, motion vectors and direction of prediction by using different λ and distortion measurement. For inter frame, RDO is mainly used in motion estimation and mode decision.

2.2 Motion Estimation

In H.264/AVC, the motion estimation is performed on integer-pixel positions followed by sub-pixel refinement. For integer-pixel motion estimation, the cost function is calculated as:

$$J_{ME-SAD}(\mathrm{m}|\lambda_{Motion}) = SAD(s, r(\mathrm{m})) + \lambda_{Motion} \cdot R(\mathrm{m}-\mathrm{p}) \tag{3}$$

For sub-pixel refinement, the cost function is calculated as:

$$J_{ME-SATD}(\mathrm{m}|\lambda_{Motion}) = SATD(s, r(\mathrm{m})) + \lambda_{Motion} \cdot R(\mathrm{m}-\mathrm{p}) \tag{4}$$

where $\mathrm{m} = (m_x, m_y)^T$ is the motion vector, $\mathbf{p} = (p_x, p_y)^T$ is the prediction for the motion vector and λ_{Motion} is the Lagrange multiplier for motion estimation which is given by

$$\lambda_{Motion} = \sqrt{0.85 \cdot 2^{(QP-12)/3}} \tag{5}$$

R(m-p) denotes the bits used for motion information. *SAD*(*s*,*r*(m)) is sum of absolute difference, and it is computed as:

$$SAD(s, r(\mathrm{m})) = \sum_{x=1, y=1}^{B_1, B_2} |s[x, y] - r[x - m_x, y - m_y]| \tag{6}$$

where $B_1 \times B_2$ is the block size used for motion estimation and mode decision; *s* is the current frame to be coded; *r* is the reconstructed reference frame. Whereas *SATD*(*s*,*r*(m)) is sum of absolute transformed differences, which can be seen as an estimate of the rate cost for encoding. Hence, the *SATD* could operate as an approximative RD cost criterion used for motion vector search [11].

2.3 Mode Decision

The mode decision process is to minimize the following Lagrangian cost:

$$J_{Mode}(s, c, Mode \mid \lambda_{Mode}, QP) = SSD(s, c, Mode \mid QP) + \lambda_{Mode} \cdot R(s, c, Mode \mid QP) \tag{7}$$

where λ_{Mode} is chosen as follow

$$\lambda_{Mode} = 0.85 \cdot 2^{(QP-12)/3} \tag{8}$$

Mode indicates a mode out of a set of potential macroblock modes:

$$Mode \in \begin{Bmatrix} Intra\,4\times4,\ Intra\,16\times16,\ Skip, \\ 16\times16, 16\times8, 8\times16, 8\times8, 8\times4, 4\times8, 4\times4 \end{Bmatrix} \tag{9}$$

the distortion measurement *SSD*(*s*,*c*,*Mode*|*QP*) denotes sum of square difference between the original signal *s* and its reconstructed signal *c*, which is given as

$$SSD(s, c, Mode \mid QP) = \sum_{x=1, y=1}^{16,16} \left(s[x, y] - c[x, y, Mode \mid QP]\right)^2 \tag{10}$$

R(*s*,*c*,*Mode*|*QP*) reflects the number of bits associated with choosing the mode and *QP*, including the bits for the macroblock header, the motion vectors and all the DCT residue blocks.

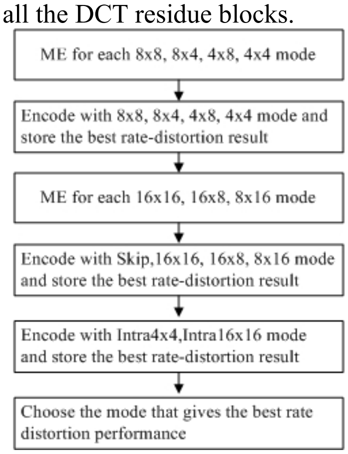

Figure 2 Inter frame mode decision in H.264/AVC

The JVT reference software [12] is based on a RDO framework for both motion estimation and mode decision. For inter modes, RDO motion estimation is done first, then given the motion vectors, RD cost can be calculated. For Intra4×4, Intra16×16 and Skip modes, RD cost can be calculated without motion estimation and motion compensation. With regard to each mode, bit rate and distortion should be calculated by actually encoding and decoding the video, then the best mode can be decided in terms of RD cost; thus, the computation complexity increases dramatically compared to the previous standards. The procedure to encode one macroblock is shown in Figure 2.

