Waseda University Doctoral Dissertation

Motion Correlation Based Low Complexity and Low Power Schemes for Video Codec

Chen LIU

Graduate School of Information, Production and Systems

Waseda University

November 2012

Abstract

Among various kinds of multimedia applications, video is the most expressive and attractive one. Owing to the fast development of multimedia communication technology, video is more and more widely used in the modern society. With the increasing demand for more convenient and higher quality video experience, more efficient video coding schemes are gathering more attention. The popularity of video applications on mobile devices makes the low complexity and low power design for video codec becoming an important research topic. To optimize the video codec system, a number of algorithms have been proposed for both encoder and decoder respectively.

This dissertation presents my research on the low complexity and low power schemes for H.264/AVC encoder and decoder based on the motion correlation. The motion correlation information, such as motion vectors (MVs) and residual which are generated during the video compression process are skillfully utilized in the proposed schemes. The motion correlation information provides more underlying spatial and temporal relation of the video content which is helpful for the motion feature analysis and optimization for the coding process.

In video encoder, the Motion Estimation (ME) costs about 70% computational amount, thus to reduce the amount of computation of the ME is an important issue. In ME, there are 8 inter modes for inter frame prediction. To reduce the computational complexity by proper mode selection while keeping the visual quality and compression efficiency at the same time becomes a research topic (fast inter

mode decision). Different from previous inter mode decision algorithms (e.g. Yu's ICASSP 2004) which mainly focus on the feature analysis of the frame contents in original video, the motion correlation information such as motion vectors and residual are utilized in proposed algorithms. The proposed motion adaptation based algorithm is introduced in Chapter 2, and the proposed residual feature based algorithm is introduced in Chapter 3.

For video decoder, to reduce the decoding time, a novel partial video decoding scheme is proposed. Usually, when the High-Definition (HD) video is played on a relatively lower resolution screen as on the smartphone, the video contents in the frame have to be fully decoded and then down-sampled to adapt the screen resolution of the device. In Chapter 4, an Object of Interest (OOI) oriented partial decoding is proposed. The OOI is defined by the user, and only the OOI related region is decoded and displayed on the device screen. In that way, the OOI region is able to keep the visual quality as the original video. To achieve partial decoding and keep the visual quality of the Decoded Partial Area (DPA), the motion correlation information such as motion vectors and residual are utilized. The proposed partial decoding is also able to decrease the decoding time.

In order to check the real effectiveness on low power of the proposed method, a new workload prediction algorithm is developed for the partial decoding scheme proposed in Chapter 4, and works together with Dynamic Voltage Frequency Scaling (DVFS) on an embedded platform to achieve low power consumption and energy cost. This envelop detection based workload prediction method is proposed in Chapter 5.

This dissertation consists of six chapters as follows:

Chapter 1 [Introduction] introduces the H.264/AVC video coding standard. The motion correlation is discussed, which is the key of this dissertation. The contributions of this dissertation are also summarized.

Chapter 2 [Motion Adaptation Based Inter Mode Decision] proposes a fast inter mode decision algorithm for H.264/AVC encoding based on the motion vector of current macroblock. The dynamically updated motion vector (MV) which is obtained from the 16×16 motion search and the absolute value of MV are used to select the modes. According to the absolute value of MV, the mode which should be chosen is judged among the 8 modes. Firstly, from the 5 modes in the first level, the candidate modes are selected according to the appropriate thresholds. And then, for the 8×8 block modes in the second level, all the 4 modes (8×8, 8×4, 4×8, 4×4) are conducted. The performance of encoding time saving is able to be increased by combining Choi's SKIP-early strategy (IEEE Trans. CSVT 2006). Comparing to the H.264 reference software JM14.1, the proposed algorithm achieves an average of 33.4% time reduction with minimal quality loss and less bit-rate increase. Moreover, the proposed algorithm reduces 19.1% encoding time comparing to the JM with Choi's SKIP-early strategy.

Chapter 3 [Residual Feature Based Fast Inter Mode Decision] proposes a fast inter mode decision algorithm based on the residual of macroblocks. The proposed algorithm optimizes for both two levels of inter mode decision. The residual is obtained after the motion search of 16×16 mode or 8×8 mode. In the previous residual based Wang's method (ICME 2007), all the residual is summed up together for selecting modes. In the proposed algorithm, to select the modes, the positive and negative residual are calculated separately. Such that, the motion

analysis is more detailed, which results in more accurate mode selection. According to the calculation of the complexity of the block (16×16, 8×8) and the similarity between two blocks (8×8, 4×4), the most appropriate mode for current block is conducted. Experiments show that, comparing to JM14.1, 60%~72% mode-search amount is reduced. As a result, the proposed algorithm achieves an average of 56.6% on time-saving comparing to JM. Comparing to Wang's method, the time-saving is 21.7%.

Chapter 4 [Encoder Unconstrained User Interactive Partial Decoding Scheme] presents an Object of Interest (OOI) oriented partial decoding scheme at only decoder side. There is no required processing at the encoder side. The OOI which is defined by the user at decoder side is tracked, and only the OOI related partial area is decoded. The OOI tracing is conducted for every P frame, and the Decoded Partial Area (DPA) is calculated and updated after decoding each I frame. In these processes, all motion vectors are used while not all the residual information, and only the residual of the macroblocks that belong to the DPA are used. Such that, only compressing domain information is used for the OOI tracking and DPA adaptation. The motion vectors are recorded to serve the OOI tracking and DPA adaptation, the residual information is selectively used for data decoding. During the partial decoding, when the reference block or pixel belong to the undecoded area (outside the DPA), for inter block, the closest block in DPA is used as the candidate reference block (named direction preferred reference block relocation (RBR) in this dissertation); for intra block, the certain pixels in the collocated block of previous frame is used as the candidate reference pixels (named co-located temporal Intra prediction (CTIP) in this dissertation). The proposed partial decoding scheme

provides an average of 50.2% decoding time reduction with 29% DPA ratio comparing to the fully decoding process, and the PSNR (Peak Signal to Noise Ratio) drop of the displayed region is only 0.09dB. This proposal is especially useful for displaying HD video on portable devices where the battery life is a crucial factor. Also, the encoder-unconstrained solution is a necessary condition of real-time broadcasting.

Chapter 5 [Envelope Detection Based Workload Prediction] presents a new workload prediction algorithm for the partial decoding scheme proposed in Chapter 4, and works together with Dynamic Voltage Frequency Scaling (DVFS) on an embedded platform to achieve low power consumption and energy cost. For workload estimation, the previous Jin's Hilbert transform based method (IEEE Trans. CAD 2012) has the drawback of relatively large computational cost and still noticeable deadline missing rate. The proposed algorithm predicts future workloads by detecting the envelope of the difference value between the previous adjacent actual-workloads of P frames. To further improve the accuracy of the envelope detection, the negative truncation is introduced by ignoring the negative difference in calculation. The deadline missing rate of Jin's method is 4.61% while the proposed one is only 0.66%. This proposal is implemented on the RP2 board and works together with DVFS. Comparing to the method without workload prediction and DVFS, the power reduction achieves about 41.7% and the energy saving is about 10.2%. Comparing to fully decoding, the energy saving is about 61.1%.

Chapter 6 [Conclusion] concludes the whole dissertation.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my dear Professor Satoshi Goto, for his constant guidance, encouragement, and support during my research. Without his guidance, help and suggestions, I cannot finish my studies smoothly.

I would also like to express my deep gratitude to my second supervisor Professor Jiro Katto, Professor Takeshi Yoshimura, and Professor Masahiro Okuda for the guidance and kind advices throughout my research.

I am also very grateful to Dr. Xin Jin, Dr. Tianruo Zhang, and Mr. Minghui Wang for working together with me in the same research group. The discussions with you always helped me a lot and inspired me. I also thank all the members of Professor Goto's laboratory for their cooperation and help.

I also express my sincere gratitude to Waseda University Ambient SoC Global COE Program of MEXT, for the support to my research. I also thank the support from the CREST project of Japan Science and Technology Agency.

Finally, I would extend my deepest thanks to my family and all my friends for their kind support over the years. Thanks to my parents and my husband Dr. Mengshu Huang, for the love, support and understanding. Thanks to my dear friend Dr. Chuan Yue, for the precious friendship during all these years.

Contents

List of Tables

List of Figures

1 Introduction

Nowadays, multimedia plays a more important role in current societies, especially at some aspects such as the internet, broad-casting, telecommunication, mobile devices, and so on. Many hot topics come out with peoples' dream of the future life. People desire friendlier human-computer interaction (HCI) interface, more vivid present approaches, and more rapid reaction speed.

Video is a very important part of multimedia systems. To fit multiform requirements of different multimedia platforms, there are amount of forms to serve the specific applications. Many video applications become more and more popular, such as VOD (Video on Demand) which can be used to watch movies on the internet, and Pay-TV which can be used for television set-top box and video conferences. Recently, with the popularity of mobile devices with embedded video cameras, the video streaming and video encoding on hand-held devices becomes increasingly popular. However, the huge data size of videos became a heavy burden for both video compression/processing and data communication. At the same time, nearly every user expects better video quality, low-power, and least cost. Thus, an efficient video codec with low computational complexity and power consumption is a crucial problem, and to achieve the efficiency and low-cost with proper qualities is really a challenge work. Many researchers have been attracted to the video compression topic, and dedicate themselves to this work.

1.1 Introduction to H.264/AVC Video Coding Standard

1.1.1 Outline of Video Codec System

Video codec system is appearing everywhere in people's daily life. Figure 1-1 shows the components of a common video codec system. Firstly, we get the video sequences by video camera or other device, and then the uncompressed data is transferred to the video encoder. At the encoder, the data is compressed to the bit-stream according to specific coding standard. Transmit channel is used to transfer the coded bit-stream.

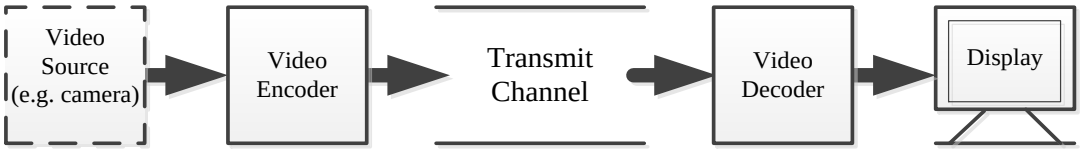


Figure 1-1 Video coding system

Through the transmit channel, coded video stream is sent from video encoder to decoder side. Video decoder receives the bit-stream, and decodes it by the coding standard that the same to the encoding one. Finally, the decoded video sequence is sent to the display, and the users can enjoy the video from the display.

Figure 1-2 shows a conventional architecture of video encoder. The encoder can be divided to three main functional units: the temporal model, the spatial model and the entropy encoder. In the temporal model, it works for reducing the temporal redundancy of current processed frame/macroblock (MB) by exploiting the similarities between neighboring video frames and/or neighboring macroblocks (MBs). There are two predictors in the temporal model, names Intra Predictor and

Inter Predictor respectively. Temporal model provides the information such as decided block modes, block types, motion vectors, residual data, etc. to the spatial model. In spatial model, residual data is processed by DCT transformation and quantization to get the quantized coefficients. It reduces the spatial redundancy.

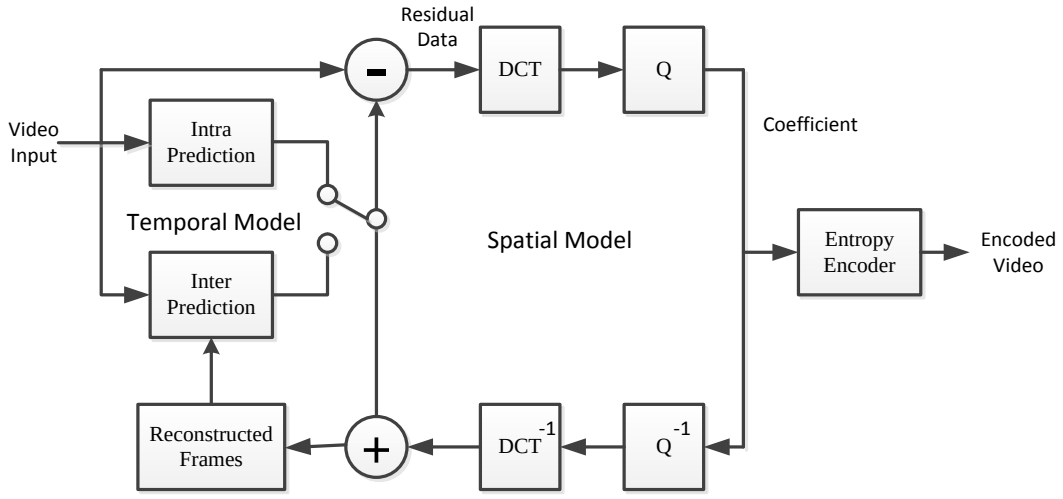


Figure 1-2 Video encoder

At the entropy encoder part, all of the data is compressed by entropy encoder to get the encoded video stream. Usually, the coded video stream composes by certain syntax, contains syntax header, parameters, information of frame types and block modes, coded motion vectors and coefficients.

Video decoder is almost the inversion of video encoder as shown in Figure 1-3. Entropy decoder performs the syntax analysis, extracts the prediction parameters, and the coefficients. The motion compensation (MC) module reconstructs the prediction data by the prediction parameters and obtained reference frames. The coefficients are processed by inverse quantization and inverse DCT to get the

residual data. Then the prediction data is combined with the residual to obtain the reconstructed frame. As a result, we can see the decoded video sequences.

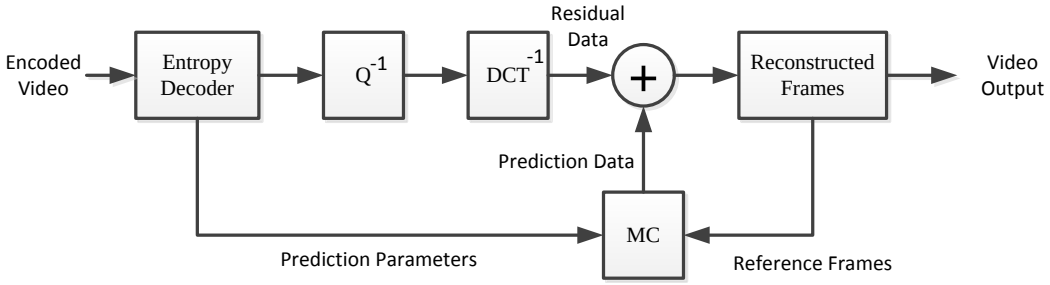


Figure 1-3 Video decoder

Compare to video encoder, video decoder is simpler from the view of architecture. It does not contain the temporal model. The time consumption and power cost of video decoder is also much lower than the encoder.

1.1.2 Overview of H.264/AVC

H.264/MPEG-4 Part 10 [1] is the latest video coding standard proposed by Joint Video Team (JVT) and together developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). Compare to previous video coding standards such as MPEG-1 [2] , MPEG-2 [3], MPEG-4 [4], H.263 [5], H.264/AVC has achieved significant improvements in compression performance, especially in the rate-distortion efficiency aspect. It is because of many new technical features which have been introduced in H.264/AVC, such as variable block-size motion compensation with small block sizes, multiple reference picture

motion compensation, the in-the-loop de-blocking filtering, context-adaptive binary arithmetic coding (CABAC), and so on [6].

H.264/AVC is designed for serving widespread applications. The four features named baseline profile, main profile, extended profile and high profile [7] can serve different applications from low-end to high-end. Figure 1-4 shows the relationship of the four profiles [8].

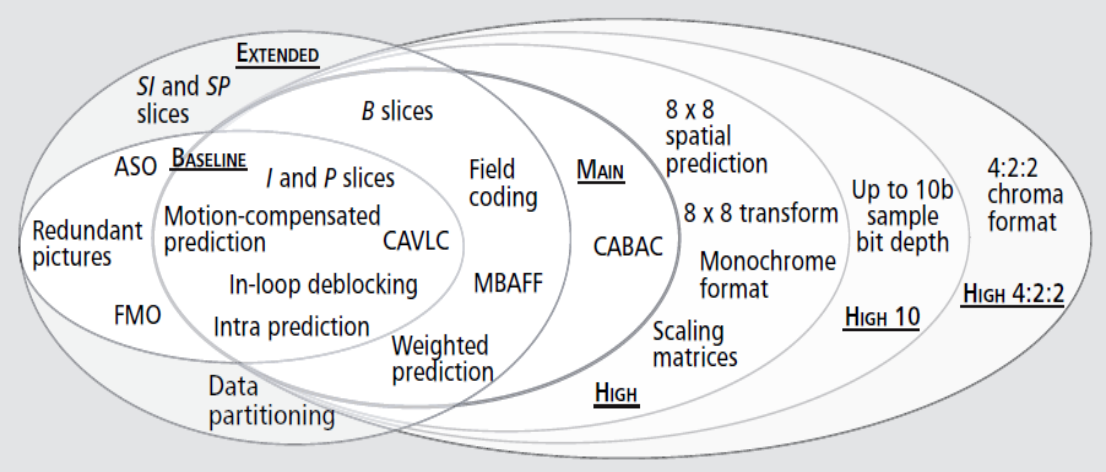


Figure 1-4 Illustration of H.264/MPEG4-AVC profiles

The baseline profile with the lowest computational complexity is the most widely used profile, it only support the very basic functions and parameters such as IME, FME, FMO, MRF, I and P slices, CAVLC, etc.. Main profile is set as the best tradeoff between the computational complexity and compression ratio, it further includes the B slices, weighted prediction, field coding, MB-AFF, CABAC, etc. High profile is added Fidelity Range Extensions (FRExt) in 2004 [7]. High profile provides the best compression ratio.

1.1.2.1 H.264/AVC bit stream

Figure 1-5 shows the nesting hierarchy of H.264/AVC bit stream [1] [9]. The top level contains the NAL syntax elements and NAL units. A NAL unit is made up of the NAL unit header and RBSP unit. The NAL unit header contains the NAL unit type. One type of NAL unit consists of slice header followed by the slice data. The slice data carries compressed information for macroblocks An macroblock is the basic coding unit which contains 16x16 luminance block and associated chrominance bocks depending on the chroma-format (e.g. the size of chrominance blocks is 8x8 for 4:2:0). Each compressed macroblock is represented by the macroblock overhead information (e.g. MB type, Mb prediction modes, motion vectors) followed by the residual information coded with CAVLC or CABAC.

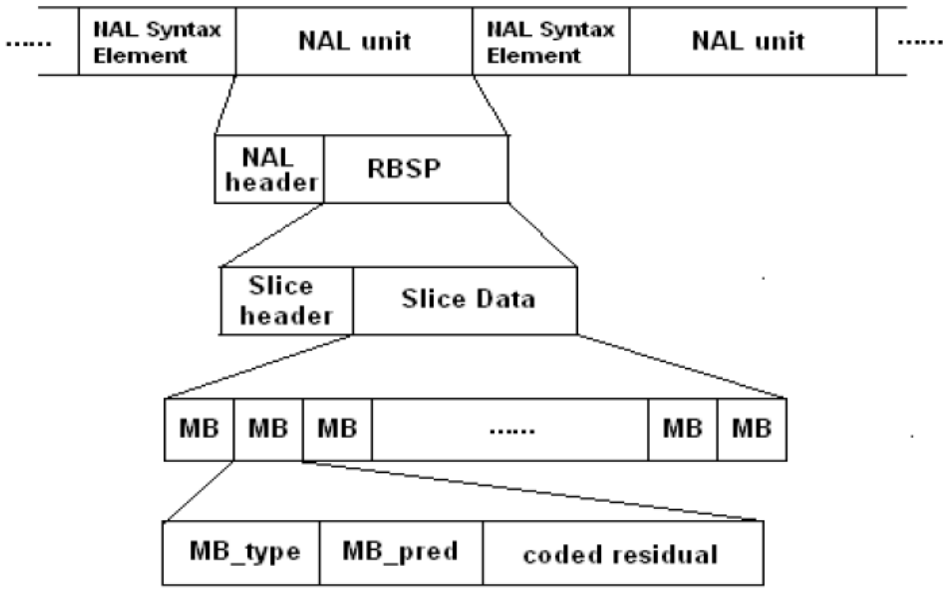


Figure 1-5 H.264/AVC bit stream structure

1.1.2.2 Inter Prediction

H.264/AVC uses block-based motion compensation. Every coded macroblock is predicted from previously-encoded data. Samples in one inter macroblock (MB) are predicted from previously-encoded pictures, and samples in one intra macroblock are predicted from samples which have already been encoded, decoded and reconstructed in current picture [9].

In H.264/AVC, to evaluate various coding modes, the Lagrange optimization technique based rate-distortion (R-D) cost is calculated for each valid mode [10]. After the motion search and rate-distortion cost calculation for each valid mode, the mode which has the minimum R-D-cost is determined as the final mode.

The inter prediction model is created from multiple previously encoded video frames.

For luminance component, the block partition modes which are defined for inter macroblocks in H.264/AVC present a 2-level tree structure as shown in Figure 1-6. The first level has five different modes: SKIP mode, P16 × 16 mode, P16 × 8 mode, P8 × 16 mode and P8 × 8 mode. In the second level, for each 8 × 8 sub-macroblock of P8 × 8 mode, it has four kinds of sub-partition modes, 8 × 8 sub-macroblock mode, 8 × 4 sub-macroblock mode, 4 × 8 sub-macroblock mode and 4 × 4 sub-macroblock mode. These sub-macroblock modes compose the second level in the inter mode hierarchy.

For each chroma component in a macroblock, it has half resolution in both horizontal and vertical of the luminance component. And the block partition of chroma block is in the similar structure as the luminance component with half size in resolution comparing to the luminance component.

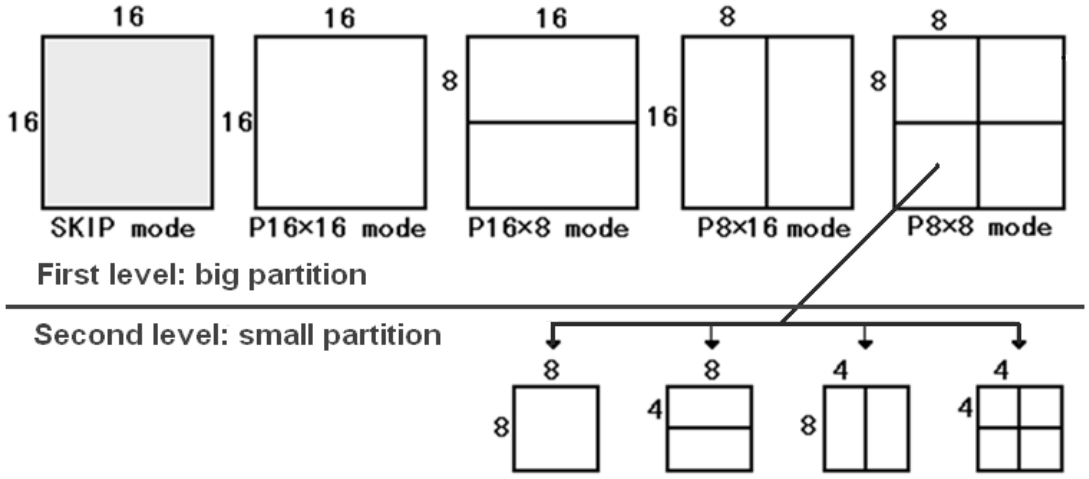


Figure 1-6 Inter mode partitions in H.264/AVC

1.1.2.3 Intra Prediction

The intra predicted MB is formed based on previously encoded and reconstructed blocks. Different from the inter prediction, in intra prediction, the luminance and chroma samples are predicted respectively [9].

For luminance samples, the macroblock is predicted by each 4x4 block or the 16x16 macroblock. For 4x4 block, there are 9 different prediction modes: vertical, horizontal, DC, diagonal down-left, diagonal down-right, vertical-right, horizontal-down, vertical-left, and horizontal-up, see Figure 1-7. For 16x16 macroblock, there are 4 modes: vertical, horizontal, DC and plane, see Figure 1-8.

For chroma samples, there are 4 prediction modes: DC (mode 0), horizontal (mode 1), vertical (mode 2) and plane (mode 3). They are very similar to the 16x16 luminance prediction modes described above, except the mode numbers are different.

