# A Fast Intra/Inter Mode Decision Algorithm of H.264/AVC for Real-time Applications

Bin Zhan, Baochun Hou, and Reza Sotudeh
School of Electronic, Communication and Electrical Engineering
University of Hertfordshire
Hatfield, AL10 9AB, United Kingdom
{b.zhan, b.hou, r.sotudeh}@herts.ac.uk

***Abstract*—A novel fast Intra/Inter mode decision algorithm for H.264/AVC is proposed in this paper. By reducing the calculations incurred in both intra- and inter-prediction in video encoding, the computational complexity is significantly decreased, which is an essential requirement for real-time video communications applications. The motion vectors of previously encoded macroblocks are exploited to detect the motion speed according to the pre-defined threshold, and the degradation of encoding efficiency can be minimized by adaptive adjustment scheme based on motion speed. The experimental results showed that the proposed fast mode decision algorithm is able to reduce 51.29% total encoding time on the average with a negligible PSNR loss less than 0.1 dB and a bitrate increase under 2%.**

***Index Terms*—H.264/AVC, video coding, fast mode decision, intra prediction, inter prediction**

## I. Introduction

Video communication is one of the hottest topics in telecommunication and broadcasting industry nowadays. However, raw video streams are difficult to transmit without compression because the limited bandwidth is still the most important bottleneck of data communication. Former video compression standards, such as H.261 [1] and H.263 [2], have to decrease the video quality in order to satisfy the real-time requirement. H.264/AVC [3] (also known as MPEG-4 part 10) is the latest video compression standard, which is jointly developed by ISO and IEC. H.264/AVC achieves extremely well encoding performance in terms of video quality and compression ratio than other standards by adopting a number of new techniques: multiple directions of intra-prediction, flexible block size based motion estimation (ME), quarter-pixel accuracy and multi-references ME, rate distortion optimization (RDO) enabled mode decision (MD), in-the-loop de-blocking filtering, and so on [4, 5]. Although the encoding performance is notably enhanced, the incurred computational complexity is the new challenge for real-time applications. The RDO enabled MD introduces major time cost, and a lot of fast MD algorithms were therefore proposed to speed up the encoding process. C. Grecos and M. Y. Yang [6] proposed a fast inter MD algorithm, which combines the optimized skip MD and the smoothness constraints based MD. F. Pan et al. presented an approach to reduce the intra candidate modes by using a pre-established local edge direction histogram in [7]. They also introduced a new fast MD scheme by means of the spatial homogeneity and temporal stationarity characteristics of video objects [8]. J. F. Wang et al. in [9] adopted the dominant edge strength to estimate the possible directions to improve the intra MD efficiency. Recently, we presented two fast MD algorithms using statistic pattern characteristic of best candidate modes in [10, 11]. In this paper, a novel fast MD algorithm is proposed based on the correlation characteristics between macroblocks in the adjacent frames. The motion speed derived from the motion vector (MV) is utilized to form a new subset candidate modes group to improve the encoding efficiency and an adaptive adjustment scheme is introduced to maintain the encoding performance. The proposed algorithm can save total encoding time 51.29% on average with negligible degradation of encoding performance.

The rest of this paper is organized as the following sections. In section II, the rate distortion optimized MD is briefly summarized. Section III presents the theory behind the proposed fast MD algorithm. The detailed algorithm is described in section IV. Experimental results and analysis are given in Section V and followed by the conclusion of this paper.