3. PROPOSED FAST MODE DECISION METHOD FOR INTER FRAME

In H.264/AVC, examining all possible modes, especially variable bock sizes motion estimation and motion compensation is a big burden on the encoder and is very time consuming. So it is very beneficial to alleviate the complexity on mode decision. In our study, we find that the RD curves of two consecutive frames with no significant scene changes may approximate to each other, and there is a high correlation between the RD costs of macroblocks in the current inter frame and those of their co-located macroblocks in the previous inter frame. Therefore, we bring up a fast mode decision scheme which search the mode in an order, and use the co-located macroblock's RD cost as a judgement to early terminate the search. Our algorithm reduces a number of potential inter modes, meanwhile, eliminates motion estimation for these modes. In addition, our algorithm can also decrease a number of tested intra modes. In this section, we first describe the ideas behind our method, and then detail the corresponding algorithm.

3.1 RD curve and RD cost

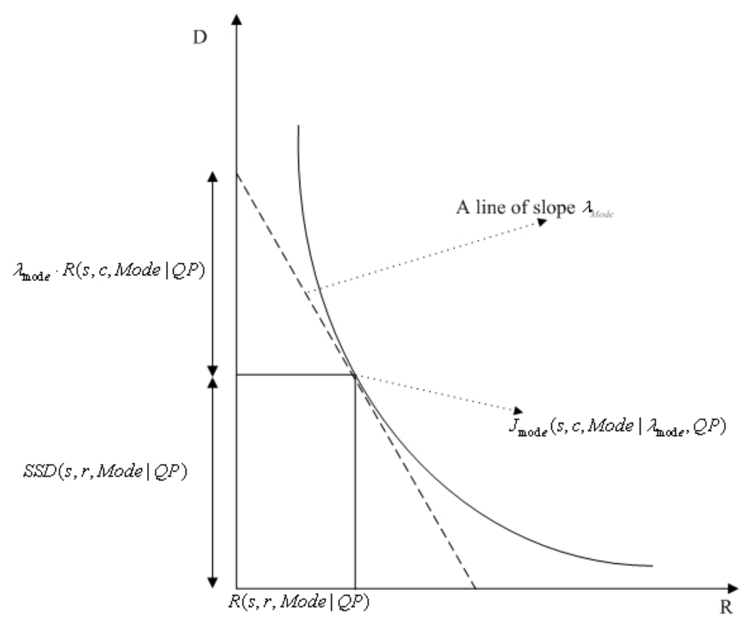

Figure 3 Get optimal RD point from RD curve

Figure 3 shows the RD plane, where the solid curve is the RD curve of one macroblock, and the dash line tangent to the solid curve is the line with slope of λ_{Mode}. To compute the RD cost of one macroblock, we first have to choose λ_{Mode} according to QP given by equation (8). In order to find the optimal RD point at the given λ_{Mode}, we start with a line of that slope passing through the origin and keep moving in the northeast direction. The point at which the sweeping line first intersects the RD curve corresponds to the optimal RD point. Therefore, the RD cost in equation (7) can be calculated.

Figure 4 Six frames picked from the sequence of "foreman" in QCIF: (a) The 273rd ; (b) The 274th; (c) The 275th (d) The 301st; (e) The 302nd; (f) The 303rd.