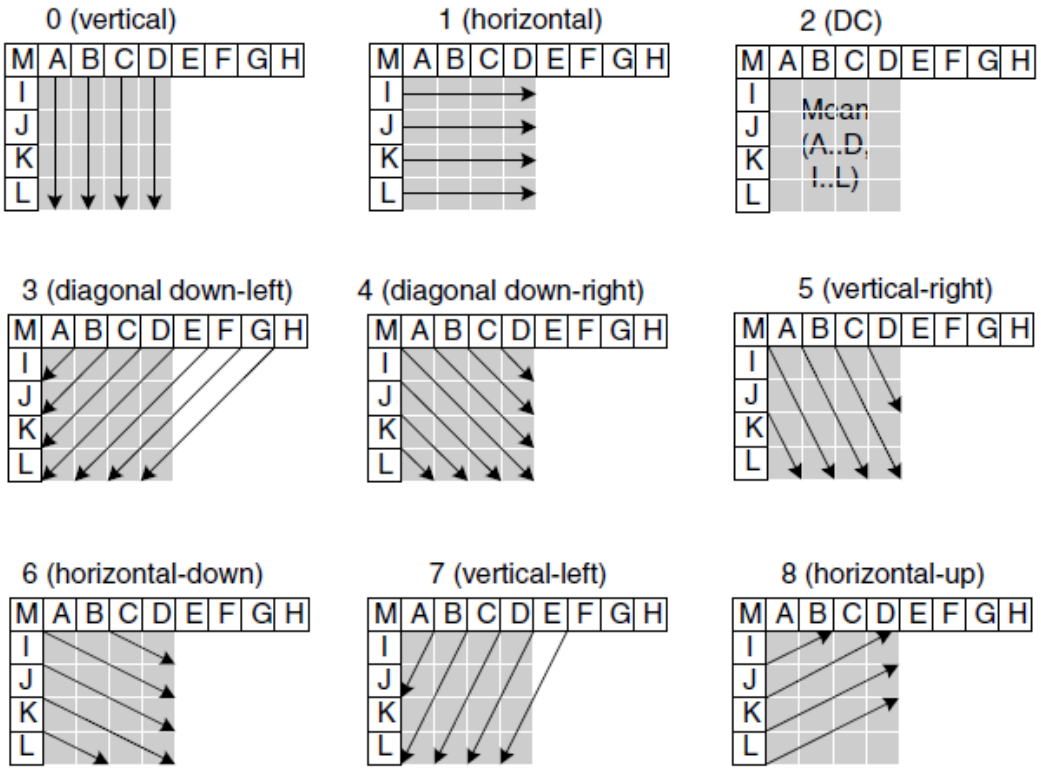


Figure 1-7 Intra 4x4 luma prediction modes

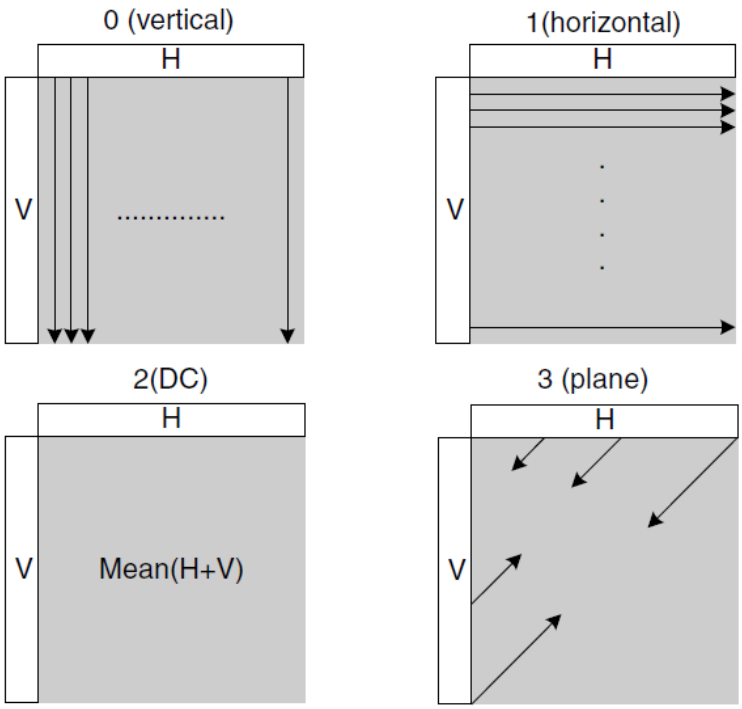


Figure 1-8 Intra 16x16 luma prediction modes

1.2 Motion Correlation

Video compression uses the temporal correlation between the macroblocks in consequent frames and spatial correlation between the macroblocks in one frame to reduce the redundancy of the video data. Amongst all the modules in video encoder, the most important and also the most time consumed part is the motion estimation (ME) module. In order to improve the prediction accurate and the prediction efficiency of the motion estimation, a number of works has been done.

Some information, such as the motion vectors and residual, which generated during the video encoding process, contains many motion correlations between the reference block and the current block, and also represents some relation between the previous frame and current frame.

In this dissertation, we are trying to analysis the motion correlation by reuse this motion information (motion vectors, residual, etc.) generated in the motion estimation process itself. On one hand, this kind of motion information is obtained by the exhausted motion estimation, thus it contains the underlying motion correlation between the reference blocks and current block. On the other hand, reuse this information is not only helpful for investigating the motion correlation of the video content itself, but also avoid extra feature extraction and analysis of the video content. Thus, the analysis based on the motion information generated by the motion estimation is an effective and efficient way to study the motion correlation in the video data.

In the following of this dissertation, we can see that, the analysis based on both motion vectors and residual, for both the encoder and decoder side optimization, the

motion correlation based analysis plays a very important and effective role. The simulation results also proved the statement above.

1.3 Summary of Contributions

This dissertation presents my research on the low complexity schemes for both H.264/AVC encoder and decoder. The motion correlation information, such as motion vectors and residual which are generated during the video compression process, is skillfully utilized in all the proposed schemes. These approaches avoid the extra fetching and analysis process of other features, and the motion correlation information also provides more underlying spatial and temporal relation of the video content which is very necessary and helpful for the motion feature analysis and optimization the coding process.

For video encoder, two kinds of fast inter mode decision algorithms are proposed to reduce the computational complexity of motion estimation (ME) by proper inter mode selection. Different from previous inter mode decision algorithms which mainly focus on the feature analysis of the frame contents in original video, the motion correlation information such as motion vectors and residual are utilized in proposed algorithms. The proposed motion adaptation based method select the mode according to the motion feature of the current motion vector (MV) within an adaptive updated MV set. Thus it works well for the sequences with different motion characteristics. The time reduction of proposed motion adaptation based inter mode decision is about 33.4%. The residual analysis based method select the mode by analyzing the similarity and complexity of the residual feature of current macroblock. Different from previous works in which the positive and negative

residual in a block is summed indiscriminately, in the proposed algorithm the positive and negative residual are extracted separately and both of the extracted results are used for evaluation. Thus the proposed algorithm is able to predict the mode more precisely, which results in a better encoding time reduction than previous works. The proposed residual feature based fast inter mode decision algorithms are able to reduce the encoding time by 56.85%.

For video decoder, a novel encoder-unconstrained user interactive partial video decoding scheme is proposed. It is the first work on enabling OOI oriented partial video decoding without encoder-side support. The proposed scheme is used to solve the resolution mismatch problem when playing high resolution videos on relatively low resolution screens, and it is designed in a user-friendly and low cost way. The moving foreground and background, as well as the reference inaccessible problem are the challenges. All motion vectors are decoded and recorded for OOI tracking and Decoded Partial Area (DPA) adaptation, and residual is selectively decoded according to the DPA. The simulation results show that, the proposed partial decoding scheme provides an average of 50.16% decoding time reduction comparing to the fully decoding process. The displayed region also presents the original HD granularity. The proposed partial decoding scheme is especially useful for displaying HD video on the devices of which the battery life is a crucial factor. To further investigate the power reduction of the partial decoding scheme, an envelope based workload prediction is proposed and work together with the Dynamic Voltage Frequency Scaling (DVFS) to reduce the power consumption of the proposed partial decoding scheme. The experimental results on the evaluation board show that, with the proposed envelope detection based workload prediction

and DVFS, the power reduction achieves about 41.65%, the energy reduction is about 10.16%. The energy reduction of the partial decoding together with proposed workload prediction and DVFS is about 61.15% compare to the original fully decoding. By further investigating and improving the DVFS, it is possible to reduce more power consumption and energy cost. The proposed envelope detection based workload prediction algorithm also can be used for general workload prediction.

1.4 Dissertation Organization

The rest of the dissertation is organized as following. Chapter 2 and 3 presents the proposed two kinds of fast inter mode decision algorithms for the encoder based on motion correlation and residual analysis respectively. Chapter 4 describes the proposed encoder-unconstrained user interactive partial decoding scheme. Chapter 5 introduces the proposed envelop detection based workload prediction. Chapter 6 concludes the dissertation.

2 Motion Adaptation Based Inter Mode Decision

This chapter presents a motion adaptation based inter mode decision method (MAIMD) for the H.264/AVC encoder. The rate-distortion optimization (RDO) module in H.264/AVC works for each mode evaluation process. Although this improving the video quality in video compression largely; at the same time, its high computational complexity brings huge time cost. Different from previous inter mode decision algorithms which mainly focus on the feature analysis of the frame contents in original video, the motion correlation information such as motion vectors is utilized in proposed algorithm. The key idea of this chapter is to reduce the candidate modes by the underlying relation between motions and macroblock partition types. Thus, a motion vector related evaluation approach with adaptive multilevel thresholds has been proposed together with SKIP mode early determination strategy. The simulation results show that the proposed method achieves 33.4% time reduction on average with only 0.06dB PSNR loss under the same bit rate, or 1.45% bit rate increase under the same PSNR. For the sequences which have more motions, the time saving of the proposed method gains about 20% than only use the SKIP mode early determination.

2.1 Background

The Motion Estimation (ME) is the most time consuming part of video encoder. H.264/AVC introduces the variable block size motion estimation (VBSME). For

each inter mode, the motion search of inter prediction is conducted for each sub-block in the mode, and the Rate Distortion (RD) cost is calculated. Finally, the mode with the lowest RD cost is set as the final/best mode.

The inter prediction model is created from multiple previously encoded video frames. The block partition modes which are defined for inter macroblocks in H.264/AVC present a tree structure. The first level has five different modes: SKIP mode, P16 × 16 mode, P16 × 8 mode, P8 × 16 mode and P8 × 8 mode; for each 8 × 8 sub-macroblock of P8 × 8 mode, it has four kinds of sub-partition modes, 8 × 8 sub-macroblock mode, 8 × 4 sub-macroblock mode, 4 × 8 sub-macroblock mode and 4 × 4 sub-macroblock mode. These sub-macroblock modes compose the second level in the inter mode hierarchy.

In H.264/AVC, to evaluate various coding modes, the Lagrange optimization technique [11] based rate-distortion (R-D) cost is calculated for each valid mode, and the mode which has the minimum R-D-cost is finally chosen [12]. The Lagrange function for calculate the R-D-cost is defined as:

$$J = D + \lambda \times R$$

Equation 2-1

where J is the cost, D is the distortion of current mode, R is the encoding rate, and λ is Lagrange multiplier which is given by:

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

Equation 2-2

where QP is the quantization parameter of macroblocks.

In order to get the rate R, first the difference between the original macroblock and the predicted one is calculated first, then, implements the 4 × 4 Hadamard transform and quantization. To obtain the distortion D, after the inverse quantization and IFFT transform，the reconstructed macro-block is compared with the original one. Thus, the running time of mode decision module is extremely large with R-D optimization turning on.

A number of fast inter-mode decision algorithms have been proposed these years [13] [14] [15] [16] [17] [18]. They are aimed at reducing the computational complexity and saving the encoding time. These algorithms can be generally categorized into three strategies. The first one is early detection of the SKIP mode. In all the modes defined in H.264/AVC, the encoded bits and computational complexity for SKIP mode MB is the simplest. All the residual, MVD and reference index of the SKIP mode macroblock are zero. Detecting SKIP mode at an early stage and judging some SKIP-like mode as SKIP mode can save the computational time [14]. However, the speed-up of this strategy is limited, and strongly dependent on the contents of video sequences. So it is typically combined with other strategies to achieve higher time-saving rate [15] [16]. Another drawback is that the PSNR loss becomes large when many miss-judgments occur. The second one is macroblock classification. Some of the inter mode decision algorithms categorize all the modes into several classes. However, the speed-up rate is not high when many modes are defined in the same class. The third one is features based evaluation [13] [15] [17]. Many inter mode decision algorithm use certain features as the criteria of mode evaluation. Choosing the effective features and performing reasonable analysis are the most important in these algorithms. Kim [15] uses the temporal

correlation between the related macroblocks to aid the current mode decision. Yu [17] uses the texture of original macroblock to serve the complexity measurement. Wu [13] evaluates the homogeneity and stationarity of regions. X. Wang [16] and Z. Wang [18] use the residual texture to evaluate current mode. For these kinds of algorithms, choosing the proper features and defining the right criterions are very important as they are directly related to the performance and efficiency of algorithms. During the inter mode decision process, overleaping some hopeless modes using some information to reduce the motion estimation is a feasible approach, such as using intrinsic complexity of the macroblock and the spatial homogeneity of the macroblock [13]. The proposed method is motion vector based motion correlation adaption, which considers both the temporal and spacial correlation, and is effectual to the inter mode decision.

2.2 Motion Analysis Based on Motion Vectors

2.2.1 Motion Vector Distribution in Different Partition Modes

There are some relation between the motion of video content and the partition mode of collocated macroblocks. To observe the relationship, I have surveyed different kinds of sequences, and used some evaluation method to get the statistic results and analysis.

Let’s take the sequence Forman and Tempete as an example. For Figure 2-1, the man is opening his mouth and raising his head, the camera is a little shaking. So we can see the macroblocks on the edges of his cap and the macroblocks on his mouth and nose, these macroblocks are apt to be decided as P8x8 mode.

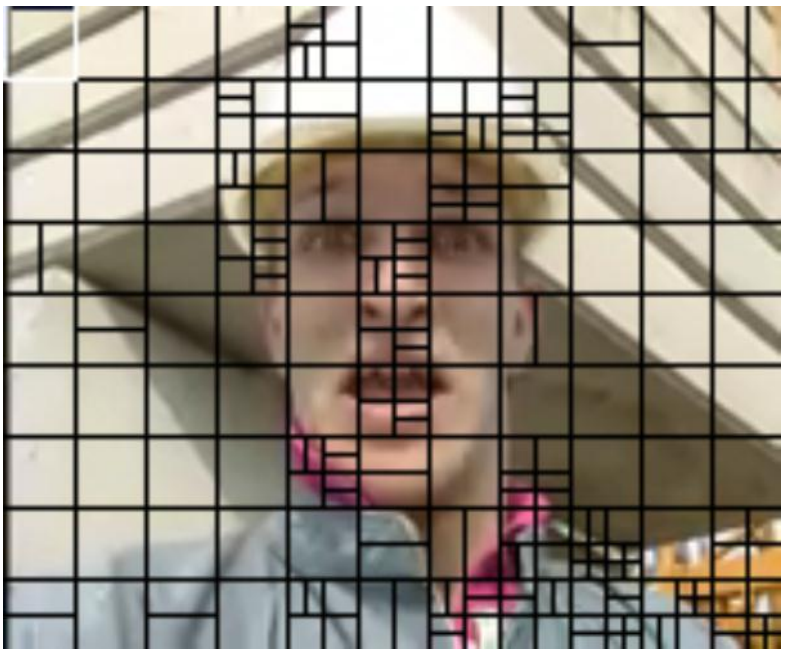


Figure 2-1 Partition mode for macroblocks in the 63rd frame of Foreman (QCIF)

Figure 2-2 Partition mode for macroblocks in the 13th frame of the sequence Tempete (QCIF)

For Figure 2-2, the wind is blowing and the leaves are waving. The lens is drawing away at the same time. From the macroblock partition we can see that the macroblocks which have more motions are apt to use P8 × 8 mode, especially at the macroblocks which were occupied by the yellow flowers area and green leaves area.

In all of the first level partition modes, P16 × 16 mode is necessary, P8 × 8 mode is the most time-consuming one for its sub-partition. So the comparisons between the motion vector distribution and the correlated block-partition decision are given between {P16 × 8 mode, P8 × 16 mode} and {P8 × 8 mode}. To evaluate different motion vectors, the arithmetic module has been used. I defined $MV_{evaluation}$ values which is calculated as the arithmetic module of motion vector Equation 2-3

$$MV_{eveluation} = \mid MV_{\mathrm{P16x16}} \mid = \sqrt{x^2 + y^2}$$

Equation 2-3

A lot of sequences have been test by the $MV_{evaluation}$ values. Figure 2-3 and Figure 2-4 show the motion vector values distribution of {P16 × 8 mode, P8 × 16 mode} in blue and {P8 × 8 mode} in red for the sequences Foreman(CIF) and News (CIF) respectively. X-coordinate represents the amount of macroblocks which belong to the partition mode, and their order of appearance; Y-coordinate denotes the $MV_{evaluation}$ values which is defined by Equation 2-3

From Figure 2-3 and Figure 2-4, it is known that the ratio and amount of a certain partition mode are different in different sequences, and much related to the video contents. It is means for difference sequences, the patterns of $MV_{evaluation}$ are changing along with different sequence contents. Thus, an adaptive approach is necessary in this case.

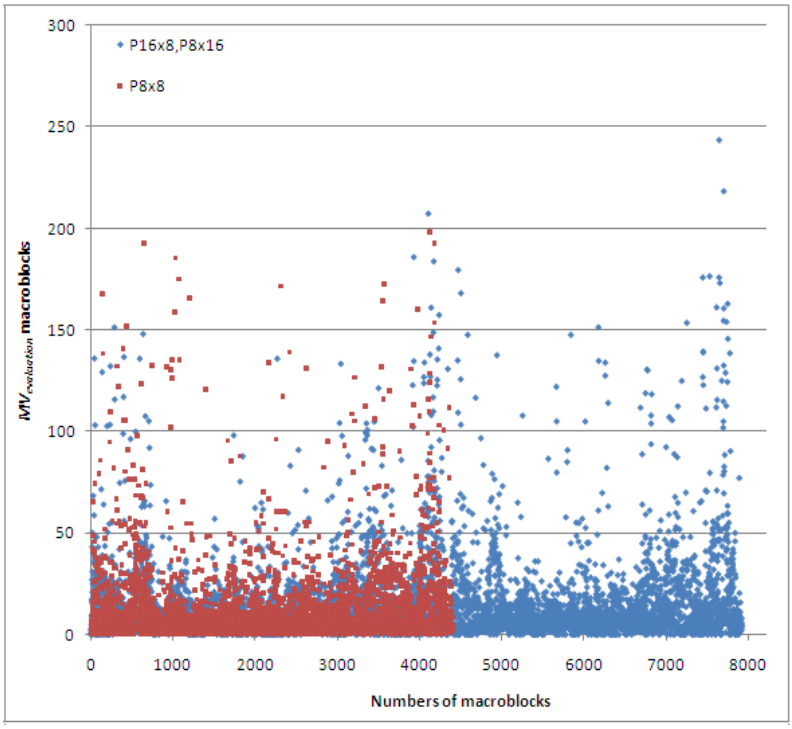


Figure 2-3 Motion vector value distribution of Foreman (CIF)

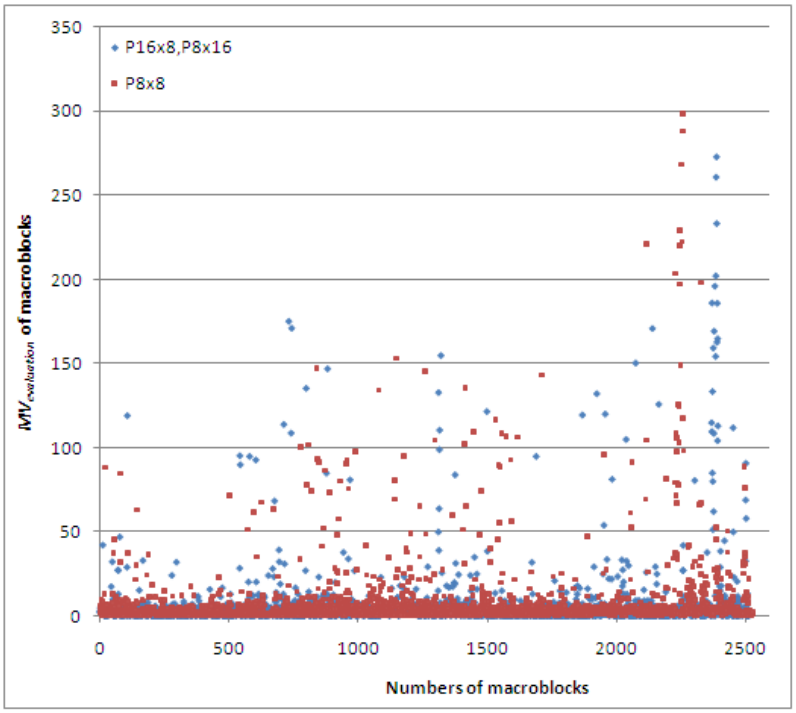


Figure 2-4 Motion vector value distribution of News (CIF)

2.2.2 Adaptive Motion Correlation Setting

The motion correlation adaptive setting process works like Figure 2-5 shows. The key strategies to implement adaptive motion correlation setting for inter mode decision:

- Make an adaptive reference MV set.
- Record $MV_{evaluation}$ in the MV set for every decided P8×8 MB.
- *Percent value*: the percentage of values less than current $MV_{evaluation}$ in current MV set.
- For accuracy, first few frames (frame number <= T_{frame}) are treated in the normal process, just update MV set with the $MV_{evaluation}$. Otherwise, the *Percent values are* calculated for mode judgment.

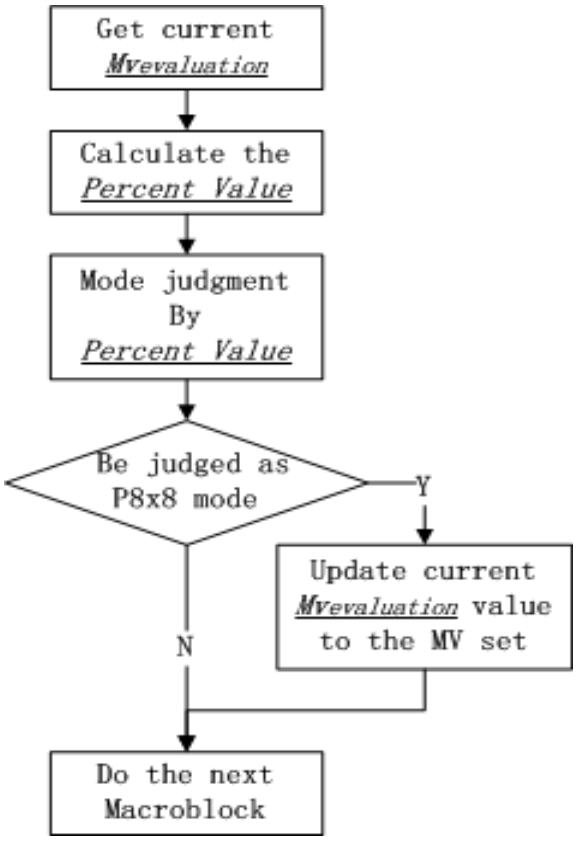


Figure 2-5 Motion correlation adaptive setting process

Tframe means only implementing early SKIP mode decision from 0 to *Tframe*, and the proposed motion correlation adaption inter mode decision works after *Tframe*.

T1 and *T2* are both the thresholds of the *Percent Value*. *Tmv* is an assistant thresholds works together with T1 to decide whether to execute the P8 × 8 motion search. T2 decides whether to overleap the P16 × 8 and P8 × 16 motion searches.

At the end, if the decided best mode is P8 × 8 mode, the motion vector set is updated.

2.2.3 Working Flow of the Proposed Algorithm

The whole flow chat of the proposed motion adaptation based inter mode decision method (MAIMD) is shown in Figure 2-6.

As above discussed, the P16x16 motion search is conducted for all the MBs. Then the $MV_{evaluation}$ is calculated. Following is the SKIP mode judgment, the SKIP-early determination [21] [22] is conducted. Then the proposed motion correlation adaptive setting process is conducted. The explanation of the thresholds is given in the Section 2.2.4, and the values set in the simulations are given in the Section 2.3.1.

The proposed MAIMD method only performs the inter mode decision to the first level of inter mode partitions. In the worst case, it conducts all the 9 modes of inter mode partitions.