## II. Rate Distortion Optimized Mode Decision

The block based encoding technique is exploited for inter-prediction, intra-prediction and quantization in most video compression standards. Usually unique block size is adopted, and the encoder only needs to compare a few results to choose the best mode. Less block size means less time is spent on the encoding process. However, the compressed video quality is worse for high-detailed video scene in the sequences. In H.264/AVC standard, the original macroblock is further partitioned into seven sub-blocks with different block size. Flexible and smaller block size can encode each frame more accurate and effective; especially there are a large number of details in the video scene. Just like a sword has its two sides, although smaller size leads to fewer residuals after the predictions, more accessory data (from MVs and directions) sometimes compromises the improvement. As specified in H.264/AVC, RDO enabled MD balances the residuals and the accessory data to select the most suitable mode for each

978-1-4244-2075-9/08/$25.00 ©2008 IEEE

macroblock according to the rate-distortion (RD) cost, which is given in (1):

$$\begin{aligned} &RD\cos t(s,c,Mode/QP) = \\ &\quad SSD(s,c,Mode/QP) + \lambda_{Mode} \cdot R(s,c,Mode/QP,\lambda_{Mode}) \end{aligned} \tag{1}$$

where $s$ indicates the original macroblock, and $c$ indicates the re-constructed macroblock; *Mode* stands for one of the seven candidate macroblocks modes {Skip, inter-16×16, inter-16×8, inter-8×16, inter-P8×8, intra-16×16 and intra-4×4}, and each inter-P8×8 mode consists of four sub-macroblock modes. The sum of the squared differences (SSD) denotes the distortion between $s$ and $c$ macroblocks. $\lambda_{Mode}$ is the Lagrangian multiplier related to the quantization parameter (*QP*), and the definition for P-frame is in (2):

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

For each candidate mode, the RD-cost has to be calculated once, especially for the inter-prediction, a series of processes are involved such as ME, forward and inverse transform and quantization, entropy coding and so on. Seven macroblock modes and four sub-macroblocks make the exhaustive MD an extremely time-consuming process, and more than 50% video encoding time is spent on the exhaustive MD when RDO is enabled [12]. Therefore fast MD algorithm plays a very important role to enhance the encoding speed to enable the real-time video communications applications.

## III. Investigation and Motivations

### A. Correlation in MD

Usually, the successive frames exhibit strong correlation in temporal domain and spatial domain. By employing that characteristic, the spatial and temporal redundancy from the video sequence can be removed during intra prediction and inter prediction to achieve further compression. Intensive experiments indicate that the macroblock modes selected for the adjacent frames also possess high correlation. An example is given in Fig. 1, where $frame_1$ is previously encoded and used as the reference for $frame_2$, and the number in the figure indicates the adopted best mode. From the figure, it is clearly shown that the macroblocks in the $frame_2$ are highly correlated with co-located or neighbouring macroblocks in the $frame_1$. Here we introduce a dynamic mode list $L_{(x,\,y)}$ for macroblock (x, y), and is defined in (3):

$$mode_i \in L_{(x,y)} \begin{cases} if\ \ mode_i\ is\ the\ best\ \ for\ co\text{-}located\ MB \\ if\ \ mode_i\ is\ the\ best\ \ for\ any\ first\ neighbouring\ MB \\ if\ \ mode_i\ is\ the\ best\ \ for\ any\ second\ neighbouring\ MB \end{cases} \tag{3}$$

where $mode_i$ is one of the seven macroblock modes. Since there are insufficient reference modes for the macroblocks on the first edge of frame, only the first condition in (3) is applied. For the same reason, only the first two conditions in (3) are satisfied for the macroblocks on the second edge of frame. The first edge, second edge of frame and the neighbouring macroblocks are illustrated in Fig. 2. According to the correlation characteristics between macroblocks, the best mode for macroblock (x, y) should be one candidate in $L_{(x,\,y)}$. Sufficient simulation results indicated that the modes in the $L_{(x,\,y)}$ are usually a small subset of entire modes. Fig. 3 and Fig. 4 are examples to show the number of candidate modes in $L_{(x,\,y)}$ for Foreman and Bridge-close two sequences in QCIF. A large amount of quiescent backgrounds and the minor motions of objects are contained in Bridge-close, the number of modes in each candidate group is therefore very small (mostly equaling to 1 or 2). This characteristic is more obvious for video sequences with low motion or smooth background. Compared with the use of the exhaustive modes, dynamic list $L_{(x,\,y)}$, is composed of less candidate macroblock modes. Using these highly correlated modes in a dynamic group, the encoding time can be reduced significantly with negligible performance degradation in terms of PSNR and bitrate.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 3 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| 2 | 2 | 2 | 2 | 8 | 8 | 2 | 1 | 1 | 8 | 2 |
| 1 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 2 | 1 | 2 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 1 | 2 |