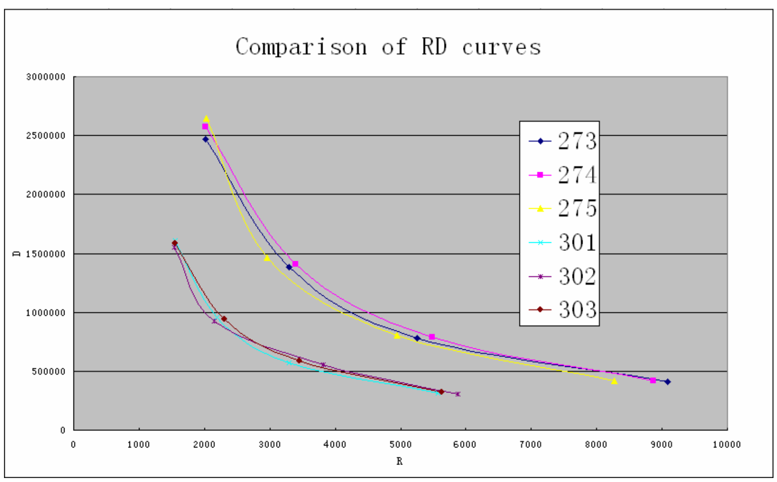

Figure 5 Comparison of RD curves

When there are no significant scene changes in the video sequence, from the view of information theory, the source signal may be no change, so its probability distribution remains unchanged, thus the RD curve of the source signal may vary slowly. For example, we pick six frames from the sequence of "foreman" in QCIF: the 273rd, the 274th, the 275th, the 301st, the 302nd, and the 303rd, as can be seen in Figure 4. The 273rd, the 274th and the 275th frame belong to the first group of three successive frames with no scene change, and the 301st, the 302nd and the 303rd frames belong to another group of three successive frames with no scene change. But compared with the first three frames, the second three frames is not in the same scene. We use QP 28, 32, 36 and 40 to get the RD curves of all the six frames, which are shown in Figure 5. The RD curves of the first group frames are approximate to each other, so do the second group frames. But the curves of the first group frames are apart from those of the second group frames. Based on the observation, we can draw the conclusion that for frames with no significant scene changes, their RD curves may be very approximate. Furthermore, we can assume that the RD curves of macroblocks in the current frame and those of their co-located macroblocks in the previous inter frame may be approach to each other in the case that there is no significant scene changes in the current frame. By Figure 3, it can be deduced that we can get similar RD costs using the same QP.

In order to verify our assumption, we plot the average RD cost ratio of macroblocks in the current inter frame to their co-located macroblocks in the previous inter frame at one given QP in Figure 6, as can be seen that the ratio approximates to a fixed value, which is around 1.0.

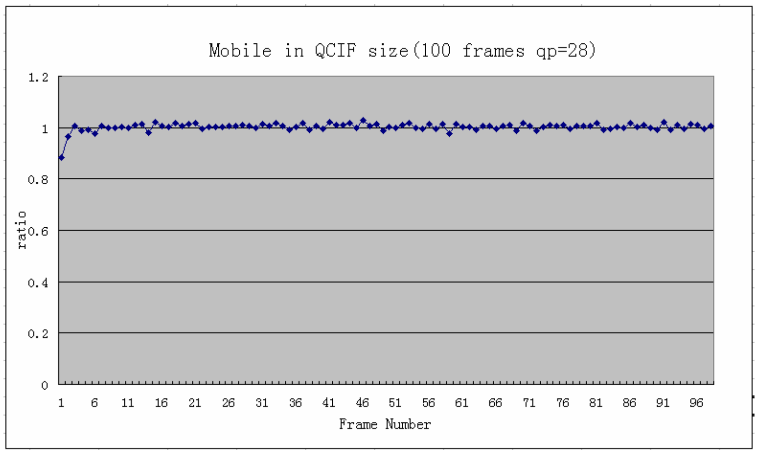

Figure 6 Average RD cost ratio of macroblocks in the current inter frame to their co-located macroblocks in the previous inter frame

3.2 Our proposed method

Based on the above assumption and observation, we propose to employ the co-located macroblock's RD cost as a judgement to select the best mode in the current macroblock and to terminate the mode decision as early as possible. Moreover, we find that the average RD cost of the Skip mode is much less than the other four modes (16×16, 16×8, 8×16, and P8×8 (including 8×8, 8×4, 4×8, 4×4)), as can be seen from Figure 7, so we could set a lower threshold in the case of selecting Skip mode additionally. This circumstance may happen when the best mode of the current macroblock is Skip while the co-located macroblock in the previous inter frame is other modes.