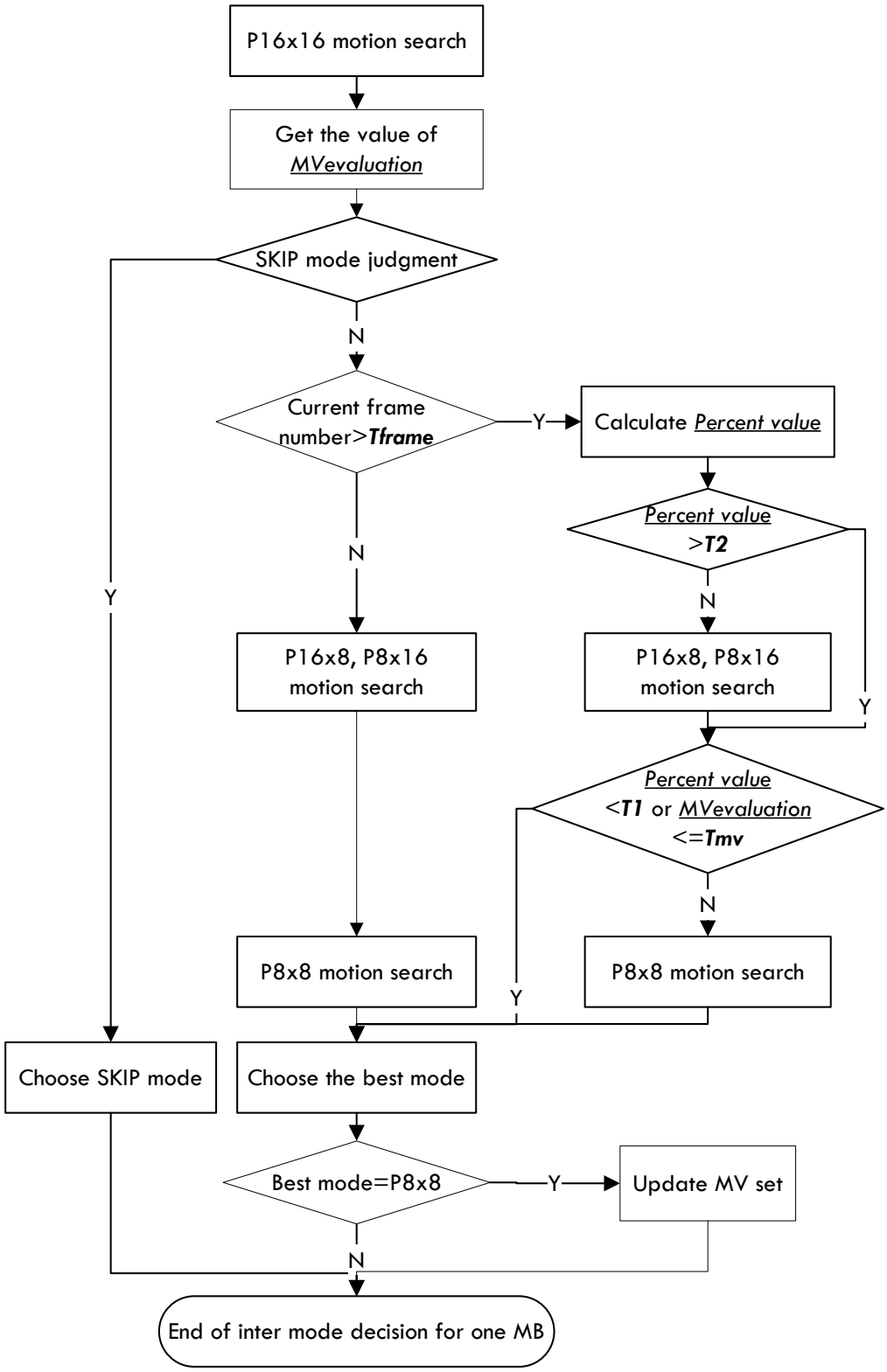


Figure 2-6 Flowchart of proposed Inter mode decision of one macroblock after T_{frame} frames

2.3 Simulation Results

2.3.1 Simulation Parameters

According to the discussion above, the threshold of motion vector is adaptive; the pre-processing frame number *Tfream* and the thresholds *T1*, *T2* of *Percent Value* and *Tmv* are set manually. By some simulation in which several parameters' combination has been tried, the thresholds are set as shown in Table 2-1.

Table 2-1 Thresholds setting for MAIMD algorithm

Name	*Value*
Tframe	8
T1	0.5
T2	0.5
Tmv	1

Table 2-2 shows the simulation conditions.

Table 2-2 Simulation conditions of MAIMD algorithm

Version	*JM 14.1*
Number of Frames	100
Group of Pictures	IPPP
RDO	ON
QP	20, 24, 28, 32
Search Range	32
Number of Reference Frames	5
Search Mode	Full Search
Motion Estimation Mode	FME
Entropy Coding Method	CABAC

All of the results are compared with the JM14.1. BDPSNR in dB and BDBR in percentage represent the equivalent difference in PSNR and bit rate respectively. The evaluation method for BDPSNR and BDBR is the average differences between the RD-curves [20]. The two values are equivalent. The ΔTotal Time is the average value of the encoding time ratio of proposed method to JM14.1 for listed four QPs.

2.3.2 Simulation Results Comparison

Table 2-3 Results of SKIP-early-only for MAIMD

Sequences		BDPSNR(dB)	BDBR (%)	ΔTotal Time (%)
CIF	Bus	-0.004460	0.079239	-3.13668
	Foreman	0.002234	-0.064720	-6.66830
	Hall monitor	0.004226	-0.125550	-14.2500
	Head with glasses	0.016145	-0.468820	-13.1403
	Highway	-0.006840	0.423012	-10.3703
	News	0.007472	-0.167390	-32.4883
	Silent	0.030608	-0.584390	-28.7347
QCIF	Akiyo	0.008018	-0.148430	-43.5214
	Coastguard	-0.010250	0.155508	-3.85014
	Container	0.007922	-0.176370	-32.9087
	Grandma	-0.007040	0.151153	-34.6901
	Foreman	-0.003630	0.088121	-6.55719
	Mother and daughter	-0.005570	0.139691	-15.2013
	Tempete	-0.009170	0.146459	-2.34613
Average		0.002119	-0.039460	-17.7045

The performance of only using SKIP-early determination [21] [22] for CIF/QCIF sequences is shown in Table 2-3.

The performance of the whole MAIMD method for CIF/QCIF sequences is shown in Table 2-4.

The time saving comparison between early-SKIP-only and whole scheme of MAIMD is shown in Figure 2-7.

Table 2-4 Simulation results of MAIMD Algorithm

Sequences		BDPSNR(dB)	BDBR (%)	ΔTotal Time
CIF	Bus	-0.10410	1.742476	-21.1355
	Foreman	-0.08347	2.007812	-21.1931
	Hall monitor	-0.03433	1.373986	-29.7337
	Head with glasses	-0.07469	2.040272	-31.1754
	Highway	-0.02424	1.241063	-26.1394
	News	-0.07009	1.392026	-45.3822
	Silent	-0.04129	0.835488	-41.0736
QCIF	Akiyo	-0.05930	1.151816	-54.4426
	Coastguard	-0.06394	1.194142	-23.0578
	Container	-0.04093	1.069401	-46.4607
	Grandma	-0.03765	0.832887	-46.4783
	Foreman	-0.09017	2.033965	-26.2096
	Mother and	-0.07748	1.603195	-32.3797
	Tempete	-0.10676	1.727835	-22.6316
Average		-0.06489	1.446169	-33.3922

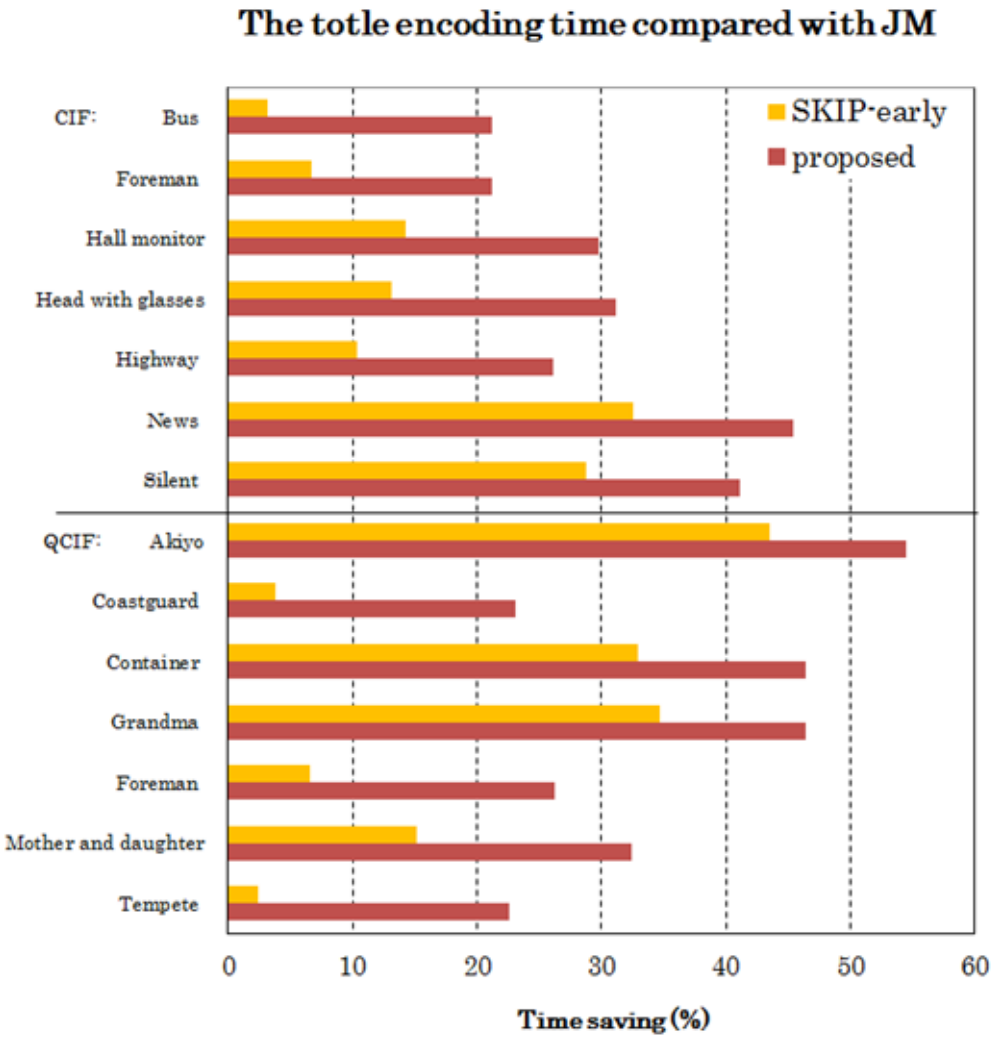
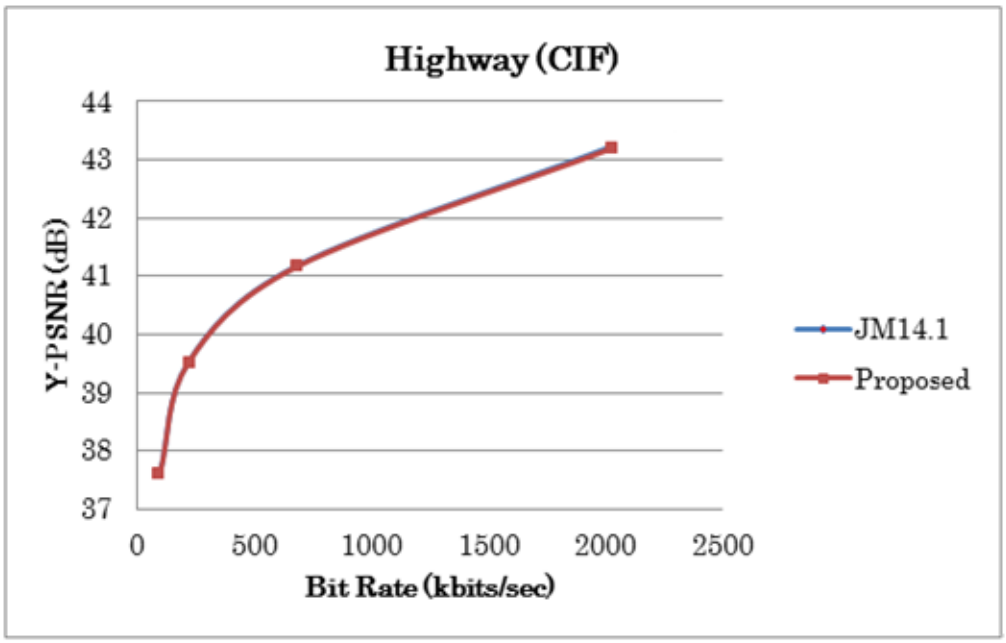
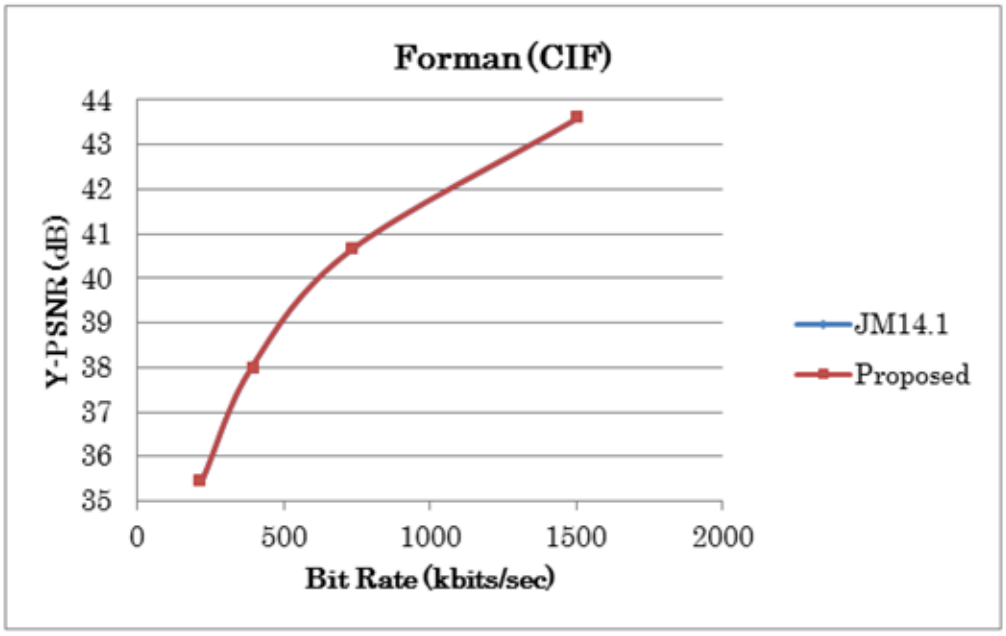


Figure 2-7 Time saving comparison between early-SKIP-only and whole scheme of MAIMD

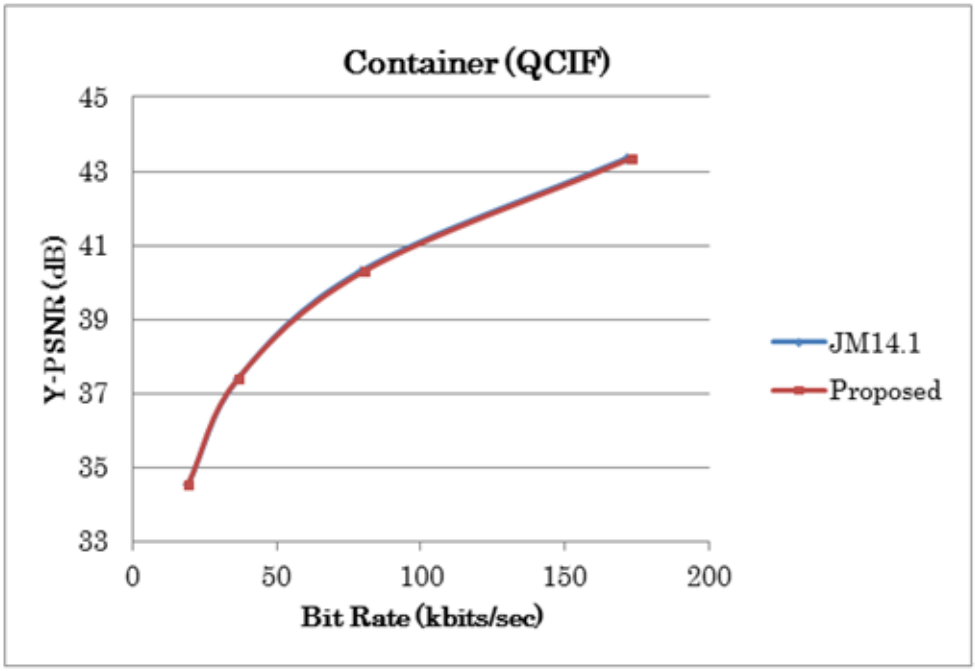
Figure 2-8 (a)—(d) show the rate-distortion curves results of the whole MAIMD algorithm. Tested sequences are Highway (CIF), Forman (CIF), Container (QCIF), Tempete (QCIF). QP values are 20, 24, 28, and 32. All the performances are compared with JM 14.1 with the parameters described above.


Highway (CIF)
Y-PSNR (dB)
44
43
42
41
40
39
38
37
0
500
1000
1500
2000
2500
Bit Rate (kbits/sec)
JM14.1
Proposed

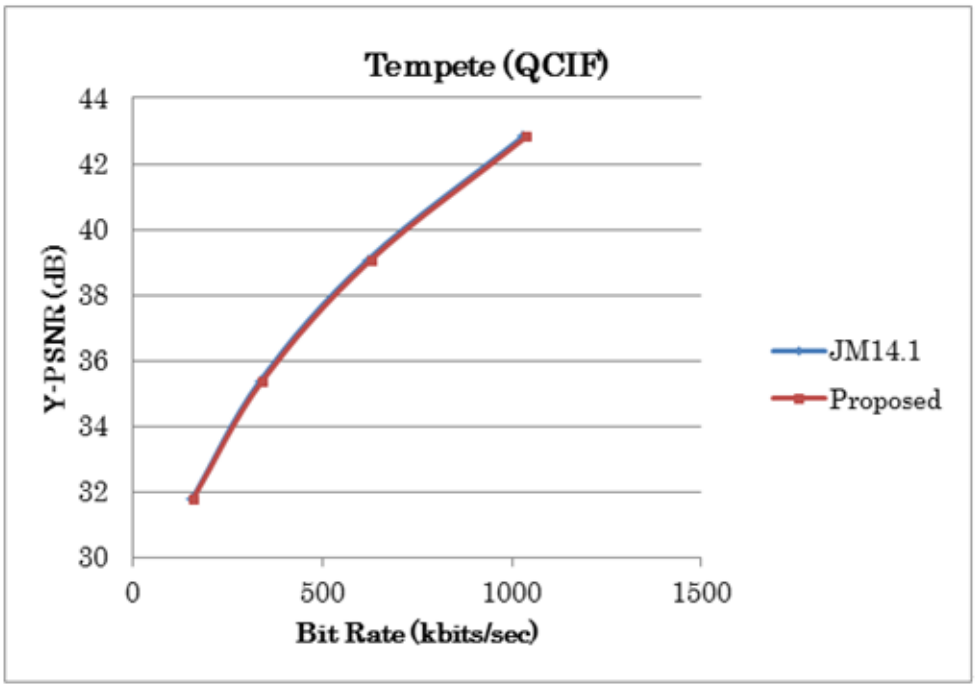
(a)


Forman (CIF)
Y-PSNR (dB)
44
43
42
41
40
39
38
37
36
35
0
500
1000
1500
2000
Bit Rate (kbits/sec)
JM14.1
Proposed

(b)



(c)



(d)

Figure 2-8 Rate-distortion curves of MAIMD algorithm

The performance comparison of proposed MAIMD algorithm and **[13]** for CIF/QCIF sequences is shown in Table 2-5. The reference paper **[13]** select the inter modes by evaluating the homogeneous and stationary of macroblocks. For time reduction, in most cases, the proposed MAIMD is able to reduce more encoding time than [13], for the other 2 cases (sequence "Paris" and "Stefan") the time reduction of the two methods are very close. For the BDPSNR and BDBR, both of MAIMD and [13] are acceptable, the quality loss is negligible. Thus, the proposed MAIMD performs better for the videos whose contents have many tiny and irregular movements.

Table 2-5 Comparison of MAIMD algorithm and [13]

Sequence		BDPSNR (dB)		BDBR (%)		ΔTotal Time (%)	
		[13]	MAIMD	[13]	MAIMD	[13]	MAIMD
CIF	Mobile	-0.005	-0.090	0.13	1.38	-9.97	-17.96
	Paris	-0.040	-0.142	0.87	2.41	-31.90	-30.55
	Stefan	-0.015	-0.072	0.33	1.29	-17.37	-16.09
QCIF	Container	-0.012	-0.040	0.30	1.06	-36.25	-46.46
	Forman	-0.062	-0.090	1.28	2.03	-25.18	-26.21
	News	-0.065	-0.130	1.18	2.23	-42.62	-50.16

2.4 Conclusion

A motion correlation adaptation method with SKIP mode early determination for H.264/AVC inter mode decision has been proposed in this chapter. The main idea of the proposed inter mode perdition is to determine the mode for inter macroblock more immediately. The modes which are evaluated as low probability to be the best mode have been ignored during the mode decision process. A lot of experiments have been done for test the performance of proposed method. On average, about 33.4% encoding time is saved by proposed adaptive inter mode decision, with only 0.06dB PSNR loss under the same bit rate, or 1.45% bit rate increase under the same PSNR. And for the sequences which have more motions, the time saving of the proposed method gains about 19.1% than only use the SKIP mode early determination. The proposed MAIMD performs better for the videos whose contents have many tiny and irregular movements.

3 Residual Feature Based Fast Inter Mode Decision

A residual feature based fast inter mode decision algorithm for H.264/AVC has been proposed in this chapter. The residual is obtained after the motion search of P16×16 mode or P8×8 mode. Different from previous works in which the positive and negative residual in a block is summed indiscriminately, in the proposed residual feature based fast inter mode decision the positive and negative residual are extracted separately. Basing on the extracted residual feature, the complexity and similarity are evaluated for the inter mode decision. According to the evaluation of similarity between different sub-blocks and the complexity of each sub-block, the most possible inter modes for current block is chosen to be conducted. In the worst case, the proposed whole scheme of inter mode decision algorithm only conducts 4 modes, which is much more effective than conducting all the 8 modes in conventional approach. The simulation results show that, comparing to JM14.1, on average, the proposed algorithm achieves 57.98% and 55.72% time-saving on CIF and 720p sequences respectively, with equivalent 0.219dB PSNR drop and 5.55% bit rate increase for CIF and 0.107dB PSNR drop and 3.53% bit rate increase for 720p. Compared the time reduction ratio to Wang's algorithm [16] which also utilize the residual information, the proposed residual feature based inter mode decision algorithm achieves 10.68% and 13.26% timing-reduction on CIF and 720p sequences respectively with less performance loss. Since the positive and negative residual are extracted separately and both of the extracted results are used for

evaluation, the proposed algorithm is able to predict the mode more precisely, which results in a better encoding time reduction than Wang's algorithm [16].

3.1 Background

H.264/AVC, also known as MPEG-4 Part 10 [19] is the latest video compression standard. Compared to previous video coding standards, H.264/AVC achieves a significant bit-rate reduction, a small PSNR gain, while increasing the computational cost by more than one order of magnitude [23]. The standout compression efficiency of H.264/AVC is attributed by a series of advanced coding techniques and features, including the multiple reference frames and variable block size motion estimation (VBSME) with Lagrange rate-distortion optimization, integer discrete-cosine transform (DCT), quantization, context-based entropy coding, in-loop-deblocking filter, and so on. All of those features are somehow computation-hungry, while the main computational complexity still comes from the motion estimation (ME) part. Motion estimation costs about 60% - 80% portion of computation in the whole encoder system [24]. The computation bottleneck in inter frame encoding is the rate-distortion optimization (RDO) based variable block size mode decision.

In this chapter, a fast inter mode decision algorithm based on the residual analysis has been proposed. The residual feature based inter mode decision is conducted for both 2 levels of inter mode decision. P16×16 motion search and SKIP mode check is firstly performed to obtain the residual of P16×16 blocks. The residual characteristic of current macroblock and all the sub-blocks are extracted and the complexity and similarity are evaluated for choosing the next mode to be

conducted. If P8×8 mode is chosen, the similar method is applied in the second level. Finally, the mode with the minimal R-D cost is determined as the final mode. The proposed residual feature based inter mode decision is able to achieve more time reduction than the previous work.