***$frame_1$***

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 1 | 1 | 2 | 1 | 1 | 3 | 0 | 1 | 1 |
| 3 | 8 | 8 | 8 | 8 | 2 | 2 | 1 | 1 | 1 | 8 |
| 1 | 0 | 1 | 10 | 8 | 0 | 1 | 1 | 1 | 3 | 2 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 |

***$frame_2$***

Fig. 1. Modes selected in two successive frames for Highway.

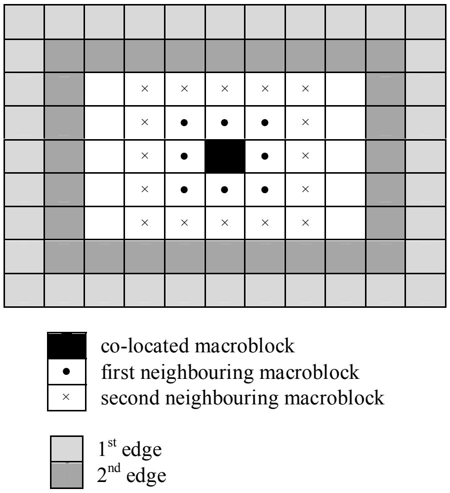


Fig. 2. Macroblocks in one QCIF frame.

### B. MD for Macroblocks on the First Edge of Frame

In section III-A, we discussed the fast MD for macroblocks and experimental results showed that about 35% encoding time can be saved. However, the macroblocks on the first edge

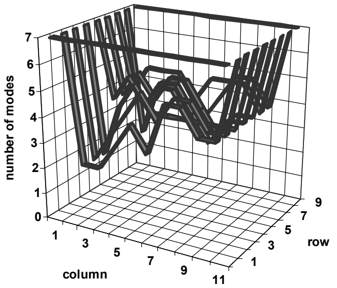


Fig. 3. Number of macroblock modes in $L_{(x, y)}$ of one frame for Foreman.

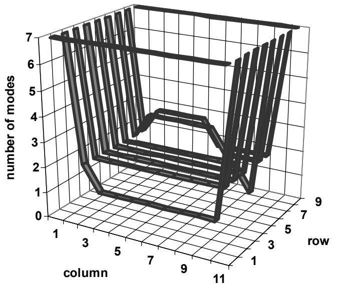


Fig. 4. Number of macroblock modes in $L_{(x, y)}$ of one frame for Bridge-close.

of frame possess notable percentage of all macroblocks: 36% for QCIF and 19% for CIF. Additional improvement of the MD process on those macroblocks could further reduce the encoding time. Since there are not sufficient reference macroblocks around the current macroblock on the edge of a frame, the spatial correlation in MD is weak. However, the temporal correlation characteristic can be exploited. Recall that skip mode is selected only when the macroblock (16×16) owes zero MVs and the residuals are also zeros, which means the macroblock is the same as its co-located macroblock in reference frame. Taking into account the high temporal correlation between successive frames, we add additional constraint into the dynamic list $L_{(x, y)}$ by modifying the fast MD scheme for macroblock on the first edge:

-- Only Skip and inter 16×16 two modes are contained in the list $L_{(x, y)}$ if the best mode of co-located macroblock is SKIP.

-- Otherwise, $L_{(x, y)}$ consists of all candidate modes.