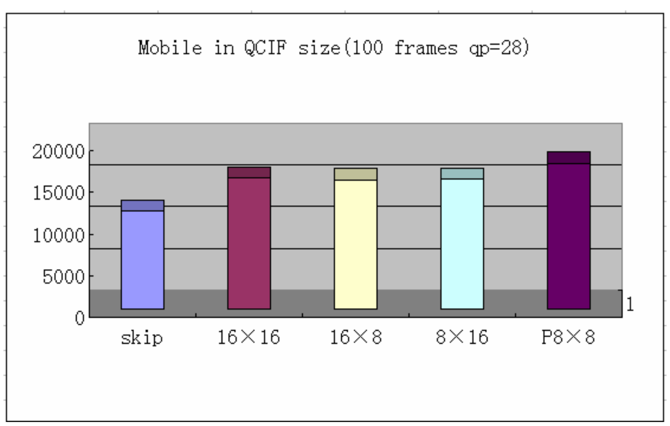

Figure 7 Average RD cost comparison for skip, 16×16, 16×8, 8×16, P8×8 modes

We have found that, in general, the Skip mode or the 16×16 mode is likely the best mode for macroblocks in the background or the smooth regions of the image. On the other hand, the 8×8 mode or the 4×4 mode is the favorite choice for macroblocks that are on the boundary of objects or in the fast moving regions of the image. In [8], statistical results show that about 50% of the modes are Skip mode; more than 10% are 16×16 mode; the occurrence of the remaining modes has such an order: Inter16×8, Inter8×16, Inter8×8 (including 8×8, 8×4, 4×8, 4×4), Intra4×4, Intra16×16. Thus, in our proposed scheme, the Skip mode is the initial guess for the mode decision, and the order of other modes is as follows: Inter16×16, Inter16×8, Inter8×16, Inter8×8 (including 8×8, 8×4, 4×8, 4×4), Intra4×4, Intra16×16.

Let T be the co-located macroblock's RD cost in previous inter frame; α be the lower threshold factor for Skip mode which can stop the mode selection as early as possible; β be the range reference factor for all modes which is used to decide whether J (mode) is near to T; c be the fluctuation of β. Our fast mode decision algorithm consists of the following steps:

Step1 Check Skip mode. If J (Skip) is lower than αT or in the range $((\beta-c)T, (\beta+c)T)$, set Skip mode as the best mode and then stop.

Step2 Check the modes in the order of {Iner16×16, Inter16×8, Inter8×16, Inter8×8 (including 8×8, 8×4, 4×8, 4×4), Intra4×4 and Intra16×16}. Once the RD cost of one mode is in the range $((\beta-c)T, (\beta+c)T)$, then set the corresponding mode as the best mode and stop.

We set the initial value of factors α, β, c to 0.70, 1.0, 0.05 respectively and update them according to the previous RD cost ratio dynamically. The flow chart of our proposed algorithm is demonstrated in Figure 8.

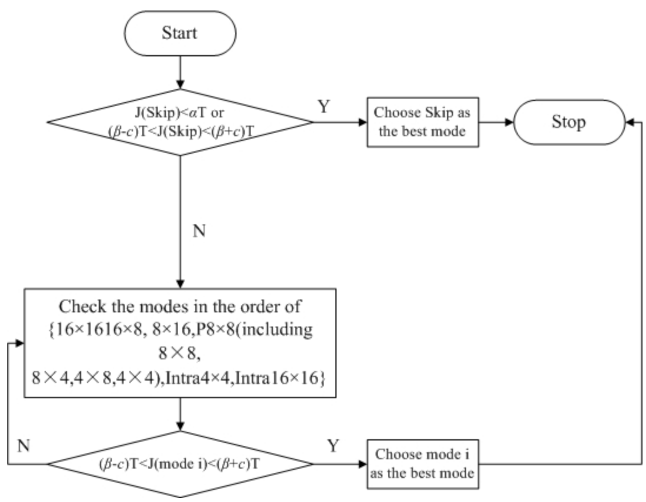

Figure 8 The proposed fast mode decision algorithm

4. SIMULATION RESULTS

In order to verify our proposed approach, we implemented it into JVT reference software JM6.1e [12]. The test conditions are as follows.

1) MV search range is 16 pels.
2) Reference frame number equals to 1.
3) CABAC is disabled.
4) MV resolution is 1/4 pel.
5) GOP structure is IPPPP….P
6) 30fps.

We have tested seven QCIF video sequences of different characteristics with QP equal to 20, 24, 28, 32, 36, and 40 respectively. The comparison results were produced and tabulated based on the percentage of reduction coding time

Δ T, the PSNR difference of Luma(Y) ΔPSNR and the bit-rate difference Δ R. To simplify our comparison, we have given the average results based on the above QP. The test platform is Pentium IV-3.0 GHz with 1.0G RAM.