3.2 Residual Based Motion Analysis

3.2.1 Residual Correlation Discussion

Some existing fast inter mode decision algorithms, such as Yu's [17], evaluate each macroblock by the complexity of texture in the uncompressed one. The macroblocks with smooth texture are judged as big-partition modes, while the ones with rough texture are judged as small-partition modes. However, this scheme is not very accurate. For the macroblocks on the still and complex background, many of those are judged as P8×8 mode under the above scheme; while in fact, most of their best mode are big partition modes in conventional full-mode-search inter mode decision. Moreover, according to the R-D cost, described in Equation 2-1, the most impacting factor is the bits R. R is mainly decided by the value of residual matrix, but not the original image. Large residual often indicates an inappropriate mode prediction. In the case of big-partition modes, large residual often means that smaller partition mode should be adopted. Thus, in the proposed algorithm, the residual based criterions is used to do the inter mode prediction.

Literature [16] and [18] use residual to assist the prediction. Both [16] and [18] evaluate the residual without distinguishing the positive and negative values. Literature [16] focuses on getting the motion edge by residual. However, edges are

not always related to the moving object. Besides, the evaluation methods in [16] cannot extract the edge information exactly. The algorithm proposed in [16] firstly sums up the residual in every 4×4 block. It causes performance loss, since the positive and negative numbers are counteracted. Through the summation, some partitions with both big positive and negative residual values are described as plain area. All of the criterions proposed in [16] are based on the sum of residual in 4×4 blocks. So the mode judging method of [16] cannot be accurate enough. In literature [18], the evaluation is also based on the residual of 4×4 blocks. It evaluates the residual data by root mean square error (RMSE) without considering the average, which also causes inaccuracy. Moreover, in both [12] and [18], the first level (16×16 block) mode decision is performed based on the characteristics of four 8×8 blocks, which are calculated by the characteristics of four 4×4 blocks. Actually, in the first level mode decision, just to evaluate the residual of 8×8 blocks is enough. In my proposed algorithm, more effective criterions are defined to avoid problems mentioned above and reduce the inaccuracy and redundancy. The proposed algorithm utilizes the more effective criterions in [25] to avoid the aforementioned problems and reduce the inaccuracy and redundancy.

3.2.2 Evaluation Method Based on Residual Analysis

The residual produced by the motion estimation is the difference of current block and the reference block, which means the best matching block of current block in the reference frames. Both the sign and the absolute value of residual are important information to get the relationship between current block and its reference block. In the residual, the positive and negative values denote different directions of

the difference between current pixels and reference pixels, which means the sign of residual indicates the motion direction. And for the value of residual, the larger absolute value indicates the bigger mismatch.

Consequently, for mode decision, both the sign and the absolute value of residual are the key attributes. First, the sign is used to distinguish the motion trend in a group of pixels. Secondly, within a specific sign, the absolute value reflects the degree of mismatch. The proposed residual feature evaluation methods work as the following steps: Assuming the size of current block is n × n. To extract its characteristics, let r = { rij , i, j ∈ [1, n]}, each pair of (i, j) represents a pixel in the n × n residual block Q. r ∈ **R n×n** , rij ∈ **R** is the residual value of the i-th row, j-th column (i, j ∈ [1, n]). Then, the function f is performed on the residual block Q to get the characteristic of current block, which is defined as:

$$
\begin{aligned}
&f : R_{n\times n} \rightarrow R_{2\times 1} \\
&q = f(r) = (q_p, q_n)^T \\
&q_p = \frac{1}{2}\sum_i \sum_j (r_{ij} + |r_{ij}|) \\
&q_n = \frac{1}{2}\sum_i \sum_j (r_{ij} - |r_{ij}|)
\end{aligned}
$$

Equation 3-1

The complexity of current block q is defined like:

$$CPX[q] = \sqrt{(q_p^2 + q_n^2)}$$

Equation 3-2

The similarity of two n×n residual blocks Q1 and Q2 is defined as distance between q1 and q2:

$$SIM[q_1, q_2] = \sqrt{(q_{1p} - q_{2p})^2 + (q_{1n} - q_{2n})^2}$$

Equation 3-3

3.3 Proposed Fast Inter Mode Decision Algorithm

3.3.1 Mode Judgment Process

Assuming the current processing block M consists of 2n×2n pixel, and M is divided into four n×n sub-blocks named A, B, C, D, as shown in Figure 3-1. The characteristics q_A, q_B, q_C, q_D of block A, B, C, D are calculated by Equation 3-1 respectively. The characteristics of M are obtained by Equation 3-4.

$$q_M = q_A + q_B + q_C + q_D$$

Equation 3-4

The judgment of the modes in one level is shown in step 1 to 6. Here n is 8 for the first level mode decision and 4 for the second level mode decision. During this process, at most two inter modes are conducted. Note that the P 8×8 mode decided during Step 2 of the first level mode decision (n = 8) is further conducted in step 1 of the second level mode decision (n = 4).

· Step 1: Do the $2n \times 2n$ block motion search.

· Step 2: If $CPX[q_M]$ is larger than T_1, or in $CPX[q_A]$, $CPX[q_B]$, $CPX[q_C]$ and $CPX[q_D]$ if there are more than one which is larger than T_2, then conducts the P $n \times n$ mode and go to Step 6.

· Step 3: If $CPX[q_M]$ is less than T_3, go to Step 6.

· Step 4: If both $SIM[q_A, q_C]$ and $SIM[q_B, q_D]$ are less than T_4, then conducts the P $n \times 2n$ mode and go to Step 6.

· Step 5: If both $SIM[q_A\ q_B]$ and $SIM[q_C, q_D]$ are less than T_4, then conducts the P $2n \times n$ mode and go to Step 6.

· Step 6: Processing of this level ends.

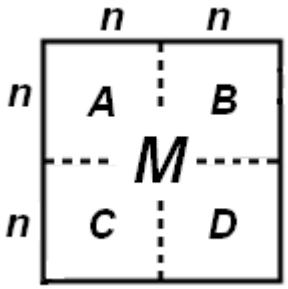


Figure 3-1 The partition of block M with 2n×2n pixel

T_1, T_2, T_3, T_4 are thresholds set before running. To get the thresholds, the sequence of Foreman with QCIF size is tested under QP 28. After exhausted experiments, the thresholds with the best performance are set as shown in Table 3-1.

Table 3-1 Thresholds setting

For first level mode decision				For second level mode decision			
T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
850	200	400	110	350	50	80	80

3.3.2 Efficiency evaluation of proposed judgment methods

The efficiency of proposed judgment methods is evaluated by matching rate. The matching rate is calculated as the percentage of macroblocks for which the proposed criterion yields the same result as the full-mode-search in conventional JM. The matching rate results of the first level modes are listed in the Table 3-2.

According to the processing order of Step 1 to 6, the P8×8 mode is judged first, so the matching rate should be the largest. The 71.34% is large enough for the first judgment step. Then the P16×16 mode and SKIP mode are calculated together simply. In fact for each macroblock the P16×16 mode is required to be conducted and then the SKIP mode is checked. So the 49.85% matching rate of Step 3 is enough for more possible modes to be conducted, which is good for the performance. Finally, the P8×16 and P16×8 mode are with 47.54% and 41.84% respectively. All the steps are with enough matching rate from higher rate to the lower rate in order. Thus the whole algorithm is fast and effective.

Table 3-2 The test results of the best mode

Judgment criterion	Matching rate (%)
Step 2 (P8×8 mode)	71.32
Step 3 (P16×16 and SKIP mode)	49.85
Step 3 (P8×16 mode)	47.54
Step 4 (P16×8 mode)	41.84

3.3.3 Working Flow of Proposed Algorithm

The overall scheme of proposed algorithm is shown in Figure 3-2.

First of all, the P16×16 motion search is performed for the first level mode decision. Then the SKIP mode is checked by the method introduced in [21]. If the conditions are satisfied, SKIP mode is chosen. Otherwise the Step 2 to 6 in the previous section are executed with n=8. The complexity and similarity of related blocks are calculated, and compared to the thresholds according to the rules defined in the Section 3.3.1.

And then, if the P8×8 mode is enabled, the second level mode decision for 8×8 sub-block is performed using Step 1 to 6 with n=4. The process is very similar to the first level mode decision, but the block size and thresholds are all different.

Finally the mode with the smallest R-D cost in all the conducted modes is set as the best mode for current macroblock.

The proposed scheme is effective. Even in the worst cases, the proposed fast inter mode decision algorithm conducts only 4 modes for one macroblock, which is much more effective than conducting all the 8 modes in conventional JM.

Also, this scheme is better than the method proposed in Chapter 2. The motion adaptation based inter mode decision method proposed in Chapter 2 only conducts the mode decision for the first level modes, and in the worst case it has to do the motion search for all the 8 modes.

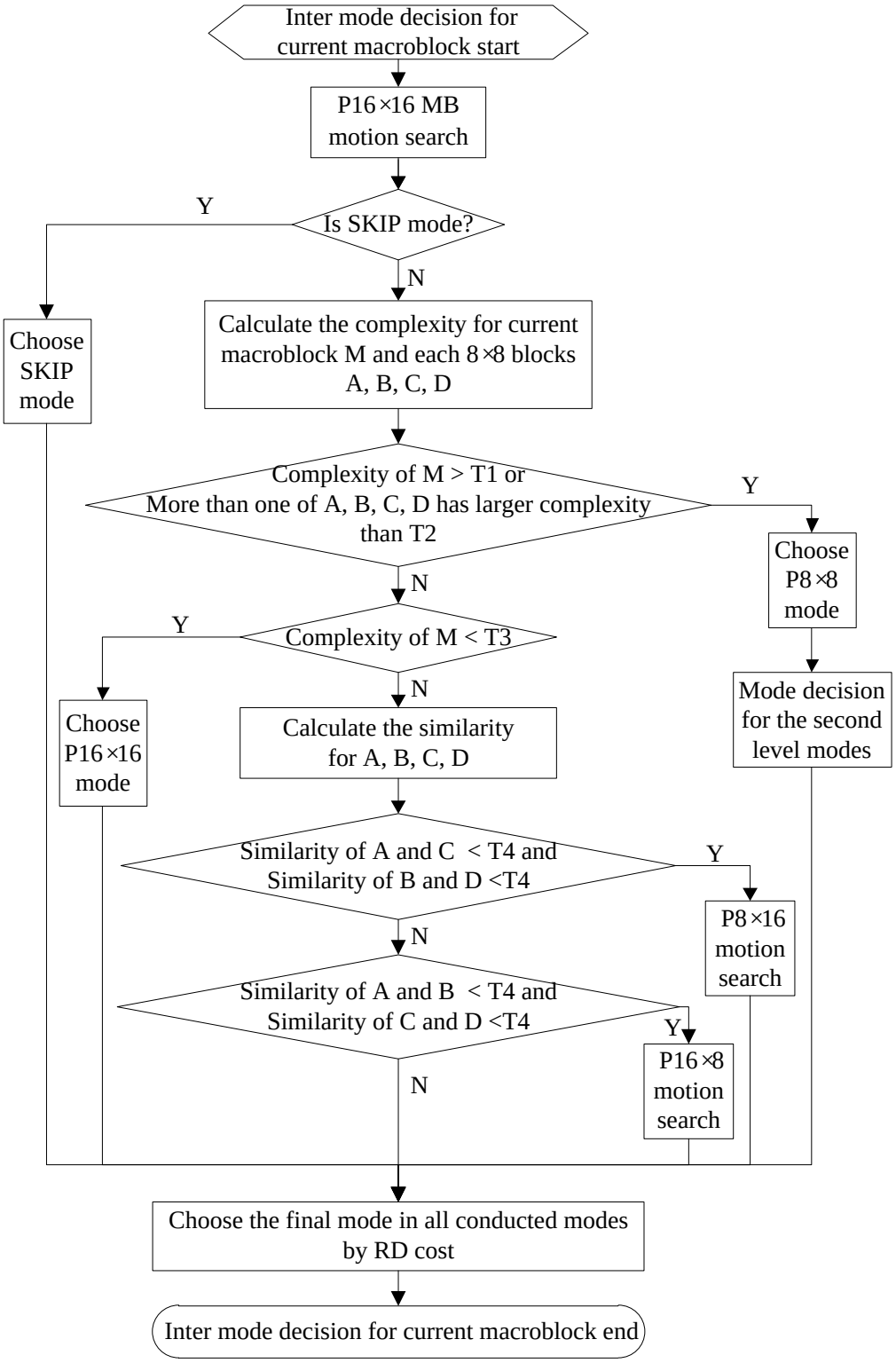


Figure 3-2 The whole flowchart of proposed residual based fast inter mode decision algorithm

3.4 Simulation Results and Discussions

3.4.1 Simulation Parameters

The proposed fast inter mode decision algorithm has been implemented in reference software JM14.1 [22] provided by JVT. CIF and 720p sequences with different characteristics are used for experiments. The encoding parameters are shown in Table 3-3 Simulation conditions. Existing fast motion estimation method EPZS [26] is adopted as the search mode for all test cases. For comparison, Wang's [16] algorithm, which also utilize residual information in the optimization, is implemented and tested under the same environment. The simulation is accomplished on a PC with Intel(R) Core™ 2 Duo CPU E8500 @3.16GHz and 3.25GB RAM.

Table 3-3 Simulation conditions

Version	JM 14.1
Frame Size	CIF (352×288), 720p (1280×720)
Number of Frames	100
Group of Pictures	IPPP…
RDO	ON
QP	20, 24, 28, 32
Search Range	CIF: 32, 720p: 64
Number of Reference Frames	1
Search Mode	EPZS
Motion Estimation Mode	FME
Entropy Coding Method	CABAC

3.4.2 Simulation Results Comparison

Table 3-4 and Table 3-5 show the performance of proposed algorithm and Wang's method [16]. For both algorithms, nine CIF sequences and nine 720p sequences are tested respectively. All of the results are compared with conventional full-mode-search JM code and use the same encoding parameters referred above. The time-reduction ratio (TRR) of the whole encoding time in percentage between fast algorithm and conventional JM is defined in Equation 3-5. BDPSNR in dB and BDBR in percentage represent the equivalent difference in PSNR and bit rate respectively. They are calculated by the method proposed in [20]. Also, the differences of the items between proposed and Wang's algorithm are given.

$$TRR = (1 - \frac{EncodingTime[\,fast\,]}{EncodingTime[JM\,]}) \times 100\%$$

Equation 3-5

Figure 3-3 (a – d) and Figure 3-4 (a – d) show the TRR and R-D performance comparison of sequences (CIF: Akiyo and Paris, 720p: ShuttleStart and Sailormen) with QP 20, 24, 28, 32 respectively.

In Table 3-4 for CIF sequences, it shows that on average, the time saving of proposed algorithm achieves 57.98% which is 10.68% higher compared to Wang's algorithm [16]. On average, the BDPSNR of proposed algorithm is -0.219dB which is 0.134dB higher than Wang's, and the BDBR of proposed algorithm is about 5.55% which is 3.18% less than Wang's. For the sequence "Akiyo" with the highest average TRR difference 15.04%, at all tested QPs, the TRR of proposed algorithm are much higher than Wang's (Figure 3-3 (a)), and the difference is up to about 19%

on QP 20. And the PSNR drop is small, the R-D curve of proposed one and Wang's are nearly the same (see Figure 3-4 (a)). For the sequence "Paris" with the minimal TRR, proposed algorithm achieves higher TRR on four different QPs than Wang's (see Figure 3-3 (b)), and the R-D performance of proposed one is obviously much better than Wang's [16] (see Figure 3-4 (b)).

In Table 3-5 for 720p sequences, it shows that on average, the time saving of proposed algorithm achieves about 55.72% which is 13.26% higher than Wang's. On average, the BDPSNR of proposed one is -0.107dB which is 0.021dB higher than Wang's, and the BDBR is also 0.67% less than the compared one. For sequence "Shuttle-Start" with the highest average TRR difference 20.23%, at all tested QPs, the TRR of proposed algorithm are much higher than Wang's (see Figure 3-3 (c)), and the difference is up to about 23% on QP 20. And the PSNR drop is small, the R-D curve of proposed one and Wang's are nearly the same, and very close to the curve of JM (see Figure 3-4 (c)). For the sequence "Paris" with the minimal TRR, proposed algorithm achieves higher TRR on four different QPs than Wang's (see Figure 3-3 (d)), and the R-D performance of proposed one is a little better than Wang's, and very close to the curve of JM (see Figure 3-4 (d)).

Table 3-4 Performance comparison of proposed algorithm and Wang's on CIF sequences

CIF Sequences	TRR (%)			BDPSNR (dB)			BDBR (%)		
	[16]	Prop.	Prop. vs. [16]	[16]	Prop.	Prop. vs. [16]	[16]	Prop.	Prop. Vs. [16]
Akiyo	46.27	61.31	15.04	-0.148	-0.153	-0.005	2.64	2.86	0.22
Container	49.17	59.81	10.64	-0.126	-0.118	0.008	3.47	3.25	-0.22
Foreman	47.16	54.08	6.92	-0.541	-0.346	0.195	14.15	8.80	-5.35
Hall_monitor	46.42	58.34	11.93	-0.189	-0.133	0.056	6.15	4.20	-1.95
Headwithglasses	46.64	55.70	9.06	-0.376	-0.222	0.153	10.92	6.32	-4.60
Highway	46.68	59.16	12.47	-0.197	-0.161	0.036	10.11	8.18	-1.92
News	48.36	61.60	13.24	-0.475	-0.325	0.151	10.04	6.71	-3.33
Paris	46.90	53.77	6.87	-0.697	-0.271	0.426	12.05	4.60	-7.45
Silent	48.09	58.06	9.97	-0.424	-0.239	0.185	9.03	4.98	-4.05
AVERAGE	47.30	57.98	10.68	-0.353	-0.219	0.134	8.73	5.55	-3.18

Table 3-5 Performance comparison of proposed algorithm and Wang's on 720p sequences

720p Sequences	TRR (%)			BDPSNR (dB)			BDBR (%)		
	[16]	Prop.	Prop. vs. [16]	[16]	Prop.	Prop. vs. [16]	[16]	Prop.	Prop. vs. [16]
City	42.97	54.91	11.94	-0.112	-0.090	0.021	3.62	2.85	-0.77
Cyclists	41.61	58.42	16.81	-0.133	-0.120	0.012	5.20	4.76	-0.44
Night	42.18	51.21	9.02	-0.236	-0.177	0.058	6.08	4.54	-1.54
Optis	40.80	55.22	14.42	-0.069	-0.060	0.009	2.30	1.98	-0.32
Raven	42.47	59.98	17.51	-0.132	-0.115	0.017	4.23	3.73	-0.49
Sailormen	42.43	50.76	8.33	-0.164	-0.141	0.024	5.13	4.31	-0.82
Sheriff	41.43	53.65	12.22	-0.129	-0.111	0.018	3.84	3.30	-0.54
Shields_ter	43.46	52.33	8.86	-0.113	-0.079	0.033	3.88	2.72	-1.16
ShuttleStart	44.82	65.05	20.23	-0.065	-0.069	-0.003	3.50	3.53	0.03
AVERAGE	42.46	55.72	13.26	-0.128	-0.107	0.021	4.20	3.53	-0.67

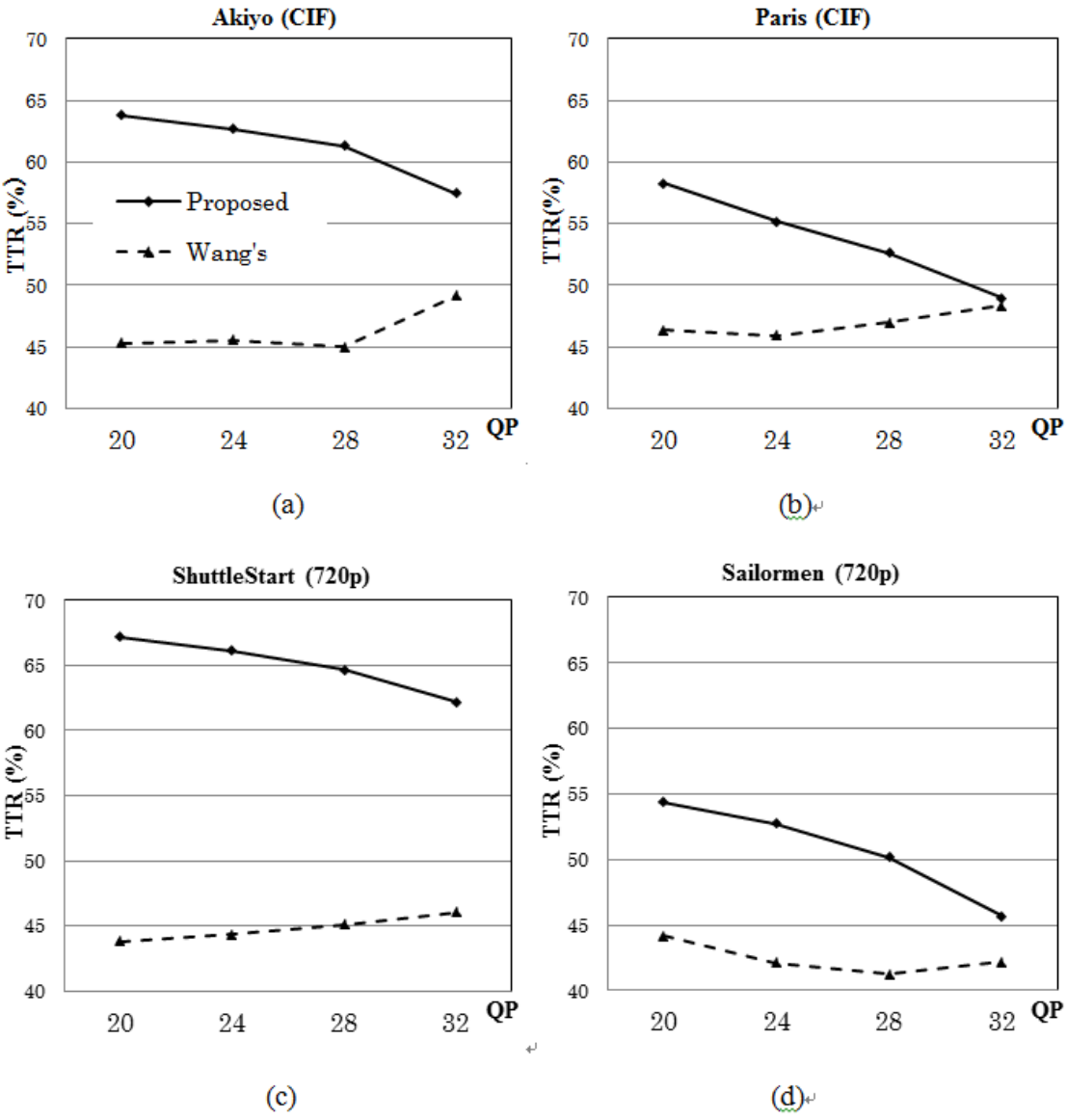


Figure 3-3 TRR of Akiyo (CIF), Paris (CIF), ShuttleStart (720p) and Sailormen (720p) at QP 20, 24, 28 and 32

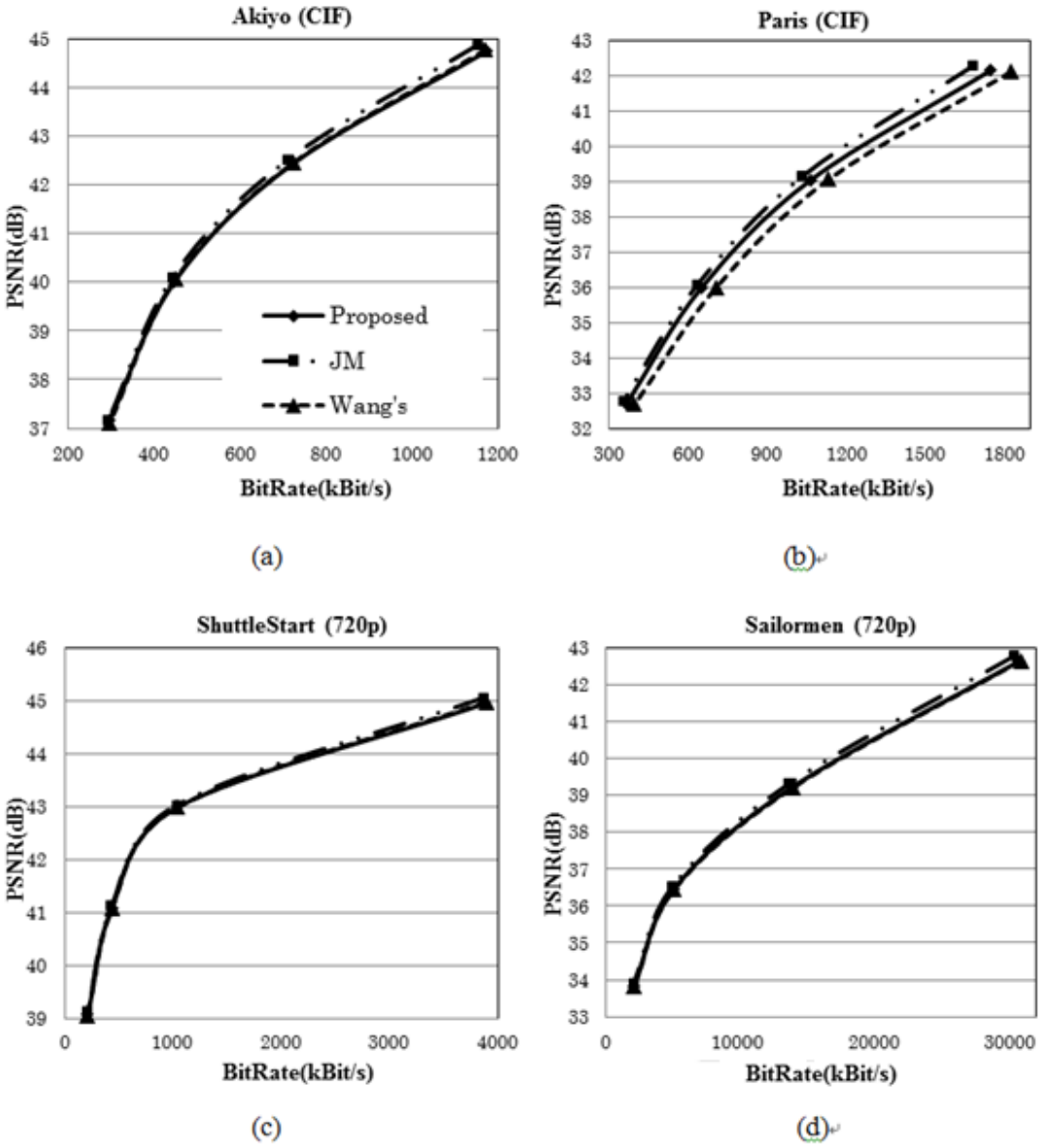


Figure 3-4 R-D performance comparison of Akiyo (CIF), Paris (CIF), ShuttleStart (720p) and Sailormen (720p)