### C. Motion Speed Based Adjustment Scheme

In most cases, the values of MVs (as shown in Fig. 5) provide information on the speed and the direction of movement of objects. Smaller values are obtained for relative static background and smoothly moving objects, and larger values indicate the speedy motion. Taking into consideration of the real life videos, the high-speed movement of objects involved from time to time. In these cases, the method introduced in section III-A could exclude the best mode from the dynamic list, and the encoding performance would be degraded if there is no further adjustment scheme deployed. In the proposed algorithm, MVs are exploited to form the criterion to adaptively adjust the current dynamic candidate mode list, $L_{(x, y)}$. We defined the displacement of MVs as the

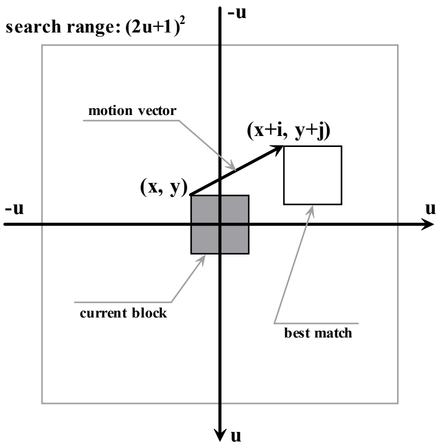


Fig. 5. Motion vector.

threshold to judge the status of movement:

-- If $MV_x >= TH$ or $MV_y >= TH$, the motion speed of objects contained in the current macroblock is regarded as fast;

where subscript $x$ indicates horizontal direction and $y$ indicates vertical direction. $TH$ is a predefined threshold used to detect the high-speed motion, and the threshold is set to 5 from large experiments. For simplicity, the best MVs for inter 16×16 mode are checked only. A flag will be set for the macroblock if MV exceeds the threshold. While encoding a macroblock, the fast MD strategy will be available if no flag is set in the co-located macroblock from the reference frame; otherwise, all modes will be enabled for the macroblock.

Based on the analysis given in this section, a novel fast MD algorithm is proposed, and the detailed algorithm is outlined in the next section.

## IV. Proposed Fast Mode Decision Algorithm

Before introducing the fast algorithm, several parameters will be defined:

- *FLAG*: indicates possible high speed motion.
- $T_0$: is mode reference table used to record the selected modes and *FLAGs* in the first frame within a group of frames (*FG*).
- $T_1$: is mode reference table used to record the selected modes and *FLAGs* in the most previously encoded frame.
- *TH*: is threshold for motion speed detection.

```
Step 1.   Initialize $T_0$ and $T_1$.
Step 2.   Encode the I-frame by using the exhaustive MD.
Step 3.   If the current macroblock (x, y) is in the first frame
          of current FG, the mode list $L_{(x, y)}$ consists of
          exhaustive candidate macroblock modes. Otherwise,
          go to Step 4.
Step 4.   If (x, y) is on the fist edge of frame
          {
            If the mode of co-located position in $T_0$ (the
            second frame of FG) or $T_1$ (the rest frames of FG
            excluding the first and second frames) is SKIP,
        $L_{(x, y)}$ consists of Skip and inter 16×16 modes.
        Otherwise, $L_{(x, y)}$ consists the exhaustive modes.
    }
    Else if FLAG at the co-located position in $T_1$ (or $T_0$ for the second frame of FG) is HIGH, $L_{(x, y)}$ consists of the exhaustive modes.
    Else
    {
        If (x, y) is on the second edge of frame, $L_{(x, y)}$ consists of the modes of the co-located and first neighbouring positions in $T_0$ and $T_1$ ($T_1$ is not available for the second frame of FG).
        Otherwise, $L_{(x, y)}$ consists of the co-located, first neighbouring and second neighbouring positions in $T_0$ and $T_1$ ($T_1$ is not available for the second frame of FG).
    }
Step 5. Encode (x, y) based on the current $L_{(x, y)}$, and update $T_0$ and $T_1$. If (x, y) is not the last macroblock in the current frame, go back to Step 3. Otherwise, go to Step 6.
Step 6. If the current frame is the last of video sequence, the encoding process is finished. Otherwise, go back to Step 3.
```