Table 1 Performance of the Proposed Algorithm

Sequence	Bus	Football	Foreman	Mobile	News	Paris	Tempete
Δ R(%)	1.70	3.23	1.93	1.00	1.87	2.58	1.23
Δ PSNR(dB)	-0.054	-0.111	-0.063	-0.061	-0.016	-0.0531	-0.056
Δ T(%)	-24.9	-19.0	-21.8	-32.0	-26.0	-31.4	-25.1

Simulation results show that our approach can reduce 20% to 30% average encoding time for all the sequences, with a negligible PSNR loss less than 0.1 dB and bit rate increase no more than 2%, as can be shown in Table 1, which demonstrate clearly that our proposed approach is very effective.

We also give the RD curves of two sequences "mobile" and "paris" in Figure 9 and Figure 11, as can be seen that our proposed algorithm has similar RD performance as that of JM6.1e. The reduction percentage of coding time can also be seen from Figure 10 and Figure 12, which prove that our approach saves much time.

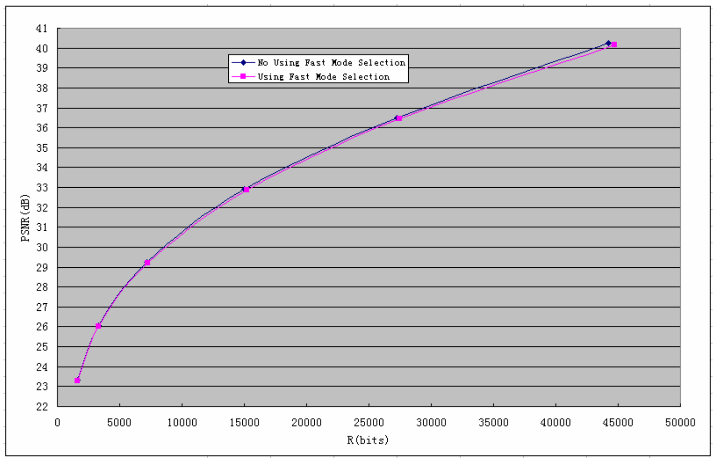

Figure 9 The R-PSNR curve comparison of JM6.1e and proposed method for "Mobile" in QCIF

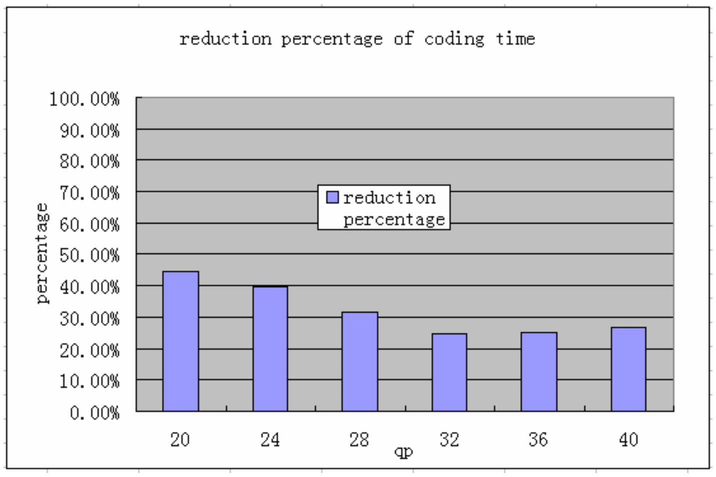

Figure 10 The reduction percentage of coding time comparison of JM6.1e and proposed method for "Mobile" in QCIF

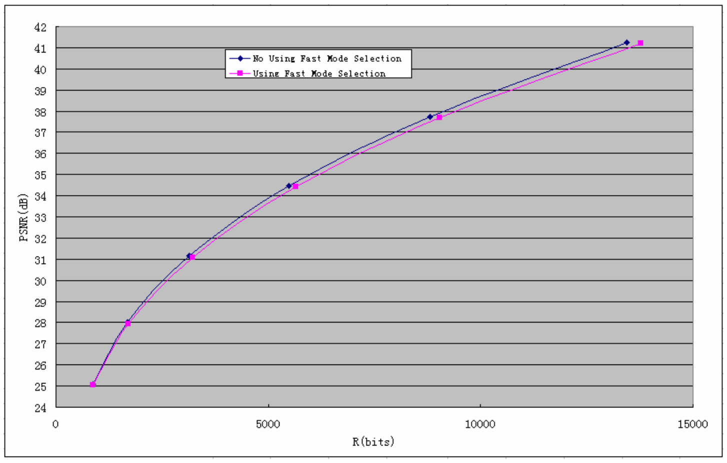

Figure 11 The R-PSNR curve comparison of JM6.1e and proposed method for "Paris" in QCIF