3.4.3 Saving Factor Comparison

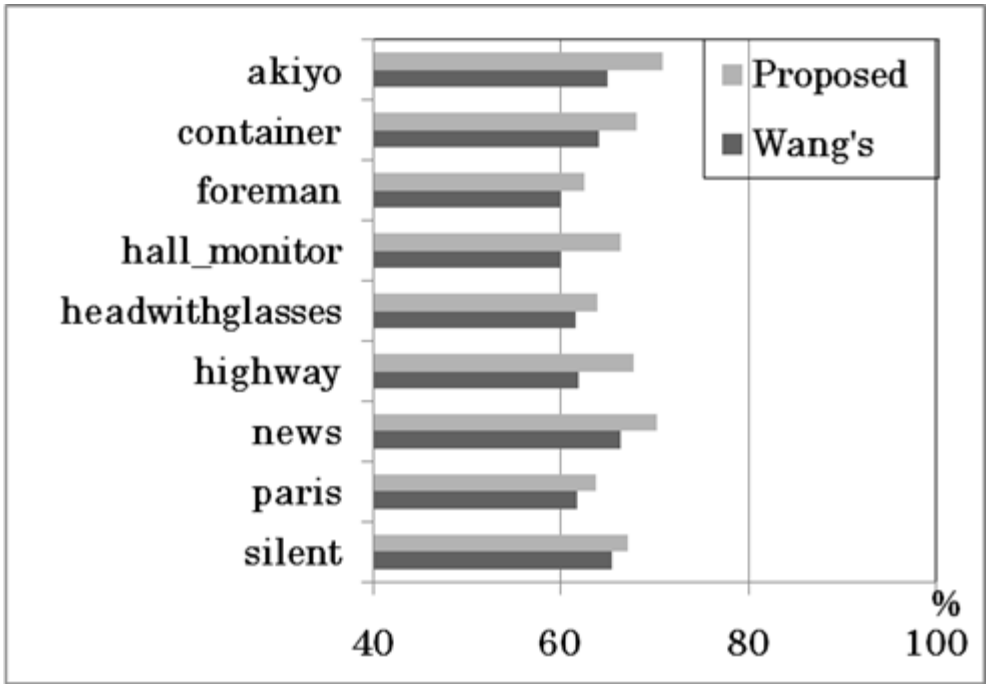
Figure 3-5 shows the saving factor comparison of proposed algorithm and Wang's [16]. The saving factor (SF), which means the percentage of saved mode-search amount, is defined in

$$SF = (1 - \frac{\#RDsearch[fast]}{\#RDsearch[JM]}) \times 100\%$$

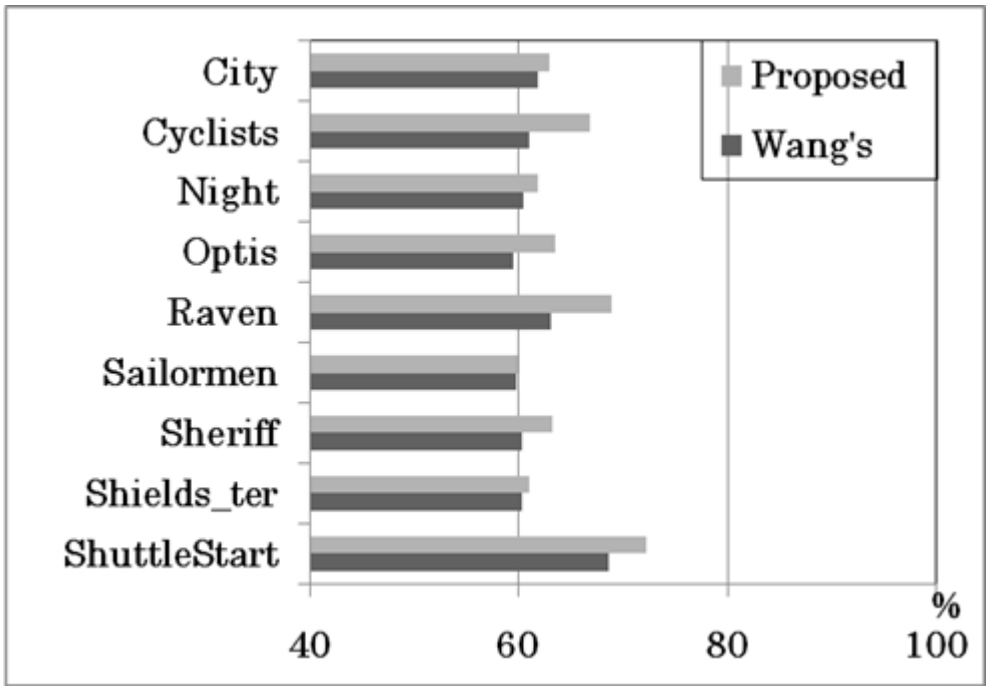
Equation 3-6

From Figure 3-5, we can see that, for all the tested sequences, the proposed algorithm achieves higher saving factor than Wang's [16]. For the sequences with more irregular moving or shaking in the video content, the proposed algorithm wins more on the saving factor than Wang's [16], for example, the CIF sequence "hall-monitor" and "highway", the 720p sequence "Cyclists" and "Raven". It means that, the proposed algorithm achieves a better analysis and utilization on the residual feature than Wang's [16]. So that, the proposed algorithm is able to avoid more redundant motion estimation, and then saving more encoding time.

The saving factor of the proposed algorithm is at least 60%, and it's up to 72% for the best case. The proposed algorithm can significantly avoid the unnecessary motion estimation. Thus the computational complexity of the inter mode decision is reduced by the proposed residual feature based inter mode decision algorithm, the whole encoding time is largely saved comparing to not only the conventional inter mode decision method, but also the previous residual based inter mode decision method Wang's [16].



(a) SF of CIF sequences



(b) SF of 720p sequences

Figure 3-5 SF comparison of proposed algorithm and Wang's

3.5 Conclusion

A Fast inter mode decision algorithm based on residual feature in H.264/AVC has been proposed. The residual are obtained from P16×16/P8×8 motion searche. The positive and negative residual in the block/sub-block are extracted respectively. The defined similarity and complexity are calculated by the extracted residual feature. By evaluating the complexity of current MB/block and the similarity of two block/sub-blocks, the candidate modes are selected to conduct. Finally the best mode is chosen as the one with the minimal RD-cost. For one macroblock, in the worst case, the proposed algorithm only conducts 4 modes which are much more efficient than conducting all the 8 modes in conventional JM code. The experimental results show that, compare to conventional JM, on average, the proposed algorithm achieves 57.98% and 55.72% time-saving for CIF and 720p sequences respectively, with equivalent 0.219db PSNR drop with 5.55% bit rate increase for CIF, and 0.128db PSNR drop with 3.53% bit rate increase for 720p. In the case of 720p sequences, proposed algorithm keeps the time-reduction ratio and gets better performance on the PSNR and bit rate than the case of CIF sequences. Compared to existing inter mode decision algorithm which also relate the residual data, proposed algorithm achieves 10.68% and 13.26% timing-reduction on CIF and 720p sequences respectively with less the performance loss. And for all the tested results with different QPs and sequences, the proposed algorithm has more timing-reduction ratio than the compared one.

4 Encoder-unconstrained User Interactive Partial Decoding Scheme

High-definition (HD) videos become more and more popular on portable devices these years. Due to the resolution mismatch between the HD video sources and the relative low-resolution screens of portable devices, the HD videos are usually fully decoded and then down-sampled (FDDS) for the displays, which not only increase the cost of both computational power and memory bandwidth, but also lose the details of video contents.

In this chapter, an encoder-unconstrained partial decoding scheme for H.264/AVC is presented to solve the problem by only decoding the object of interest (OOI) related region, which is defined by users. It is the first work on enabling OOI oriented partial video decoding without encoder-side support. The moving foreground and background, as well as the reference inaccessible problem are the challenges. All motion vectors are decoded and recorded for OOI tracking and Decoded Partial Area (DPA) adaptation, and residual is selectively decoded according to the DPA. A simplified compression domain tracking method is utilized to ensure that the OOI locates in the center of the display area. The decoded partial area (DPA) adaptation, the reference block relocation (RBR) and co-located temporal Intra prediction (CTIP) methods are proposed to improve the visual quality for the DPA with low complexity.

The simulation results show that the proposed partial decoding scheme provides an average of 50.16% decoding time reduction comparing to the fully decoding

process. The displayed region also presents the original HD granularity of OOI. The proposed partial decoding scheme is especially useful for displaying HD video on the devices of which the battery life is a crucial factor.

4.1 Background and Related Work

Along with the development of multimedia technologies, high definition (HD) videos are widely used in multimedia devices. The digital video contents with 720p, 1080p, 4k×2k and even higher resolutions also become popular. To be compatible with HD/super-high-definition (SHD) video contents, the digital display technology has been improved in terms of screen size and pixel density. While for portable devices, since the screen size is typically limited to 3–5 inches within one-hand grasp, and the pixel density that human eyes can accommodate is also limited to about 300 pixels per inch (ppi) for a comfortable viewing distance around 12 inches [27], the resolution of general-use displays for portable devices is usually lower than the HD/SHD standard. Taking the latest popular iPhone 4s mobile phone [28] as an example, iPhone 4s is equipped with a so-called Retina Display of 960×640 resolution in a 3.5 inch screen with 326 ppi. Although the Retina Display reaches the limit in ppi for human eyes, it is still not able to provide enough resolution for displaying HD/SHD video contents. In additon to the video quality, the power consumption is also an important issue for portable devices of which the battery capacity is relative limited. The latest video coding standard H.264/AVC [19] not only provides high compression efficiency but also introduces heavy computational workloads. Thus, a video decoder with lower computational cost and less memory access is desirable.

A typical solution for displaying the HD/SHD video source on a portable device is first running a fully decoding process and then down-sampling the decoded video to the display resolution. (The process is denoted by FDDS in the following). The fully decoding process costs not only lots of computing power but also memory bandwidth, while only a relatively low resolution video is required for displaying. On the other hand, the down-sampling decreases the visual quality, and consumes un-negligible power as well. Furthermore, the conventional FDDS solution usually introduces additional quality degradation such as picture freezing [29], which results in large quality loss of the HD content, especially a severe problem for portable devices due to the limited screen size.

In order to provide terminal adaptation, the Scalable Video Coding (SVC) [30] was proposed to encode a high-quality video bit-stream consisting of some subsets of bit-stream [31]. The subset bit-stream that accommodated to the relatively lower resolution of portable device is able to derive by dropping the high-quality video bit-stream. However, SVC needs the encoder side supporting and costs larger storage, while only few kinds of specific resolution supported by the encoder are able to be played. To solve this problem, a down-sampling transcoding method [32] is proposed for H.264/AVC decoder to offer the complexity scalability by directly decoding the compressed video at a downsized resolution, while this method introduces a heavy error propagation of 5dB and the details in the video contents are lost. For all above mentioned approaches, the video contents presented to the users are no longer HD/SHD resolution, and many details are lost comparing to its original resolution.

As a matter of fact, when a segment of video content is played, people are generally focus on a certain object of interest (OOI) in the field of vision, and prefer more details in the OOI than other portions. Although it is possible to only display partial of the video content in the frame which includes the OOI with the original granularity, it costs a waste of fully decoding the whole frame. A Flexible Macroblock Order (FMO) based decoder [33] is proposed to decode part of the video content at the cost of an encoder side modification to satisfy the bit-stream format requirements of the decoder. However, the OOI is fixed by the encoder, and it is unable to serve live video such as real-time Internet protocol television (IPTV). Besides, it increases the bit-rate of bit-stream. A real-time communication based CODEC system [34] is also proposed to support the user defined region decoding and display. The video source is encoded to several slices, and the certain portion, which decided by the user at the mobile terminal, is sent to the server via network. Then, the bit-stream of the slices related to the user-defined certain portion are grouped together to one slice as a frame, and sent to the mobile terminal for decoding. Thus, it is possible to zoom the certain portion of the video. Although the network traffic can be reduced by sending less information to the decoder, the quality loss introduced by the slice merging is large. Moreover, the decoded region is not always able to follow the object precisely, and the encoder side supporting is always necessary as well.

In this chapter, to provide a smooth and high quality OOI oriented decoding with low computational burden at decoder side, we propose an encoder-unconstrained H.264/AVC partial decoding scheme to only decode the OOI related portion from the single layer HD video stream without any support from the

encoder side. The OOI is interactively defined by the users. Instead of a precise object tracking which increase much computational complexity, a simplified compression domain tracking method is utilized to ensure that the OOI locates in the center of the region for display (RFD) without the information about the contour of the OOI. To reduce the error propagation at the boundary of the displayed region, an decoded partial area (DPA) adaptation method is proposed to assign the extra area to be decoded around the region for display (RFD). The reference block relocation (RBR) method is also proposed to solve the reference block inaccessible problem when the decoded motion vector (MV) points to an un-decoded area. The co-located temporal Intra prediction (CTIP) is proposed to avoid the reference pixel inaccessible for the Intra MBs in P-frame. The proposed solution is illustrated in Figure 4-1. The FDDS output is shown Figure 4-1 (a) and the output of the proposed solution is shown in Figure 4-1 (b). As shown in the figure, the FDDS output is lost many details in the video contents. The input of the proposed solution is a high resolution compressed bit-stream, and the user-defined OOI is the region enclosed by a red rectangle on the man's head in Figure 4-1 (a).

(a) FDDS with a user-defined OOI (b) The proposed solution

Figure 4-1 Comparison between solutions of resolution mismatach

The output of the proposed solution is the region centered by the OOI, of which the resolution equals to that of the display to provide the user with a clear view of the OOI. Compared with the existing solutions, the proposed method provides the following advantages:

- No requirement on any modifications or support from the encoder side.
- Much lower computational complexity and power consumption compared with FDDS because of partial decoding and skipping the down-sampling process.
- Much better visual quality of OOI compared with FDDS because of keeping the granularity of OOI to make it more visible and clear for users.
- OOI that displayed is fully determined by the user freely.

The rest of this Chapter is organized as follows. From Section 4.2 to 4.5, the proposed interactive partial decoding scheme is introduced, including the definitions, working flow and the adaptation methods, etc. Section 4.6 provides the simulation results and discussions. Finally, the conclusions are drawn in Section 4.7.

4.2 Proposed H.264/AVC Partial Decoding Scheme

4.2.1 Definitions and Pre-defined Parameters

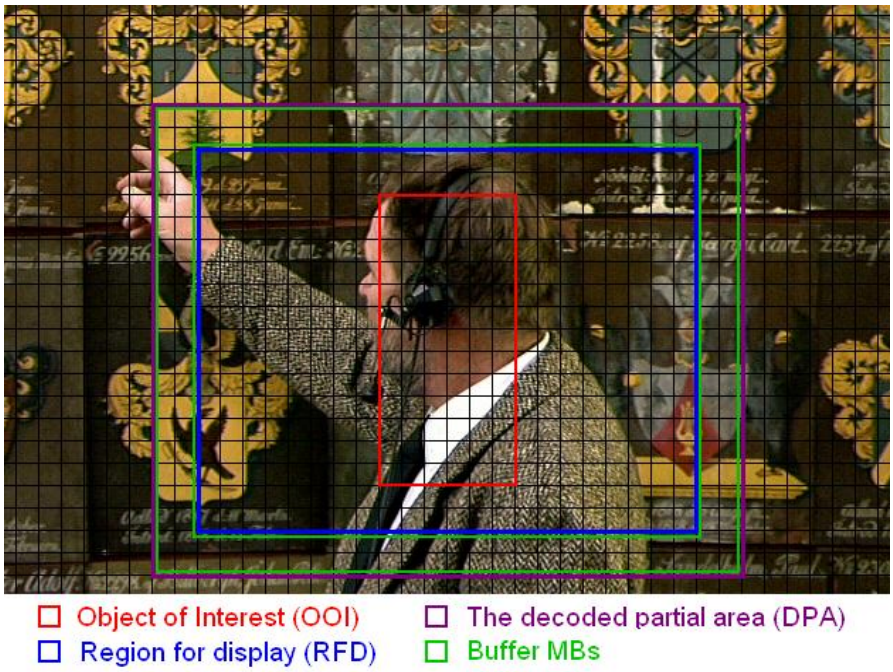


Figure 4-2 Definitions of OOI, RFD, DPA and Buffer MBs

The three key areas for the proposed H.264/AVC partial decoding scheme are shown in Figure 4-2. The definition and initialization of them are given in the following.

The object of interest (OOI) is a user defined rectangle within the displayed region like the red rectangle shown in Figure 4-2, and specified by the coordinates of the top-left and bottom-right corners of the rectangle. The location of OOI is updated frame by frame by a simplified compression domain tracking algorithm.

The region for display (RFD) is the region displayed on the screen, which is shown as the blue rectangle in Figure 4-2. The size of RFD equals to the display resolution of portable devices, defined by the width and height (W×H). The RFD in

the frame is defined by allocating the OOI in the center of RFD and is updated in a frame by frame manner.

The decoded partial area (DPA) is the specific decoded region in this partial decoding scheme, shown as the purple rectangle in Figure 4-2. In I-frames, DPA is the whole frame. For P-frames, in order to reduce the error propagation from the un-decoded region, the size of DPA is larger than RFD. In the first P-frame, the DPA is obtained by extending the RFD by N MBs in each direction. These extended MBs are called Buffer MBs in this chapter, which is the region enclosed by green lines in Figure 4-2. The number of Buffer MBs is discussed in the latter section. The location of DPA is decided after each I-frame is decoded, and then remains unchanged in the whole group of picture (GOP).

For the H.264/AVC bit-streams coded with de-blocking filter (DF) at the encoder side, DF is skipped for some MBs to further improve the efficiency in the proposed partial decoder. The region performed DF is one MB smaller than DPA in each direction. So that all the pixels required by the DF are available.

4.2.2 Structure of the Proposed Partial Decoder

The block diagram of the proposed H.264/AVC partial decoder is shown in Figure 4-3. Bit-stream parsing is also performed in the proposed solution and all the NAL units are entropy decoded. Thus the macroblock (MB) overhead information such as MB type, prediction mode, motion vector (MV), as well as the residual are obtained.

Comparing with general H.264/AVC decoder [10], the proposed one has a judgment module after entropy decoding module. Whether the residual of the current MB is going to be further decoded by IQ, IT and the following processing is

determined by decoded partial area (DPA), which is the really decoded region updated every Intra period.

The MVs of all the MBs in the current frame are recorded for compression domain OOI tracking and DPA adaptation which are described in Section 4.3 respectively. For those MBs that need to be decoded, the reference blocks have to be relocated when the motion vectors (MVs) point to the un-decoded area in the reference frames. This process is performed by the proposed reference block relocation (RBR), which is described in detail in Section 4.4. For the Intra prediction, if the reference pixels are inaccessible, the co-located temporal Intra prediction (CTIP), which is also described in Section 4.4, is proposed to solve this issue. For those MBs that are determined not to be decoded, only the MV recording is performed, and all the other decoding modules are bypassed.

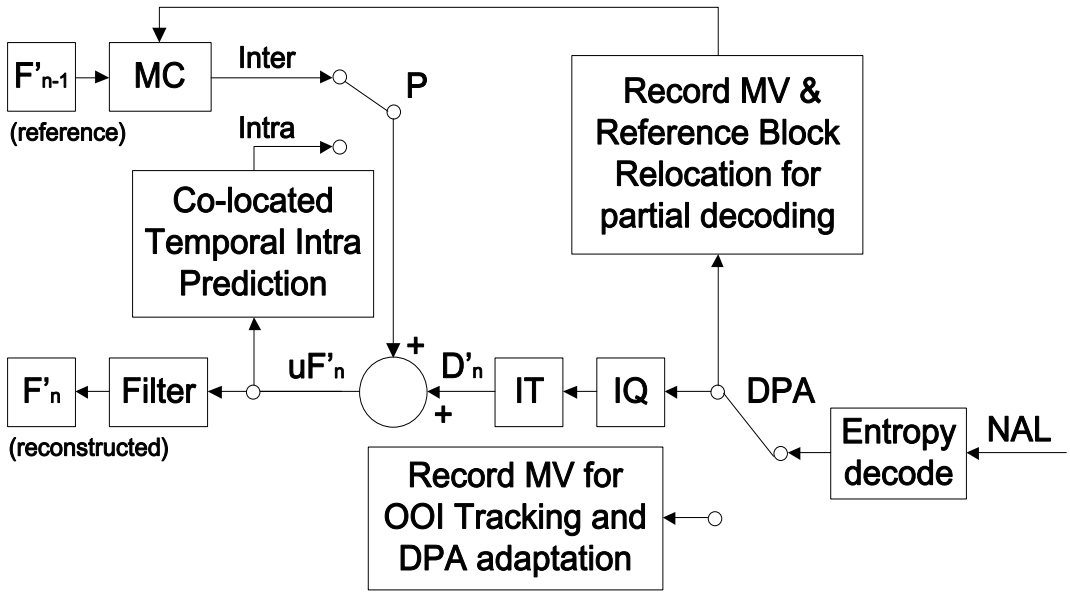


Figure 4-3 Block diagram of the proposed H.264/AVC partial decoder

4.2.3 Decoding Flow of the Proposed Partial Decoding Scheme

The decoding flow of the proposed partial decoding scheme is a two-level structure, MB level and Frame level, as shown in Figure 4-4. In the frame level, the OOI is tracked by a compression domain tracking algorithm to update the OOI and RFD every frame. The DPA is updated every GOP by estimating the size of error control Buffer MBs after the I-frame is decoded. In the MB level, different elements are obtained by parsing MB. Only the residues of MBs in the DPA are decoded. No succeeded processing is performed for the residues of the MBs that belong to the non-DPA.