## V. Experiments and Discussion

The test experiments of the proposed fast MD algorithm were performed on a Pentium IV 2.66 GHz PC with 2GB memory. The proposed fast MD algorithm was integrated into the standard JM10.1 [13] software and the configurations are listed below:

| |
|---|
| Baseline profile |
| The search range was ±16 |
| The type of sequence was IPPP |
| Full search ME scheme was deployed |
| One frame was used as the reference for ME |
| The test sequence consisted of 300 frames |

Ten commonly recommended video sequences were used in the test experiments, and they were divided into three groups according to their characteristics: distance, backgrounds, and motion. Bridge-far and Bridge-close are in the first group, and their distances to the camera are different. The second group consists of sequences with medium motion under different backgrounds where Foreman contains medium detailed background and scene changes, fast moving background is in Carphone, and the background of Claire is homochromous and steady. In the third group, Mobile involves medium speed motion and high detailed contents. Dual-direction movement of two boats is in Coastguard. The motion speed of Highway is fast. Stefan includes fast motion and detailed background, and a barge with slow motion is in Container. In order to evaluate the proposed algorithm, three parameters (encoding time reduction rate, variation of PSNR and bitrate increase rate) were defined in (4), (5) and (6):

$$\Delta Time = \frac{Time_B - Time_A}{Time_A} \times 100\% \tag{4}$$

$$\Delta PSNR = PSNR_B - PSNR_A \tag{5}$$

$$\Delta Bitrate = \frac{Bitrate_B - Bitrate_A}{Bitrate_A} \times 100\% \tag{6}$$

where subscript $A$ indicates the result under the exhaustive MD algorithm specified in JM10.1, and subscript $B$ indicates the result under proposed fast MD algorithm. Table I - Table

TABLE I
Experimental Results (QP=28)

| Sequence | ΔTime (%) | ΔPSNR (dB) | ΔBitrate (%) |
|---|---|---|---|
| bridge-far | -80.69 | -0.04 | 1.08 |
| bridge-close | -68.77 | -0.02 | 0.85 |
| foreman | -26.53 | -0.03 | 1.10 |
| carphone | -32.26 | -0.05 | 0.50 |
| Claire | -69.06 | -0.07 | 1.17 |
| mobile | -31.25 | 0 | 0.14 |
| coastguard | -28.59 | -0.02 | 0.41 |
| highway (CIF) | -44.66 | -0.06 | 0.40 |
| Stefan (CIF) | -20.37 | -0.01 | 0.11 |
| container (CIF) | -61.01 | -0.03 | 0.47 |

TABLE II
Experimental Results (QP=32)