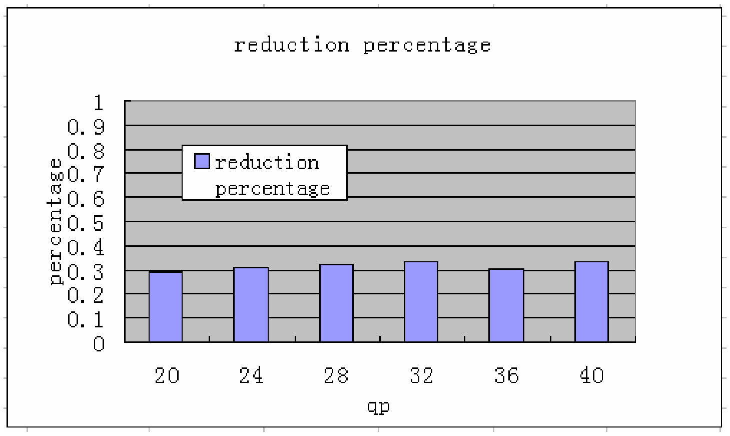

Figure 12 The reduction percentage of coding time comparison of JM6.1e and proposed method for "Paris" in QCIF

5. CONCLUSION

In this paper, we propose a novel inter-prediction mode decision for H.264/AVC by exploiting a high correlation between the RD costs of macroblocks in the current inter frame and those of their co-located macroblocks in the previous inter frame. The experimental results show that our proposed method is very effective. However, in this paper, our assumption and observation is based on the frames without significant scene changes, so for the video sequence which has violent motion, our method seems to be not so effective. How to tackle such problem should be our future work. Moreover, in this paper, we only exploit the temporal correlation of the RD cost. As a future work, we would try to induce the spatial correlation of the RD cost into this scheme.

ACKNOWLEDGEMENT

This research is partially supported by NSFC under contract No. 60333020 and open fund of MOE-Microsoft Key Laboratory of Multimedia Computing and Communication under contract No. 05071803.

REFERENCE

1. Yi Wang, Houqiang Li, Chang Wen Chen, "A novel video coding scheme for mobile devices," Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia, Vol. 83, Pages: 125 – 131, College Park, Maryland, 2004.
2. "Draft ITU-T Rec. and FDIS of Joint Video Spec. (H.264 | ISO/IEC 14496-10 AVC)" *JVT of MPEG and VCEG, Doc. JVT-G050r1*, May 2003
3. Andy Chang, Oscar C. Au, Y. M. Yeung, "A Novel Approach to Fast Multi-Block Motion Estimation for H.264 Video Coding, " *Proc. of IEEE Int. Conf. on Multimedia & Expo (ICME' 03)*, Baltimore, Maryland, USA, July 2003.
4. Jing, X.; Chau, L.-P. "Fast approach for H.264 inter mode decision, " *Electronics Letters* Volume 40, Issue 17. Aug. 2004
5. Ahmad, A.; Khan, N.; Masud, S.; Maud, M.A. "Efficient block size selection in H.264 video coding standard ," *Electronics Letters* Volume 40, Issue 1. Jan. 2004
6. P. Yin, H. Y. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for H.264," *IEEE Int'l Conference on Image Processing*, vol. III, pp. 853-856, Sept. 2003.
7. D. Wu, S. Wu, K. P. Lim, F. Pan, and X. Lin, "Block inter mode decision for fast encoding of H.264," *IEEE Int'l Conference on Speech, Acoustics, and Signal Processing*, vol. III, pp. 181-184, May 2004.
8. C.-Y. Chang, C.-H. Pan, and H. H. Chen, "Fast mode decision for P frames in H.264," *Picture Coding Symposium*, San Francisco, Dec. 2004
9. 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, pp. 560–576, July 2003.
10. G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74–90, Nov. 1998.
11. Wien, M, "Variable block-size transforms for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 604 - 613, 2003.
12. JVT Reference Software version 6.1 http://bs.hhi.de/~suehring/tml/