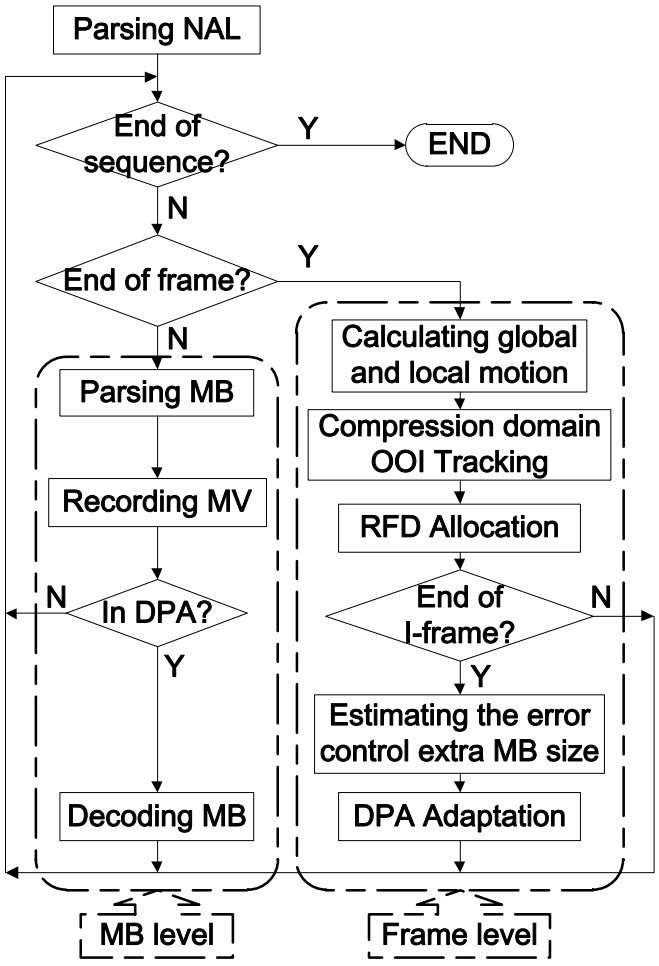


Figure 4-4 Decoding flow of the proposed partial decoding scheme

4.3 Compression Domain Motion Correlation Based Adaptation

4.3.1 Tracking of Object of Interest

In order to locate the OOI in the forward frames to adapt the motions in the video, a simplified tracking method based on the motion correlation of the compression domain is conducted. This tracking method has the possibility to be replaced by other tracking algorithms to increase the accuracy in the future.

Unlike the existing tracking methods which require the contour of the specific object, the tracking algorithm in the proposed scheme only needs a rough region in blocks. The proposed method exploits the MVs of OOI and non-OOI (the region outside of OOI) to track the moving of foreground (local motion) and background (global motion).

To accommodate multiple reference frames of P-frames, all the MVs are scaled by the reference distance between the current frame and its reference frame. The average scaled MVs in the OOI represents the local motion which is expressed in Equation 4-1, and the average scaled MVs in non-OOI represents the global motion which is expressed in Equation 4-2, where m, m' are the MB index, frame is the frame index, and dis_m represents the reference distance of the MB m.

$$LM_{frame,x} = avg(MV_{m,x} / dis_m)$$
$$LM_{frame,y} = avg(MV_{m,y} / dis_m)$$
$$m \in OOI_{frame}$$

Equation 4-1

$$GM_{frame,x} = avg(MV_{m',x} / dis_{m'})$$
$$GM_{frame,y} = avg(MV_{m',y} / dis_{m'})$$
$$m' \in \overline{OOI}_{frame}$$

Equation 4-2

The local motion is calculated every frame to help the location prediction of OOI, and the global motion is used in the decoded partial area (DPA) adaptation in the next section. The coordinates of OOI in the next frame are calculated by adding the local motion ($LM_{frame,x}$, $LM_{frame,y}$) to the coordinates of OOI in the current frame.

4.3.2 Decoded Partial Area Adaptation

The RFD location is adjusted in each frame according to the current OOI location and the DPA. The RFD must be located within the DPA, and the OOI should be in the center of RFD.

The DPA adaptation is proposed to keep the visual quality. For all the I-frames, DPA is the whole frame. The location of DPA in the P-frame is updated every GOP, and makes change after each I-frame is decoded. The motion direction and speed of the OOI/non-OOI are indicated by the difference of local/global motion during a whole Intra period. The movement of the OOI/non-OOI in one Intra period can be predicted by the previous Intra period. The DPA is updated according to the movement prediction of both OOI and non-OOI. Denoting the local and global motion of the *i*-th Intra period by (LMX*i*, LMY*i*) and (GMX*i*, GMY*i*) respectively. The definition of (LMX*i*, LMY*i*) and (GMframe,x, GMframe,y) are given by Equation 4-3 and Equation 4-4, where K is the number of frames in a Intra period, i is the Intra period index.

$$LMX_i = \sum_{frame=iK}^{i(K+1)-1} LM_{frame,x}$$
$$LMY_i = \sum_{frame=iK}^{i(K+1)-1} LM_{frame,y}$$

Equation 4-3

$$GMX_i = \sum_{frame=iK}^{i(K+1)-1} GM_{frame,x}$$
$$GMY_i = \sum_{frame=iK}^{i(K+1)-1} GM_{frame,y}$$

Equation 4-4

The extended region in each direction for DPA in the next GOP, which are denoted by Δxleft, Δxright, Δyup and Δydown respectively, are given by Equation 4-5. Finally, the DPA for P-frames in the coming Intra period is calculated by the current RFD plus the Buffer MBs in each direction and the extended region calculated by DPA adaptation. Finally, the DPA location is rounded up to the units of MB.

$$\Delta x_{left,i+1} = \max(-LMX_i, -GMX_i, 0)$$
$$\Delta x_{right,i+1} = \max(LMX_i, GMX_i, 0)$$
$$\Delta y_{up,i+1} = \max(-LMY_i, -GMY_i, 0)$$
$$\Delta y_{down,i+1} = \max(LMY_i, GMY_i, 0)$$

Equation 4-5

4.4 Solutions of Relocate the Inaccessible Reference

For the partial decoding of P-frames, in some cases, the referenced blocks/pixels may be inaccessible science the referred blocks/pixels are belonging to the undecoded area in the reference frame. In this situation, these kinds of unreachable reference blocks/pixels are largely disturbed in the motion compensation and intra block perdition during the decoding process. The visual quality in the DPA is also largely decreased by the MBs that cannot get the right reference blocks/pixels. And the error propagation will influence the visual quality in the following frames.

To decrease the influence of visual quality that caused by the inaccessible reference blocks/pixels, some approaches are proposed for inter and intra blocks respectively.

4.4.1 Reference Block Relocation for Inter Blocks

4.4.1.1 Inaccessible Reference block for Inter Blocks

For inter MBs, if the decoded MVs point to the un-decoded area in the reference frames, the reference block should be relocated to avoid the reference inaccessible problem.

As shown in Figure 4-5, solid blocks are the MBs within DPA, and dashed blocks are the ones outside of DPA. The MVs represented by dashed arrows point outside the DPA in the reference frame, so that part of reference blocks become unavailable. In this case, two approaches of reference block relocation (RBR) are proposed. The corresponding reference blocks in the decoded area of the reference frame are fetched for motion compensation after relocation.

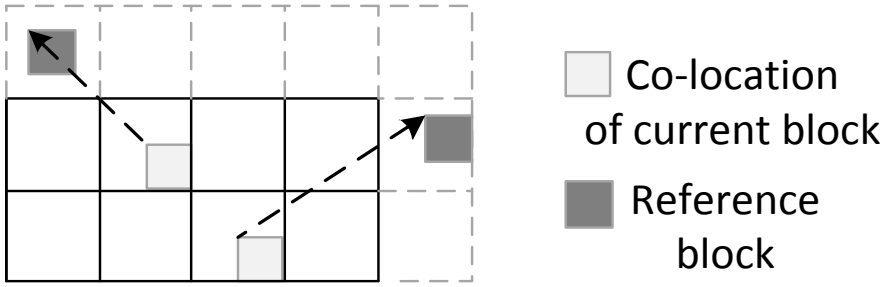


Figure 4-5 Reference block inaccessible

4.4.1.2 Direction Preferred Reference Block Relocation

The first solution of reference block relocation for inter blocks is named direction preferred RBR.

The idea of direction preferred RBR is relocating the reference block along with the MV directions until all the pixels in the reference block are within DPA, as shown in Figure 4-6.

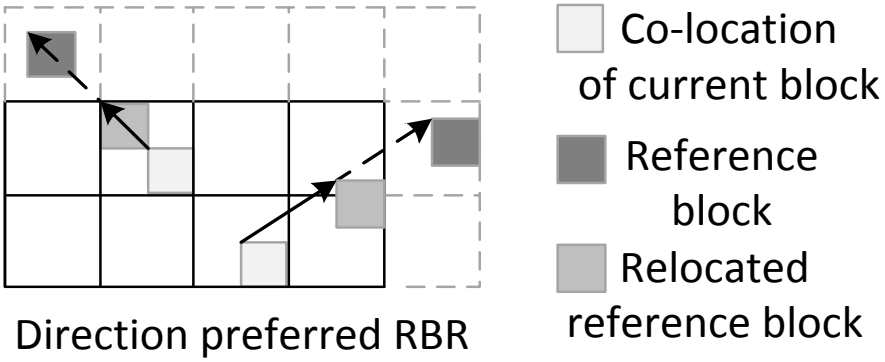


Figure 4-6 Direction preferred reference block relocation

4.4.1.3 Distance Preferred Reference Block Relocation

The second solution of reference block relocation for inter blocks is named distance preferred RBR.

The relocation method of distance preferred RBR is to simply relocate the reference block horizontally or vertically until all the pixels in the reference block are within DPA, as shown in Figure 4-7.

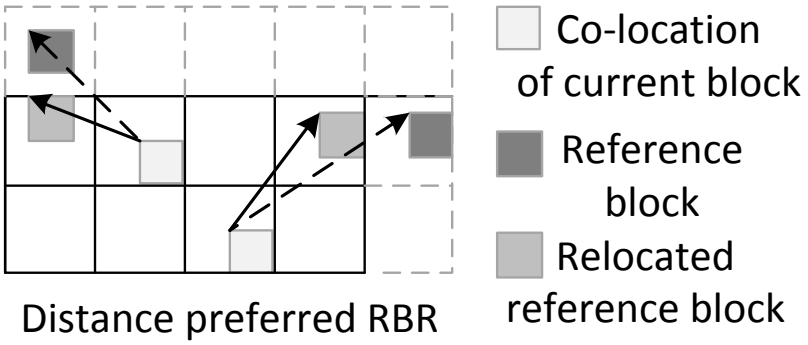


Figure 4-7 Distance preferred reference block relocation

4.4.1.4 Performance Comparison

To compare the two approaches, sequences with different content and motion features are tested, with the RFD of CIF size and 0 Buffer MBs. The experimental results of PSNR in RFD are listed in the Table 4-1. Compared with the direction preferred RBR (Dir. Pref. RBR), the distance preferred RBR (Dis. Pref. RBR) approach provides an average of 0.35dB higher quality in PSNR. Thus, the distance preferred RBR is adopted in this work due to its better performance.

Table 4-1 PSNR comparison of Distance Preferred RBR and Direction Preferred RBR.

Sequence	PSNR of Dir. Pref. RBR (dB)	PSNR of Dis. Pref. RBR (dB)	ΔPSNR (dB)
City (720p)	32.61	32.80	0.19
Shields_ter (720p)	29.06	29.49	0.43
Crowd_run (1080p)	33.38	33.81	0.43
Average	31.68	32.03	0.35

4.4.2 Reference Pixels Relocation for Intra Blocks

4.4.2.1 Inaccessible Reference for Intra Blocks

As the description in Section 1.1.2.3 about the intra predication, the reference pixels of the intra block are the neighboring pixels in the upper block and the left block. For the Intra MBs in P-frame, if the MB locates in the first row or the first column of DPA, the reference pixels of the current MB are outside DPA, thus they are inaccessible for Intra prediction. This situation is illustrated in Figure 4-8.

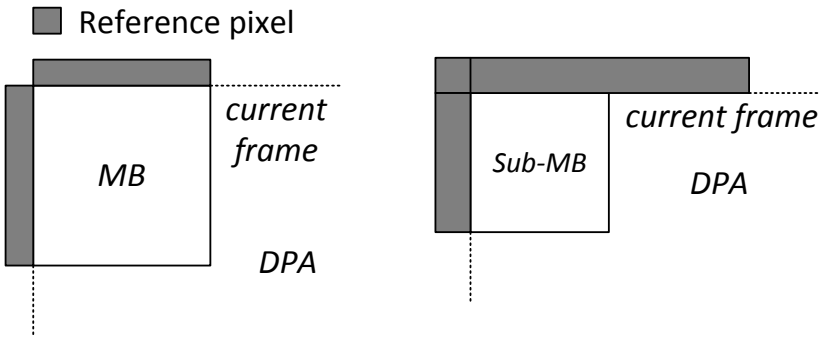


Figure 4-8 Reference pixels inaccessible

4.4.2.2 Co-located Temporal Intra Prediction

The first solution for the intra reference pixel relocation is named co-located temporal Intra prediction (CTIP), which is shown in Figure 4-9.

The idea of CTIP is to use the pixels in the first row and in the first column of the co-located MB/sub-MB in the last frame to replace the unavailable reference pixels for Intra prediction. The left-up pixel of the co-located MB in the last frame is referenced twice. The candidate reference pixels in CTIP are always accessible. Because I-frame is fully decoded (whole frame DPA) and the DPA is updated every intra period after I-frame decoding, the co-located region of DPA in the last frame is always DPA.

The advantage of CTIP is that the candidate reference pixels are in the last frame which is already decoded, thus the CTIP is able to be conducted in the normal decoding order.

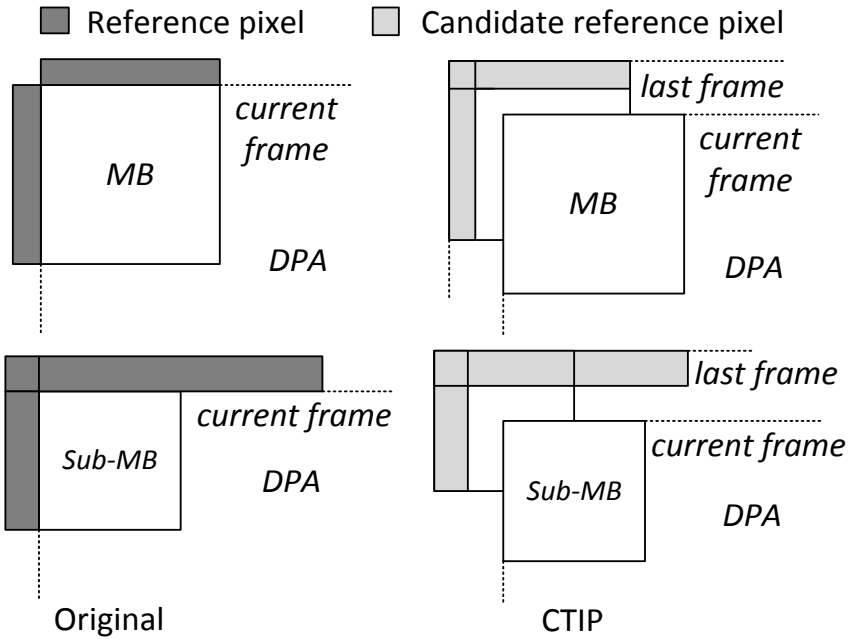


Figure 4-9 Co-located temporal intra prediction (CTIP)

4.4.2.3 Opposite Spatial Intra Prediction

The second solution for the intra reference pixel relocation is named opposite spatial Intra prediction (OSIP), which is shown in Figure 4-10.

The basic idea of OSIP is to use the first column of pixels in the neighboring MB/sub-MB on the right and the first row of pixels in the neighboring MB/sub-MB below the current one to replace the unavailable reference pixels for Intra prediction.

The disadvantage of OSIP is that the candidate reference pixels are in the right and the bottom microblock/block of current frame which is going to be decoded. Thus, to achieve the OSIP, rescheduling the decoding order is necessary. The current block has to be decoded after its right and bottom neighbor MBs being decoded.

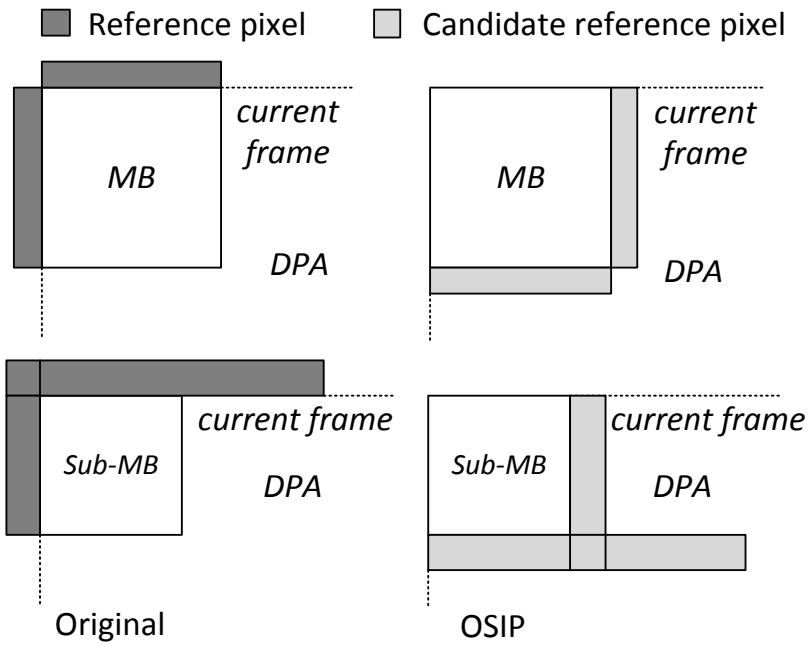


Figure 4-10 Opposite spatial intra prediction (OSIP)

4.4.2.4 Correlation Comparison

To compare the proposed two approaches OSIP and CTIP for intra reference inaccessible, the pixel correlation between the original reference pixels and the candidate reference pixels are evaluated. The pixel correlation is calculated by the sum of absolute difference (SAD) for each MB.

In order to compare OSIP and CTIP, the statistics as following is conducted. If the candidate reference pixels provided by CTIP have a SAD value less than that provided by OSIP, which corresponds to CTIP having higher correlation with the original reference pixels than that of OSIP, the MB will be recorded as a "CTIP preferred MB". Videos with different content features are evaluated. Table 4-2 lists the evaluation results for Intra_16×16 and Intra_4×4 of one sequence with 300 frames as an instance. As shown in the Table 4-2, within 14960 Intra_16x16 MBs, 77.8% of them are "CTIP preferred MB". Also, more than 79% of Intra_4x4 MBs prefer CTIP. Consequently, CTIP is finally selected as Intra prediction method when the reference pixels are outside of DPA.

Table 4-2 Pixel correlation comparison of CTIP and OSIP

Intra mode	Intra_16×16	Intra_4×4	Total
No. of MBs	14960	1739	16699
No. of CTIP preferred MBs.	11636	1381	13017
Ratio of CTIP preferred MBs	77.8%	79.4%	78.0%

4.5 Analysis of the Buffer Macroblocks within the Decoded Partial Area

To decide the proper number of Buffer MBs within DPA, some experiments have been done. Because the Buffer MBs have to co-work with the DPA adaptation and RBR that proposed in the previous sections, different Buffer MBs number are tested together with DPA adaptation and RBR. A typical example, the 120^{th} frame of sequence Sheriff is selected for the experiments with an RFD of CIF size. Table 4-3 shows the PSNR values of the RFD under different number of Buffer MBs. Comparing No. 4 and No. 5, 1 Buffer MBs improves 3.17dB PSNR. Then, the PSNR is increased by 0.33dB and 0.01dB for No. 6 and No. 7 comparing to its last one respectively. For No. 7, the PSNR of RFD is the same to the one of fully decoding solution. Thus in No. 8, there is no more PSNR increase. The visual quality comparison is also shown in Figure 4-11. Figure 4-11 (a) shows the frame of Approach No. 1 in Table 4-3. The green part indicates the loss of the reference pixels. Figure 4-11 (b) shows the frame of Approach No. 2 in Table 4-3. The area of loss is less than that in Figure 4-11 (a). Figure 4-11 (c) shows the frame of Approach No. 3 in Table 4-3. The visual quality is much better than previous 2 approaches, but there are still some errors along the edge of the frame. Figure 4-11 (d) shows the frame of Approach No. 7 in Table 4-3. Among all the previous tests, the visual quality of Figure 4-11 (d) is the best. Setting the number of Buffer MBs as 3 is a good tradeoff between good visual quality and low computational complexity. It is used for all the test sequences in this work.

Table 4-3 PSNR comparison of different number of Buffer MBs

No.	Approaches	PSNR (dB)
1	0 Buffer MBs	30.95
2	DPA Adaptation with 0 Buffer MBs	33.40
3	RBR with 0 Buffer MBs	35.27
4	DPA Adaptation and RBR with 0 Buffer MBs	35.58
5	DPA Adaptation and RBR with 1 Buffer MBs	38.75
6	DPA Adaptation and RBR with 2 Buffer MBs	39.08
7	DPA Adaptation and RBR with 3 Buffer MBs	39.09
8	DPA Adaptation and RBR with 4 Buffer MBs	39.09

(a) 0 buffer MBs

(b) DPA Adaptation with 0 Buffer MBs

(c) RBR with 0 Buffer MBs

(d) DPA Adaptation and RBR with 3 Buffer MBs

Figure 4-11 Visual quality of different number of Buffer MBs

4.6 Simulation Results and Discussions

4.6.1 Simulation Parameters

The proposed partial decoding scheme has been implemented into H.264/AVC reference software JM 14.2 [22]. Seven video sequences, including four 720p (1280×720) and three 1080p (1920×1080), which correspond to different content features, are tested. The encoding parameters are: main profile, group of pictures (GoO) as IPPP…IPPP… with 30 Intra period, 2 reference frames, deblocking filter on, three kinds of QP values (I-frame/P-frame): 27/28, 31/32, 35/36. The decoded frames and the pre-defined OOI of each sequence are listed in Table 4-4. There are four different RFD sizes: QCIF (176×144), CIF (352×288), VGA (640×480) and WVGA (800×480). Each sequence is tested at each RFD size under the three different QP values.

Table 4-4 Simulation parameters of sequences

Index	Sequences	Frames	OOI
I.	City (720p)	0 to 249	(560, 464), (740, 560)
II.	Sheriff (720p)	0to 249	(656, 360), (784, 432)
III.	Shields_ter (720p)	3 to 302	(700, 432), (770, 640)
IV.	Shuttle_Start (720p)	0 to 249	(656, 336), (680, 448)
V.	Crowd _run (1080p)	0 to 249	(1440, 560), (1600, 640)
VI.	Crowd _run (1080p)	0 to 249	(464, 320), (640, 512)
VII.	Old_ town_ cross (1080p)	0 to 249	(912, 464), (976, 592)

4.6.2 Simulation Results

The simulation results are listed in Table 4-5. The OOI location for each sequence is defined in Table 4-4. For a certain test case, the OOI is the same for all QP values and RFD sizes. The #V and #VI test case use the same sequence for experiments, but their OOI are different. Because of the difference of OOI, the movements of foreground and background are different, and their results are also very different, as shown in the results in Table 4-5. All the results are compared with original fully decoding. The PSNR between the RFD of partial decoding and the co-located region in the original video source before encoding is calculated. The PSNR between the RFD co-located region in the fully decoded sequence and the co-located region in the original video source is also calculated. Then, the ΔPSNR is calculated by the difference of the two PSNR values. Thus the ΔPSNR represents the quality loss in RFD comparing to the co-located region of the fully decoded sequence. TRR represents the decoding time reduction ratio between partial decoding and fully decoding. The decoding time of partial decoding includes the time of partial decoding and the overhead introduced by the proposed scheme. The decoding time of fully decoding is obtained from the original JM 14.2. The DPA ratio calculates the percentage of all the really decoded area in the whole sequences. The listed 7 sequences are tested under 3 different kinds of QP values for 4 different RFD size (QCIF, CIF, 720p, 1080p) respectfully. Thus, finally, there are totally 84 sets of results including the ΔPSNR, TRR, and DPA ratio that listed in Table 4-5 (a) and (b). From the simulation results, we can see that, on average, the proposed partial decoding scheme achieves 50.16% TRR with a negligible PSNR drop equals to 0.09 dB.