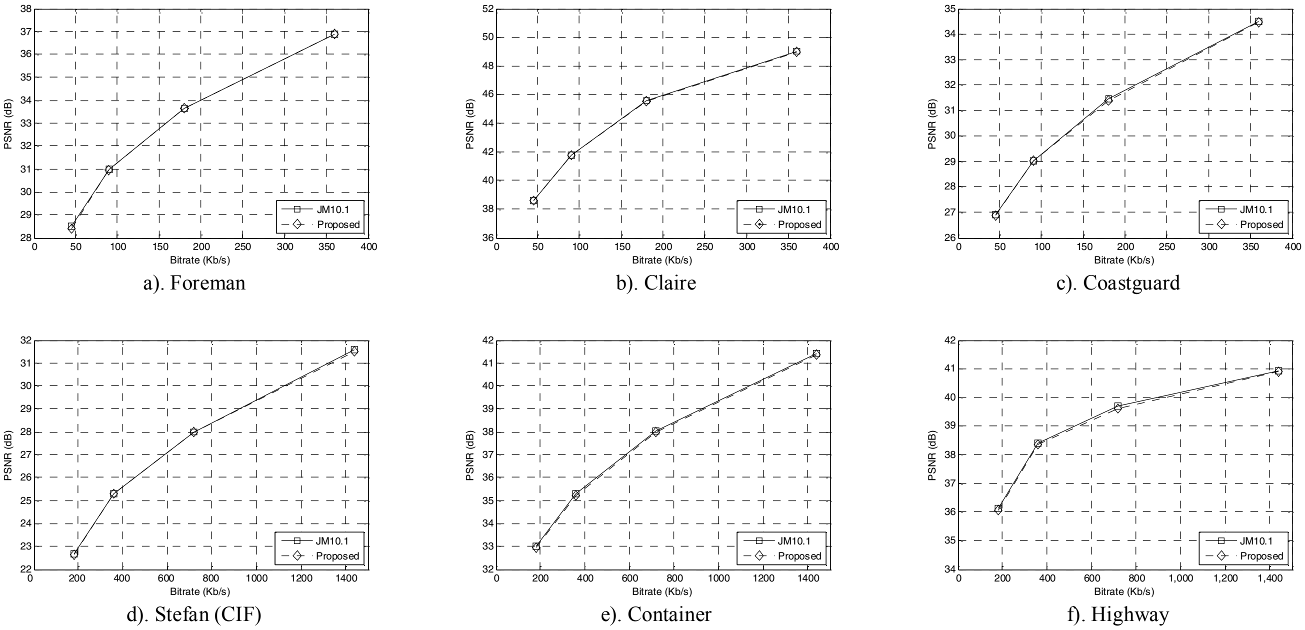
| Sequence | ΔTime (%) | ΔPSNR (dB) | ΔBitrate (%) |
|---|---|---|---|
| bridge-far | -81.02 | 0 | 0 |
| bridge-close | -76.02 | -0.04 | 1.31 |
| foreman | -29.42 | -0.04 | 2.25 |
| carphone | -37.61 | -0.09 | 0.14 |
| Claire | -66.92 | -0.16 | 1.01 |
| mobile | -30.53 | -0.02 | 0.01 |
| coastguard | -32.90 | -0.05 | 0.70 |
| highway (CIF) | -52.63 | -0.08 | 1.02 |
| Stefan (CIF) | -20.13 | -0.01 | 0.25 |
| container (CIF) | -65.32 | -0.04 | 0.74 |

TABLE III
Experimental Results (QP=36)

| Sequence | ΔTime (%) | ΔPSNR (dB) | ΔBitrate (%) |
|---|---|---|---|
| bridge-far | -79.82 | 0 | 0 |
| bridge-close | -79.10 | -0.03 | 1.76 |
| foreman | -34.35 | -0.09 | 2.34 |
| carphone | -42.95 | -0.09 | 0.67 |
| Claire | -67.73 | -0.14 | 1.44 |
| mobile | -30.98 | -0.02 | 0.50 |
| coastguard | -39.97 | -0.05 | 0.63 |
| highway (CIF) | -68.43 | -0.19 | 1.34 |
| Stefan (CIF) | -21.26 | -0.02 | 0.24 |
| container (CIF) | -64.79 | -0.04 | 0.75 |

TABLE IV
Experimental Results (QP=40)

| Sequence | ΔTime (%) | ΔPSNR (dB) | ΔBitrate (%) |
|---|---|---|---|
| bridge-far | -80.38 | 0 | 0 |
| bridge-close | -78.79 | -0.03 | 1.08 |
| foreman | -45.43 | -0.10 | 0.90 |
| carphone | -53.07 | -0.15 | 0.18 |
| Claire | -67.05 | -0.15 | 0.80 |
| mobile | -38.81 | -0.08 | -0.51 |
| coastguard | -47.53 | -0.07 | -1.70 |
| highway (CIF) | -65.77 | -0.21 | -0.07 |
| Stefan (CIF) | -24.55 | -0.03 | 0.29 |
| container (CIF) | -65.32 | -0.05 | 0.24 |



a). Foreman

b). Claire

c). Coastguard

d). Stefan (CIF)

e). Container

f). Highway

Fig. 6. RD curves comparison between the exhaustive MD and the proposed fast MD algorithm.