Table 4-5(a) Simulation results

Sequence Index	QP (I/P)	RFD size					
		QCIF			CIF		
		ΔPSNR (dB)	TRR (%)	DPA Ratio (%)	ΔPSNR (dB)	TRR (%)	DPA Ratio (%)
I.	27/28	0.00	65.79	12.44	-0.01	58.47	24.90
	31/32	0.00	65.16	12.49	0.00	58.85	24.62
	35/36	0.00	66.01	12.47	0.00	55.33	24.63
II.	27/28	-0.03	65.91	11.49	0.00	59.10	23.45
	31/32	0.00	61.26	11.64	0.00	53.35	23.37
	35/36	0.00	55.64	11.59	0.00	49.67	23.18
III.	27/28	-0.43	59.70	13.95	-0.45	49.25	26.06
	31/32	-0.22	55.12	13.85	-0.43	47.66	26.09
	35/36	-0.13	54.69	13.89	-0.26	47.08	26.01
IV.	27/28	0.00	54.51	11.64	0.00	48.50	23.52
	35/36	0.00	55.57	11.65	0.00	50.01	23.36
	37/38	0.00	57.14	11.63	0.00	49.88	23.52
V.	27/28	0.00	63.38	7.28	0.00	59.45	12.62
	31/32	0.00	66.33	7.30	-0.01	61.79	12.62
	35/36	0.00	67.21	7.31	-0.01	62.63	12.61
VI.	27/28	0.00	61.18	7.11	0.00	58.70	12.32
	31/32	0.00	60.89	7.16	0.00	57.78	12.36
	35/36	0.00	60.32	7.19	0.00	56.87	12.41
VII.	27/28	0.00	63.59	7.34	0.00	61.17	12.75
	31/32	0.00	67.12	7.40	0.00	64.08	12.81
	35/36	0.00	67.81	7.39	0.00	64.81	12.81
Average		-0.04	61.63	10.20	-0.06	55.93	19.33

Table 4-5(b) Simulation results

Sequence Index	QP (I/P)	RFD size					
		VGA			WVGA		
		ΔPSNR (dB)	TRR (%)	DPA Ratio (%)	ΔPSNR (dB)	TRR (%)	DPA Ratio (%)
I.	27/28	-0.35	42.51	51.14	-0.38	38.09	60.92
	31/32	-0.21	42.38	51.03	-0.24	38.69	60.81
	35/36	-0.22	44.76	50.91	-0.23	39.11	60.69
II.	27/28	0.00	40.89	50.81	0.00	34.09	60.72
	31/32	0.00	33.52	50.69	0.00	27.86	60.60
	35/36	0.00	32.25	50.41	0.00	27.03	60.32
III.	27/28	-0.58	36.18	54.64	-0.60	31.05	63.93
	31/32	-0.47	28.51	54.68	-0.47	24.59	63.97
	35/36	-0.30	29.86	54.56	-0.33	24.08	63.86
IV.	27/28	0.00	33.41	50.93	0.00	26.83	61.00
	35/36	0.00	36.13	50.63	-0.01	32.10	60.64
	37/38	0.00	40.86	50.93	0.00	37.24	61.00
V.	27/28	0.00	50.84	24.45	-0.38	48.09	28.32
	31/32	0.00	52.91	24.45	-0.26	50.06	28.32
	35/36	-0.01	53.06	24.33	-0.16	49.95	28.24
VI.	27/28	0.00	50.58	24.38	0.00	47.63	28.75
	31/32	0.00	50.08	24.43	0.00	47.39	28.80
	35/36	0.00	49.11	24.50	0.00	46.20	28.87
VII.	27/28	-0.01	53.38	25.03	0.00	51.51	29.46
	31/32	0.00	56.36	25.12	0.00	54.19	29.55
	35/36	0.00	57.01	25.12	0.00	54.14	29.55
Average		-0.10	43.55	40.15	-0.15	39.52	47.54

4.6.3 Visual Quality Comparison

About the visual quality in RFD, the Table 4-5 (a) and (b) show that, for most of the sequences, there is no PSNR drop in the RFD. For the sequences with PSNR loss, the error pixels are not obvious, and locate on the background while the OOI region is still clear. Here gives two examples amount the tested sequencese to show the visual quality of the partial decoding results.

The first example is a lossy case of sequence III. Figure 4-12 shows the visual quality of sequence III at QP 27/28 with RFD size of VGA, which has a large PSNR drop among the results. Figure 4-12 (a) shows the 1st frame of the fully decoding result. Figure 4-12 (b) shows the 1st frame of the partial decoding result. Figure 4-12 (c) shows the RFD of the last frame of a Intra period, which has the worst visual quality in all the RFD frames. It is observed that, even in the worst case, there is no loss in OOI, and the lossy MBs are very limited.

The second example is a lossless case of sequence VII. Figure 4-13 (a) and Figure 4-13 (b) show the 1st frame of the fully decoding result and partial decoding result respectivelly. Figure 4-13 (c) shows the visual quality of sequence VII at QP 27/28 with RFD size of CIF, and there is no loss in the whole RFD sequence either. This picture also show that, the tracking method also works well for 120 frames.

(a) The 1st frame of fully decoded sequence

(b) The 1st frame of RFD sequence

(c) The 30th frame of RFD sequence

Figure 4-12 Visual quality of sequence III

(a) The 1st frame of fully decoded sequence

(b) The 1st frame of RFD sequence

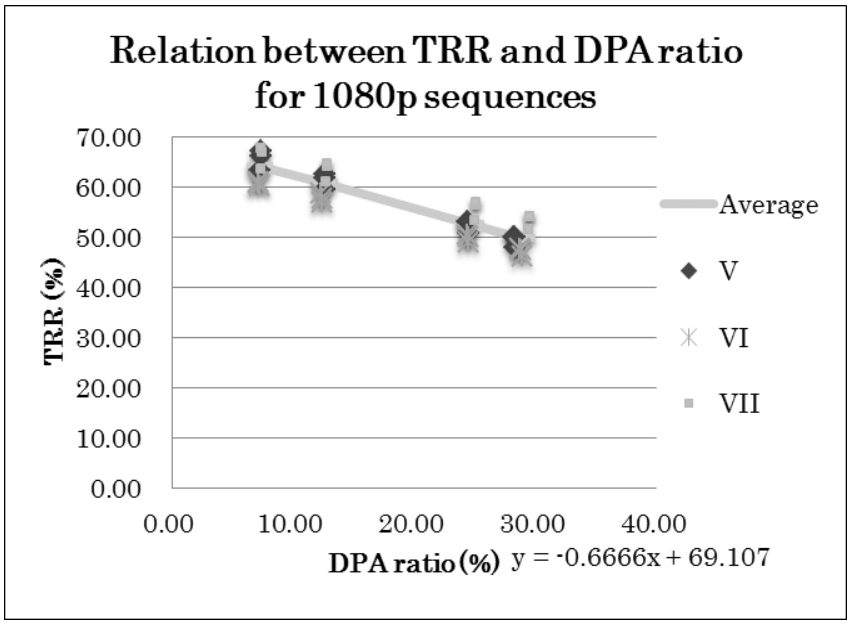
(c) The 120th frame of RFD sequence

Figure 4-13 Visual quality of sequence VII

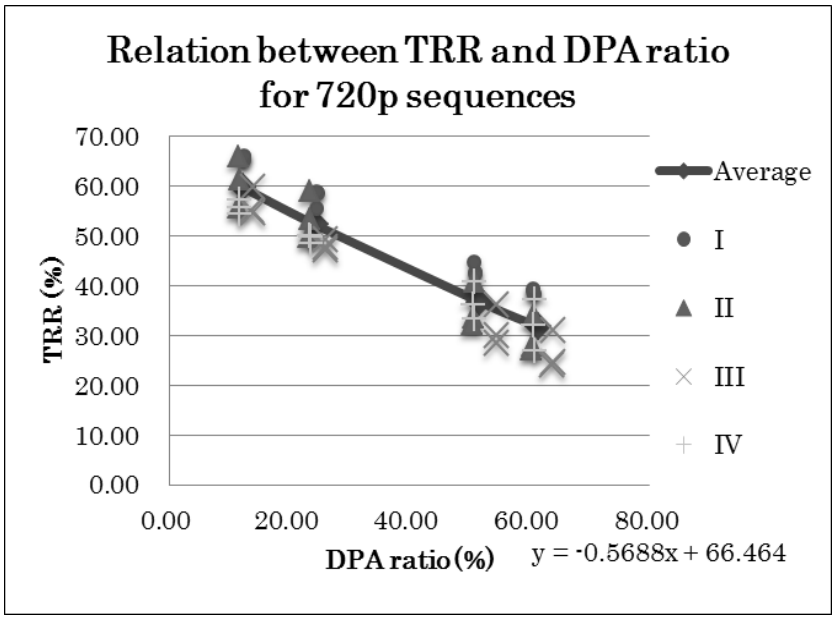
4.6.4 Relation between Time Reduction and Decoded Partial Area

On average, the proposed partial decoding scheme achieves 50.16% TRR with a negligible PSNR drop equals to 0.09 dB. From the simulation results, it can be observed that the TRR drops along with the increase in the RFD size. When the RFD size increases, the DPA ratio is also increased, and the TRR is decreased.

To further investegate the relation between the time reduction ratio and DPA ratio, the data is arranged into figures. Figure 4-14 (a) and (b) shows the relation between TRR and DPA ratio for 720p and 1080p sequences respectively. From the figures, we can see that, the TRR has the approximate linear relation with DPA ratio. When the DPA ratio is 0 which means that no video contents are actually decoded, the TRR is able to achieve its maximum value of 66.464% and 69.107% in 720p and 1080p video sequences respectively. Because the entropy decoding has to be fully conducted in the proposed partial decoding scheme, and the entropy decoding consumes about 1/3 computational cost in the whole decoding process (fully decoding). The limitation of proposed partial decoding scheme is about 2/3 time reduction.



(a) 720p sequences



(b) 1080p sequences

Figure 4-14 Relation between TRR and DPA

4.7 Conclusion

In this chapter, an encoder-unconstrained partial decoding scheme for H.264/AVC is proposed to achieve low-complexity and good visual quality of decoding and display. The proposed scheme only decodes the user defined Object of Interest (OOI) related region. The compression domain tracking method is utilized to ensure that the OOI locates in the center of the display area. To reduce the error propagation at the boundary of the displayed region, an decoded partial area (DPA) adaptation method is proposed to assign a buffer area to be decoded around the region for display (RFD). The reference block relocation (RBR) and the co-located temporal Intra prediction (CTIP) are proposed to improve the visual quality by assigning the blocks/pixels which refer to an undecoded region to the decoded region. The proposed OOI oriented partial decoding scheme is able to achieve better visual quality and lower power consumption. The simulation results show that the proposed partial decoding scheme achieves an average 50.16% decoding time reduction with only 0.09dB PSNR drop. It means lower computational complexity and power consumption comparing to fully decoding, with negligible objective quality loss comparing to the co-located region of fully decoding. The visual quality in RFD is enhanced compared to the conventional FDDS solution. This proposal is especially useful for the displaying of HD video on portable devices in which battery life is a crucial factor. Also, the encoder-unconstrained solution is a necessary condition of real-time broadcasting.

In this chapter, the partial decoding scheme is designed to decoding the sequences with “IPPP…IPPP…” frame structure. But it is possible to adapt B-frame by improving and extending current proposals. The B-frame support of partial

decoding scheme can be one of the future works. Comparing to current “P-frame” partial decoding, the “B-frame” partial decoding is expected to achieve more time reduction, and the bi-direction motion vectors provided by B-frame are more helpful for OOI tracking and DPA adaption.

Besides reducing the power consumption purely at the decoder side, reducing the data that need to be transferred in the meanwhile is also very important, especially for video streaming applications. For the proposed system, it is possible to add a feedback scheme, like that in [34], between the encoder side and the decoder side. So that, the user defined OOI can be sent to the encoder by the network and the encoder can only encode and transmit the DPA related bit-stream to reduce the transmission rate. However, how to smoothly adapt the decoder to the bit-stream variation and how to adapt the encoder to the user requirements in real-time could be very challenging items, which need further investigations.

5 Envelope Detection Based Workload Prediction

An envelope detection based workload prediction is proposed in this chapter. The workload estimation is based on the envelope of the difference value between the previous adjacent actual-workloads, and the envelope is obtained by an absolute value and low pass filter method. To further improve the performance of the envelope detection, the negative truncation is introduced by ignoring the negative difference.

This workload prediction method is proposed for the partial decoding scheme and works together with the Dynamic Voltage Frequency Scaling (DVFS) to reduce the power consumption of video decoder. It provides accurate estimation results with only 0.66% deadline missing rate, which is much lower than the previous methods.

The power consumption results on the evaluation board show that, with the proposed envelope detection based workload prediction and DVFS, the power reduction achieves about 41.65%, the energy reduction is about 10.16%. The energy reduction of the partial decoding together with proposed workload prediction and DVFS is about 61.15% compare to the fully decoding without DVFS.

By further investigating and improving the DVFS, it is possible to reduce more power consumption and energy cost. The proposed envelope detection based workload prediction algorithm also can be used for general workload prediction.

5.1 Background and Related Work

Nowadays, the video contents playback on portable devices is normally and widely used. Since the performance always trades off with power consumption and battery life is also a crucial factor, smooth video decoding with lower power consumption becomes a very important topic.

For general tasks, the Dynamic Voltage Frequency Scaling (DVFS) shows its effectiveness on the reduction of power consumption [35]. The model of DVFS is showed in the following equation [36]:

$$\mathrm{P} \propto C_{eff} V_{DD}^2 f_{CLK}$$

Equation 5-1

where, the P is the power consumption, the C_{eff} is the capacitance, the V_{DD} is the input voltage and the f_{CLK} is the clock frequency. In order to adjust the clock frequency/input voltage and minimize idle time dynamically without performance hit, an accurate workload prediction is necessary and becomes the most important issue for the DVFS.

Many methods have been proposed for the workload prediction. The past value (PAST) based workload prediction is using the last actual-workload as the predicted value [37] [38]. The moving average (MA) method predicts the workload by the average value of the actual-workload during certain intervals in the past [37], and the exponential weighted average (EWA) improves the MA by introducing weighted summation [39]. The Hilbert transformation based workload prediction [40] is proposed for the video encoding workload estimation by calculating the upper

envelope of the actual encoding signals. The bit-stream analysis based workload prediction [41] is used for MPEG video decoding.

For the low computational cost video decoding, a partial decoding scheme for video decoder is already proposed [42]. The workload profile of the partial decoding model is different from the generally full decoding. In the partial decoding scheme, all the I-frames are fully decoded. For the P-frames, the entropy decoding of the bit-stream are fully conducted, the inverse quantization, inverse transformation, motion compensation, and etc. of the residual data are conducted only for the macroblocks belonging to the Decoded Partial Area (DPA). The DPA is larger than the pre-defined Region-for-Display (RFD), and it is dynamically updated according to the video contents. That means, the workloads are highly affected by the DPA. Thus, previous workload prediction methods [38] [37] [39] [40] [41] perform not well for the partial decoding scheme. To achieve lower power consumption for partial decoding scheme, an effective workload prediction method with DVFS is necessary.

In this chapter, an envelope detection based workload prediction for partial decoding scheme is proposed. The workload estimation is based on the envelope of the difference value between the previous adjacent actual-workloads, and the envelope is obtained by an absolute value and low pass filter method. To further improve the performance, the negative truncation (negative trun.) is introduced by ignoring the negative difference. Comparing to the Hilbert transform based method [40], which is also an envelope detection based approach, the proposed method provides better performance, and low computational complexity.

5.2 Partial Decoding Workload Evaluation

5.2.1 Workload Definition for Partial Decoding Scheme

For the partial decoding scheme, the workload prediction and DVFS is only conducted for the P-frames. For the simulation on PC, the workload is directly obtained from the decoding time. For the experiments on the RP2 board, the workload of one frame is calculated by the decoding time of this frame multiply the CPU frequency. Take the case of simulation on PC as the example, the workload of *i*-th P-frame within time slot *n* is defined as:

$$w_{n,i} = T_{n,i}/T_M$$

Equation 5-2

where $T_{n,i}$ is the decoding time of *i*-th P-frame; T_M is the worst-case decoding time of the system. The predicted workload of time slot *n* is defined as:

$$\bar{w}'_n = \bar{T}'_n/T_M$$

Equation 5-3

where $\bar{T}'_n$ is the average value of $T_{n,i}$ in the following time slot.

5.2.2 Evaluation by the performance constraint deviation (PCD)

In this paper, the prediction performance is evaluated by the Performance Constraint Deviation (PCD) [40]:

$$D_k = \sum_{n=kf_s/N}^{(k+1)f_s/N-1} (\bar{w}'_n - \bar{w}_n), \qquad k \in [1, f_T/f_s]$$

Equation 5-4

where $\bar{w}'_n$ is the predicted workload, $\bar{w}_n$ is the actual workload, f_s is the performance constraint, and the f_T is the total frame number.

The PCD is illustrated in Figure 5-1.

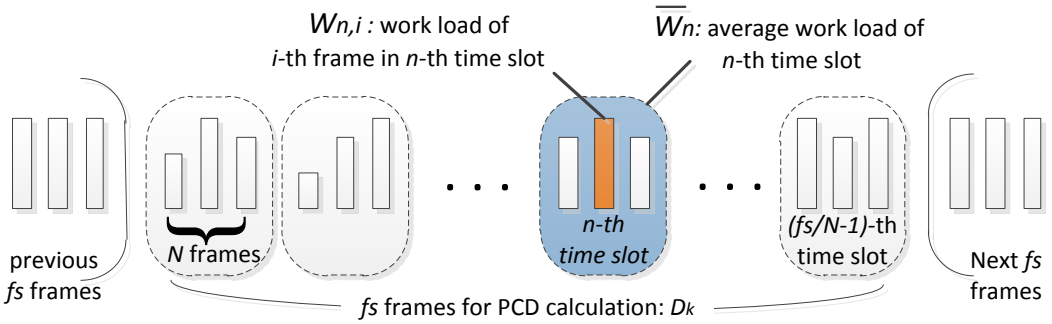
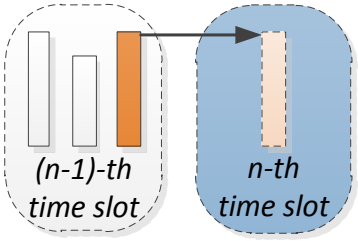
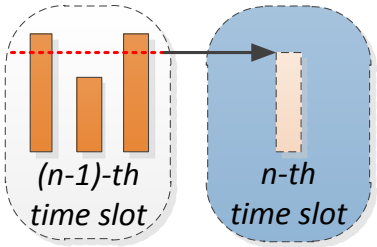
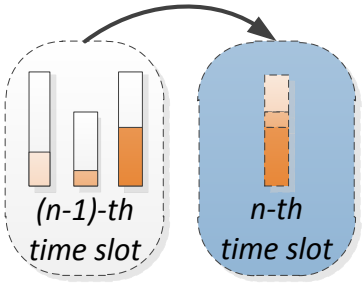
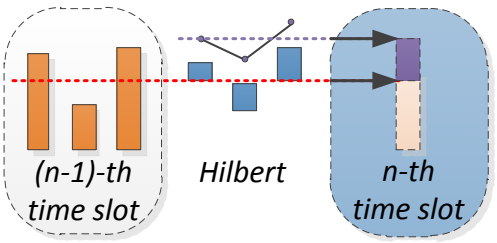


Figure 5-1 The definition of PCD

In this paper, N=3 and f_s=12. The smaller deviation between D_k and 0 presents higher accuracy in estimation, while D_k less than zero indicate the performance hit, which means the deadline missing in video decoding. The performance hit should be avoided in the workload prediction.

Basing on the definition of partial decoding workload, four different workload prediction methods listed in Table 5-1 are evaluated by PCD. The results are shown in the Figure 5-2. All of these methods cannot avoid the performance hit. The Hilbert transform based prediction method shows relative better performance, but it still have 15.79% deadline missing rate (DMR).

Table 5-1 Previous workload prediction methods

Methods	Illustrations
Past value based prediction (PAST) $\bar{w}'_n = w_{n-1,N-1}$	
Moving Average based prediction (MA) $\bar{w}'_n = \frac{1}{L}\sum_{i=N-L}^{N-1} w_{n-1,i}$	
Exponential weighted average prediction (EWA) $\bar{w}'_n = \frac{1}{2}w_{n-1,N-1} + \frac{1}{4}w_{n-1,N-2} + \frac{1}{4}w_{n-1,N-3}$	
Hilbert Transform based prediction (Hilbert) $G(i) = w_{n-1,i} - \bar{w}_{n-1}$ $A(i) = \lvert G(i) + j \cdot H(G(i)) \rvert$ $\bar{w}'_n = \frac{1}{2} \cdot \sum_{i=0}^{N-1} A(i) + \bar{w}_{n-1}$	 $g(t) * \frac{1}{\pi t} = F^{-1}\left(F\left(g(t)\right) \bullet \left(-j\,\mathrm{sgn}(f)\right)\right)$

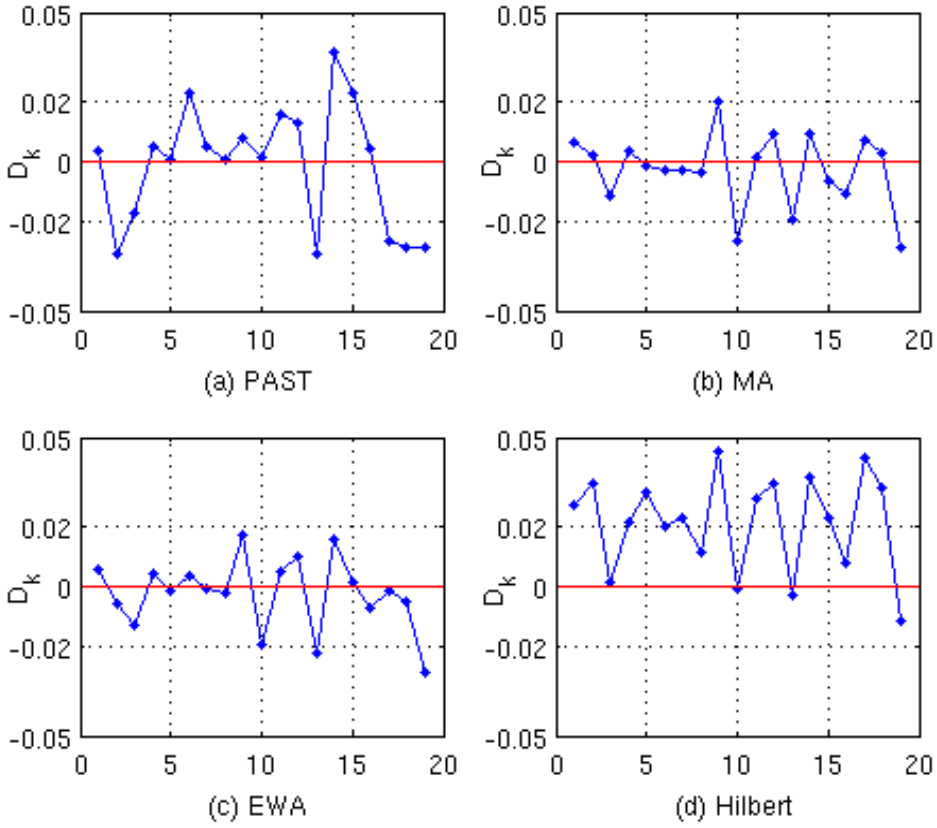


Figure 5-2 PCD of the previous works (Shuttle_Start, RFD: CIF)

5.3 Envelope Based Workload Prediction

5.3.1 Analysis of Partial Decoding Workload

According to the evaluation in the last section, previous workload prediction methods perform not well for partial decoding workload model. The first three methods cannot avoid the performance hit, especially when the signal keeps increasing. The Hilbert transform based prediction method shows relative better performance because it tracks the upper envelope of the given signal, which makes the predicted workload statistically a little larger than the actual workload. However,

the Hilbert transform is computational cost, especially when the window length is large.

5.3.2 Envelope Based Workload Prediction

The proposed workload prediction is also based on the envelope detection method. Ideally, the envelope detection can be achieved by an absolute operation with a following low pass filter, which is similar to the diode and RC filter unsynchronized envelope detector circuit. Basing on this concept, the envelope of input signal *v(t)* is defined as:

$$m(i) = \frac{1}{L}\sum_{k=0}^{L-1}|v(i-k)|$$

Equation 5-5

where *L* is the window length for the low pass filter.

To improve the prediction performance, the workload difference is predicted instead of the workload itself, since it contains the information of signal variation. The workload difference of the *i*-th frame in the *n*-th time slot is defined as:

$$V_n(i) = \begin{cases} w_{n,0} - w_{n-1,N-1}, & i = 0 \\ w_{n,i} - w_{n,i-1}, & i \geq 1 \end{cases}$$

Equation 5-6

And the envelope of the workload difference is obtained by the envelope detector according to Eq. (6):

$$M_n(i) = \begin{cases} \frac{1}{N}\left(\sum_{k=i+1}^{N-1}|V_{n-1}(k)| + \sum_{k=0}^{i}|V_n(k)|\right), & 0 \le i < N-1 \\ \frac{1}{N}\sum_{k=0}^{i}|V_n(k)|, & i = N-1 \end{cases}$$

Equation 5-7

The averaged $M_n(i)$ is calculated to predict the variation in workload for the next time slot. Finally, the averaged $M_n(i)$ is added back to the average workload in the previous time slot to generate the predicted workload for the current time slot n, $\bar{w}'_n$ as:

$$\bar{w}'_n = \frac{1}{N}\sum_{i=0}^{N-1} M_{n-1}(i) + \bar{w}_{n-1}$$

Equation 5-8

The proposed method (w/o negative trun.) is illustrated in Figure 5-3. The workload prediction and PCD results of the proposed method (w/o negative trun.) are shown in Figure 5-4 and Figure 5-5 respectively.