TABLE V
COMPARING WITH GRECOS'S ALGORITHM (QP=28)

| Sequence | Fast MD | ΔTime (%) | ΔPSNR (dB) | ΔBitrate (%) |
|---|---|---|---|---|
| bridge-far | [6] | -72.82 | -0.05 | 0 |
| | proposed | -80.69 | 0 | 1.08 |
| foreman | [6] | -12.50 | 0 | 0.67 |
| | proposed | -26.53 | -0.03 | 1.10 |
| carphone | [6] | -13.21 | -0.13 | -0.11 |
| | proposed | -32.26 | -0.05 | 0.50 |
| coastguard | [6] | -11.70 | -0.01 | -0.35 |
| | proposed | -28.59 | -0.02 | 0.41 |
| Stefan | [6] | -10.69 | 0 | 0.30 |
| | proposed | -20.37 | -0.03 | 0.47 |

IV list the experimental results when different QP values were deployed. According to the results, the encoding time can be saved 51.29% on average. In most cases, the decrease of PSNR is under 0.1dB with less than 2% increase of bitrate. Especially for some test sequences with low-speed motion and steady background, the encoding time is reduced up to 80%, and the decrease of encoding performance is negligible. Fig. 6 illustrates the rate distortion (RD) curves for several test sequences, which illustrate that the proposed fast MD algorithm performs close to the exhaustive MD algorithm. In contrast with Grecos's algorithm [6], which is under the same experimental configurations, the proposed algorithm can gives an impressive performance, and the comparison results are in Table V.

## VI. CONCLUSIONS

This paper presents a novel fast MD algorithm by exploiting the correlation between macroblocks. Combined with the adjustment based on the motion speed, the proposed algorithm can achieve a reduction of 50% total encoding time on average, and the degradation of encoding efficiency is negligible. The new algorithm does not involve complex calculations, and it is well suited for real-time video communications applications.

## REFERENCES

[1] ITU-T Recommendation H.261, CDM XV-R-37-E, *Video Codec for Audiovisual Services at p×64 kbits/s*, Dec. 1990.

[2] ITU-T Recommendation H.263, SG15, *Video Coding for Narrow Telecommunication channels at < 64kbits/s*, Apr. 1995.

[3] "Draft ITU-T Rec. and FDIS of Joint Video Spec. (H.264 | ISO/IEC 14496-10 AVC)," JVT of MPEG and VCEG, Doc. JVT-G050rl, May 2003.

[4] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression, Video Coding for Next-generation Multimedia.* John Wiley & Sons, England, 2003.

[5] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, Jul. 2003.

[6] C. Grecos, and M. Y. Yang, "Fast inter mode prediction for P slices in the H.264 video coding standard," *IEEE Trans. Broadcasting*, vol. 51, pp. 256-263, Jun. 2005.

[7] F. Pan, Xi Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, pp. 813-822, Jul. 2005.

[8] 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 and Systems for Video Technology*, vol. 15, pp. 953-958, Jul. 2005.

[9] J. F. Wang, J. C. Wang, J. T. Chen, A. C. Tsai, and A. Paul, "A novel fast algorithm for intra mode decision in H.264/AVC encoders," in *Proc. IEEE ISCAS*, 2006, pp. 3498-3501.

[10] B. Zhan, B. Hou, and R. Sotudeh, "Fast mode decision for Inter-prediction in H.264/AVC," accepted by *IEEE ISCIT*, 2007.

[11] B. Zhan, B. Hou, and R. Sotudeh, "Statistical data based low-complexity mode decision for P-frame encoding in H.264/AVC," accepted by *IEEE ICICS*, 2007.

[12] Y. F. Ling, Z. H. He, and I. Ahmad, "Analysis and design of power constrained video encoder," in *Proc. IEEE 6$^{th}$ CAS Symp. on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, 2004, pp. 57-60.

[13] JVT reference software unofficial version JM10.1, http://iphome.hhi.de/suehring/tml/download/jm10.1.zip.