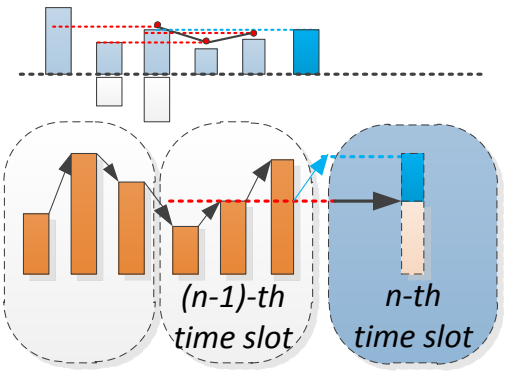


Figure 5-3 The Illustration of the proposed envelop detection based workload prediction method (w/o negative trun.)

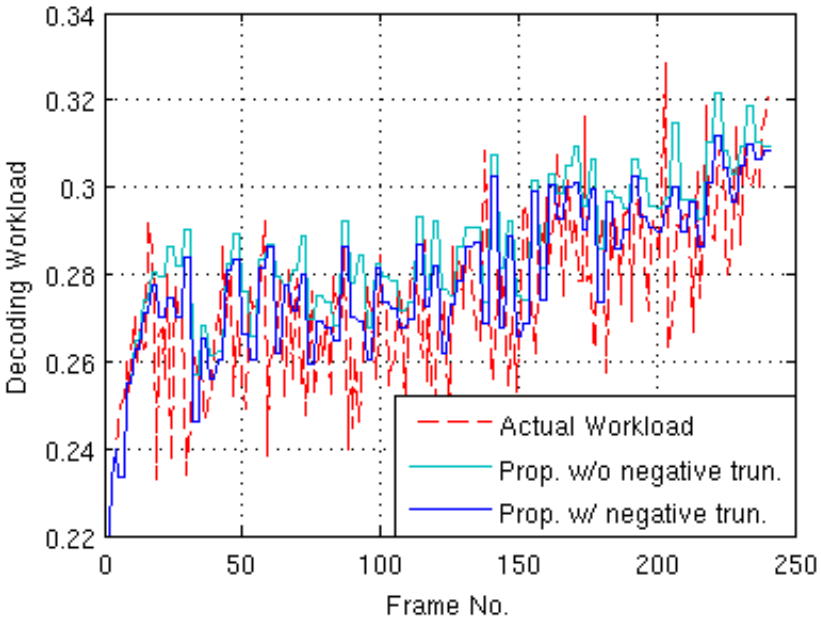


Figure 5-4 The workload prediction results of proposed envelop detection based workload prediction method

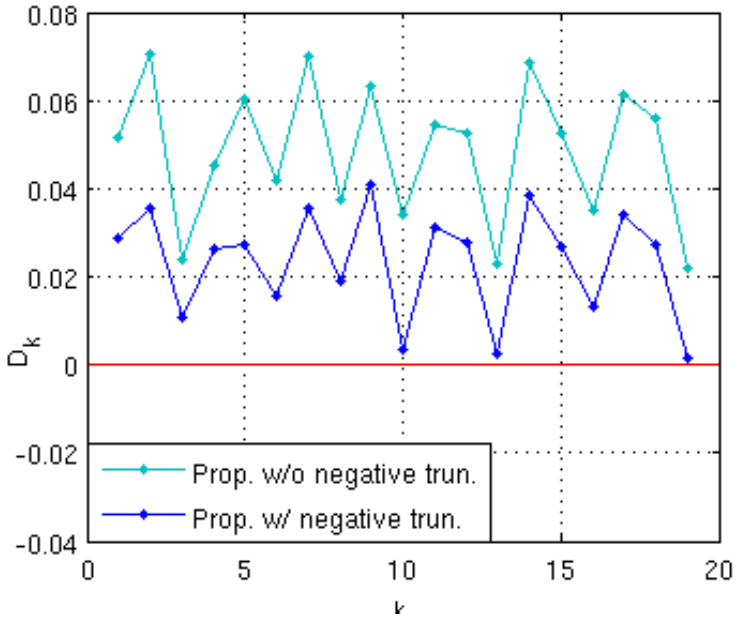


Figure 5-5 The PCD results of proposed envelop detection based workload prediction method

5.3.3 Negative Truncation Based Improvements

From Figure 5-4 and Figure 5-5, we can see that, the proposed method results in a large margin comparing to the actual-workload, because the negative difference values are all absolute operated. If the actual workload experiences a continuously falling trend, the prediction result would be opposite to the actual trend. Although this effect does not cause a performance hit, it costs extra power consumption. To solve this problem, we improve the proposed prediction method by introducing a negative value truncation. Instead of taking the absolute operation of negative values, the negative values are truncated in the calculation. Thus, in the improved method (w/ negative trun.), the definition of $M_n(i)$ is updated in the following:

$$M_n(i) = \begin{cases} \frac{1}{2N}\sum_{k=i+1}^{N-1}(|V_{n-1}(k)| + V_{n-1}(k)) + \frac{1}{2N}\sum_{k=0}^{i}(|V_n(k)| + V_n(k)), & 0 \le i < N-1 \\ \frac{1}{2N}\sum_{k=0}^{i}(|V_n(k)| + V_n(k)), & i = N-1 \end{cases}$$

Equation 5-9

The improved proposal (w/ negative trun.) is illustrated in Figure 5-6. The workload prediction and PCD results of the improved method (w/ negative trun.) are shown in Figure 5-4 and Figure 5-5 respectively. From Figure 5-5, it can be seen that the improved approach with the negative truncation shows smaller D_k without performance hit, namely, better performance, and is conducted in the workload estimation for partial decoding scheme.

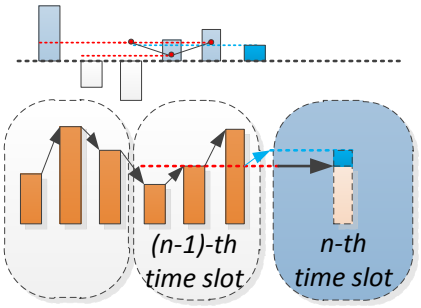
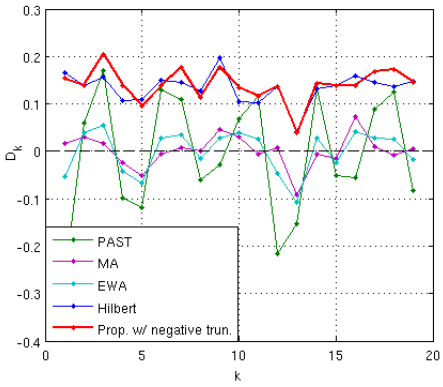


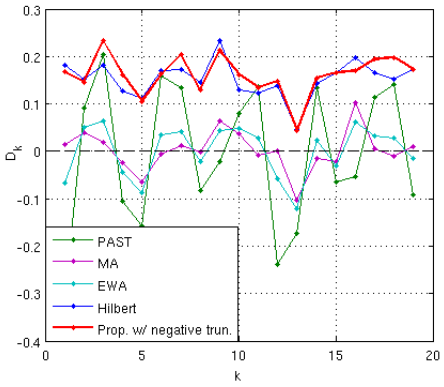
Figure 5-6 The illustration of the proposed envelop detection based workload prediction method (w/ negative trun.)

5.4 Simulation Results

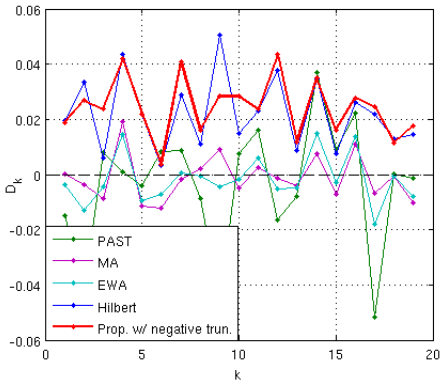
Four 720p sequences with different characteristics with RFD size QCIF and CIF for partial decoding are tested respectively. The PCD comparison is shown in Figure 5-7. The comparison of the DMR is listed in Table 5-2. From the simulation results, we can see that, the proposed workload prediction has the best performance comparing to all the previous works. The DMR is much less than others.



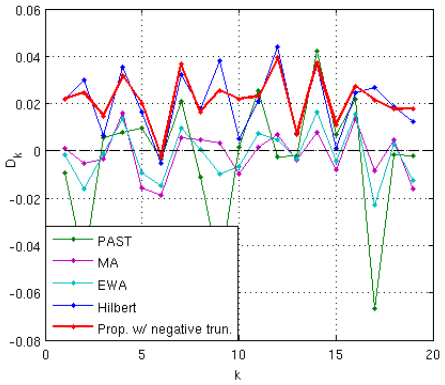
(a) City (RFD: QCIF)



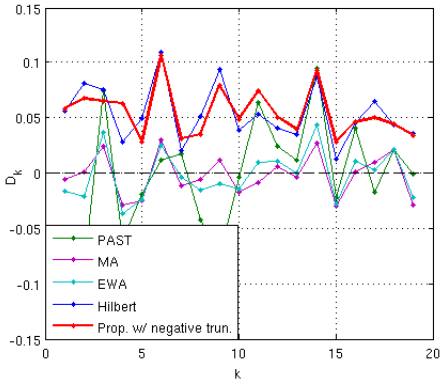
(b) City (RFD: CIF)



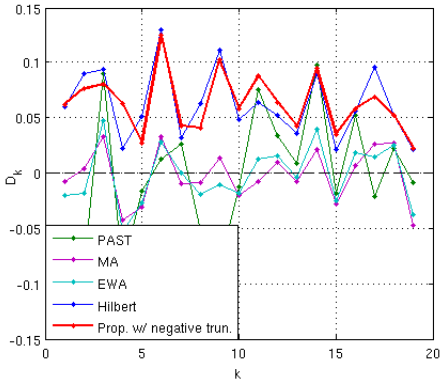
(c) Sheriff (RFD: QCIF)



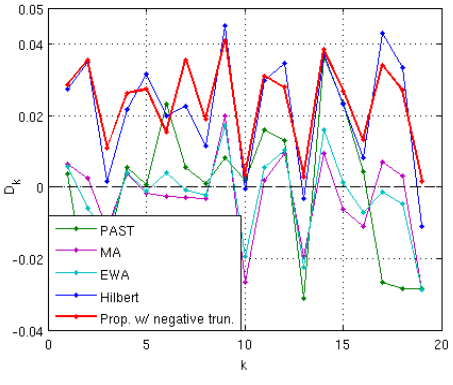
(d) Sheriff (RFD: CIF)



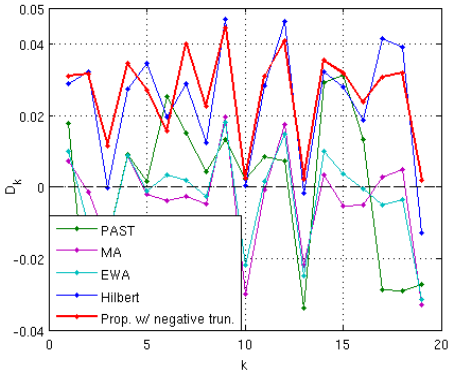
(e) Shields_ter (RFD: QCIF)



(f) Shields_ter (RFD: CIF)



(g) Shuttle_Start (RFD: QCIF)



(h) Shuttle_Start (RFD: CIF)

Figure 5-7 Comparison of PCD

Table 5-2 The comparison of DMR (%)

Seq. (720p)	RFD	PAST	MA	EWA	Hilbert	Prop.
City	QCIF	52.63	42.11	42.11	0	0
	CIF	52.63	47.37	42.11	0	0
Sheriff	QCIF	47.37	63.16	73.68	0	0
	CIF	52.63	47.37	57.89	5.26	5.26
Shields_ter	QCIF	52.63	52.63	52.63	0	0
	CIF	52.63	52.63	57.89	0	0
Shuttle_Start	QCIF	31.58	52.63	57.89	15.79	0
	CIF	31.58	63.16	52.63	15.79	0
AVERAGE		46.71	52.63	54.605	4.61	0.66

5.5 Power Evaluation

The proposed envelope detection based workload prediction (EDWP) with the DVFS used in [40] together with the partial decoding scheme proposed in [42] are all implemented in the H.264 reference software JM14.2 [22], and it is run on the Hitachi evaluation board with a Renesas SuperH SH-4A processor [43] [44]. The power consumption of the CPU can be measured in real time. The working frequency of the processor is scalable among 600MHz, 300MHz, 150MHz and

75MHz, and the input voltage is automatically scaled corresponding to the frequency with 1.4V, 1.2V, 1.0V and 1.0V.

Firstly, the fully decoding of original JM, and the partial decoding with proposed EDWP and DVFS are evaluated on the evaluation board respectively. The results of frame rate comparison are listed in the Table 5-3. The results show that, the frame rate of partial decoding is always larger than the original fully decoding, which means there is no deadline missing.

Then, the power consumption of partial decoding with the proposed EDWP and DVFS are tested. The power reduction is compared to the partial decoding w/o workload prediction and DVFS, the comparison results are listed in the Table 5-4. For all the tested results, the average power reduction is about to 41.65%. Since the frame rate is still faster than the fully decoding, the power reduction is able to be further improved by optimize the DVFS strategy.

Table 5-3 Frame rate (fps) comparison

Seq. (720p)	Original Fully Decoding (720p)	Partial Decoding (QCIF)	Partial Decoding (CIF)
City	0.749	0.964	0.837
Sheriff	0.683	0.984	0.862
Shields_ter	0.733	0.988	1.074
Shuttle_start	1.020	1.670	1.344

Table 5-4 Power reduction performance of EDWP and DVFS

Seq. (720p)	RFD: QCIF	RFD: CIF
City	48.82%	45.57%
Sheriff	43.75%	41.64%
Shields_ter	43.37%	23.53%
Shuttle_start	43.28%	43.28%
AVERAGE	44.80%	38.50%

Table 5-5 Energy saving performance of EDWP and DVFS

Seq. (720p)	v.s. Partial decoding w/o EDWP and DVFS		v.s. Original Fully Decoding	
	RFD: QCIF	RFD: CIF	RFD: QCIF	RFD: CIF
City	9.21%	6.50%	60.23%	51.30%
Sheriff	11.56%	11.51%	60.96%	53.76%
Shields_ter	10.81%	5.30%	57.99%	47.81%
Shuttle_start	13.35%	13.01%	65.35%	56.95%
AVERAGE	10.16%		61.15%	

For further evaluation, the energy saving performance is compared in Table 5-5. The energy is obtained by the power consumption multiplies the decoding time. For the partial decoding scheme, the EDWP and DVFS provides extra 10.16% energy saving. Comparing to the original fully decoding, the whole scheme of partial decoding together with EDWP and DVFS achieves 61.15% energy saving.

5.6 Conclusion

An envelope detection based workload prediction has been proposed. The workload is estimated by calculating the envelope of the difference value between the previous adjacent actual workloads, and the envelope is obtained by the absolute value and low pass filter method. To further improve the performance, the negative truncation is introduced by ignoring the negative difference. The DMR of the proposed envelope detection based workload prediction is only 0.66% which is much lower than previous works. Moreover, the proposed envelop detection based workload prediction has lower computational complexity than the Hilbert transform based one. For partial decoding, with EDWP and DVFS, the power reduction achieves about 41.65%, energy saving is about 10.16. The partial decoding scheme together with the EDWP and DVFS, the total energy saving achieves 61.15% comparing to the original fully decoding. In the future work, by further investigating and improving the DVFS, it is possible to reduce more power consumption and energy cost. The proposed envelope detection based workload prediction algorithm also can be used for general workload prediction.

6 Conclusion and future work

6.1 Conclusion

This dissertation presents my research on the low complexity schemes for both H.264/AVC encoder and decoder. The motion correlation information, such as motion vectors and residual which are generated during the video compression process, is skillfully utilized in all the proposed schemes. These approaches avoid the extra fetching and analysis process of other features, and the motion correlation information also provides more underlying spatial and temporal relation of the video content which is very necessary and helpful for the motion feature analysis and optimization the coding process.

For video encoder, two kinds of fast inter mode decision algorithms is proposed to reduce the computational complexity of the motion estimation part in H.264/AVC encoder. Firstly, the motion adaptation based inter mode decision algorithm is proposed for H.264/AVC encoding based on motion information provided by motion vectors. The key idea is to reduce the candidate modes by the underlying relation between motions and macroblock partition types. Thus, a motion vector related evaluation approach with adaptive multilevel thresholds has been proposed. The simulation results show that the proposed method achieves 33.4% time reduction as average with minimal quality loss and less bit-rate increase. The proposed motion adaptation based inter mode decision algorithm performs better for the videos whose contents have many tiny and irregular movements. Secondly, the

residual feature based inter mode decision algorithm is proposed. Different from previous work (Wang's on ICME 2007), in which the positive and negative residual in a block is summed indiscriminately, the positive and negative residual are extracted separately in the proposed algorithm. Based on the extracted residual feature, the complexity and similarity are evaluated for the inter mode decision. According to the evaluation of similarity between different sub-blocks and the complexity of each sub-block, the most possible inter modes for current block is chosen to be conducted. In the worst case, the proposed whole scheme of inter mode decision algorithm only conducts 4 modes, which is much more effective than conducting all the 8 modes in conventional approach. The experimental results show that, compare to conventional JM, on average, the proposed algorithm achieves 57.98% and 55.72% time-saving for CIF and 720p sequences respectively, with equivalent 0.219db PSNR drop with 5.55% bit rate increase for CIF, and 0.128db PSNR drop with 3.53% bit rate increase for 720p.

For video decoder, a novel encoder-unconstrained user interactive partial video decoding scheme is proposed. The proposed scheme is used to solve the resolution mismatch problem when playing high resolution videos on relatively low resolution screens. The proposed method only decodes the object of interest (OOI) related region, which is defined by users. A simplified compression domain tracking method is utilized to ensure that the OOI locates in the center of the display area. The decoded partial area (DPA) adaptation, the reference block relocation (RBR) and co-located temporal Intra prediction (CTIP) methods are proposed to improve the visual quality for the DPA with low complexity. The simulation results show that, the proposed partial decoding scheme provides an average of 50.16% decoding

time reduction comparing to the fully decoding process. The displayed region also presents the original HD granularity. The proposed partial decoding scheme is especially useful for displaying HD video on the devices of which the battery life is a crucial factor. To further investigate the power reduction of the partial decoding scheme, an envelope based workload prediction is proposed and work together with the Dynamic Voltage Frequency Scaling (DVFS) to reduce the power consumption of the proposed partial decoding scheme. The experimental results on the evaluation board show that, with the proposed envelope detection based workload prediction and DVFS, the power reduction achieves about 41.65%, the energy reduction is about 10.16%. The energy reduction of the partial decoding together with proposed workload prediction and DVFS is about 61.15% compare to the original fully decoding. By further investigating and improving the DVFS, it is possible to reduce more power consumption and energy cost. The proposed envelope detection based workload prediction algorithm also can be used for general workload prediction.

6.2 Future work

For the inter mode decision topic, the residual feature based inter mode decision is possible to be improved by setting motion feature oriented thresholds to further improve the performance, but defining the detailed strategies needs more study and experiments. Moreover, the motion correlation based inter mode decision for the next generation video coding standard High Efficiency Video Coding (HEVC) is able to be a research topic.

For partial video decoding topic, the B-frame support is able to achieve by extending and improving current proposals and strategies. The lossless partial

decoding is able to achieve according to the pre-recoded macroblock dependency between frames through a pre-analysis of the bit-stream, but how to storage and access of the dependency file effectively is a problem need to be study. Also, reducing the data traffic between encoder and decoder for the OOI oriented codec solution is another important issue for the codec system.

For low power issue, further investigating and improving the DVFS is possible to achieve a better energy reduction result. Also, a suitable experimental platform is necessary.

References

[1] ISO/IEC 14496-10 and IUT-T Rec. H.264, "Advanced Video Coding," 2003.

[2] ISO/IEC 11172, "Information technology – coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," 1993, (MPEG-1).

[3] ISO/IEC 13818, "Information technology – generic coding of moving pictures and associated audio information," 1995, (MPEG - 2).

[4] ISO/IEC 14496-2, "Coding of audio-visual objects – Part 2: visual," 2001, (MPEG-4).

[5] ITU-T Rec. H.263, "Video coding for low bit rate communication, version 2," 1998.

[6] T. Wiegand., G. J. Sullivan., G. Bjontegaard., and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560-576, 2003.

[7] G. Sulliva, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," *Proc. SPIE-Applications of Digital Image Processing XXVII*, vol. 5558, pp. 454-474, Augest 2004.

[8] D. Marpe, T. Wiegand, and G. J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *Communications Magazine, IEEE*, vol. 44, no. 8, pp. 134-143, Augest 2006.

[9] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression-Video Coding for Next-generation Multimedia*.: Wiley.

[10] G. Sullivan. and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine, IEEE*, pp. 77-90, November 1998.

[11] H. Everett, "Generalize Lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, pp. 399-417, 1963.

[12] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-TVCEG, "Joint Model Number 1 (JM-1)," January 2002.

[13] D. Wu et al., "Fast intermode decision in H.264/AVC video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 7, pp. 953-958, July 2005.

[14] C. Kannangara and et al., "Low-complexity skip prediction for H.264 through Lagrangian cost estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 8, pp. 202–208, Feburary 2006.

[15] Byung-Gyu Kim, "Novel Inter-Mode Decision Algorithm Based on Macroblock (MB) Tracking for the P-Slice in H.264/AVC Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 2, pp. 273-279, Feburary 2008.

[16] X. Wang, J. Sun, Y. Liu, and R. Li, "Fast Mode Decision for H.264 Video Encoder Based on MB Motion Characteristic," *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 372-375, July 2007.

[17] Andy C. Yu, "Efficient block-size selection algorithm for inter-frame coding in H.264/MPEG-4 AVC," *Acoustics, Speech, and Signal Processing, 2004.*

Proceedings. (ICASSP '04). IEEE International Conference on, vol. 3, pp. 169-172, May 2004.

[18] Z. Wang, Q. Peng, and Y. Zeng, "Residual Texture Based Fast Block-Size Selection for Inter-Frame Coding in H.264/AVC," *Parallel and Distributed Computing, Applications and Technologies, 2005. PDCAT 2005. Sixth International Conference on*, pp. 853-855, December 2005.

[19] JVT-G050, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264-ISO/IEC 14496-10 AVC)," March 2003.

[20] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-Curves," *ITU SG16 Doc. VCEG-M33*, April 2001.

[21] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 12, pp. 1557-1561, December 2006.

[22] Joint Video Team (JVT) Reference Software. [Online]. Http://iphome.hhi.de/suehring/tml/download/

[23] Implementation Group, "A Computational Complexity Comparison of MPEG4 and JVT Codecs," *MPEG Doc. M8696, ISO/IEC JTC1/SC29/WG11*, 2002.

[24] ISG Group, "Memory Complexity Analysis of the AVC Codec JM 1.7.," *MPEG Doc. M8378, ISO/IEC JTC1/SC29/WG11*, 2002.

[25] C. Liu, T. R. Zhang, M. H. Wang X. Jin, and S. Goto, "Residual Analysis Based Fast Inter Mode Decision for H.264/AVC," *Intelligent Signal Processing and*

Communication Systems, 2009. ISPACS 2009. International Symposium on, pp. 339-343, December 2009.

[26] A. M. Tourapis, "Enhanced predictive zonal search for single and multiple frame motion estimation," *Proceedings of Visual Communications and Image Processing (VCIP) 2002*, pp. 1067-1079, January 2002.

[27] P. M. Alt, "Displays for Electronic Imaging," *Micro, IEEE*, vol. 18, no. 6, pp. 42-53, November/December 1998.

[28] Retina Display of the iPhone 4s on the official site of Apple Inc. [Online]. http://www.apple.com/iphone/features/retina-display.html

[29] M. Horowitz, A. joch, and F. Kossentini, "H.264/AVC baseline profile decoder complexity analysis ," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 704-716, July 2003.

[30] ITU-T; ISO/IEC JTC1; JVT-W201, "Joint Draft 10 of SVC Amendment," April 2007.

[31] H. Schwarz, D. Marpe, and and T. Wiegand Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol. 17, no. 9, pp. 1103-1120, September 2007.

[32] H. M. Nam, J. Y. Jeong, K. Y. Byun, J. O. Kim, and S. J. Ko, "A complexity scalable H.264 decoder with downsizing capability for mobile devices ," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 2, pp. 1025-1033, May 2010.

[33] P. Lambert et al., "A real-time content adaptation framework for exploiting ROI

scalability in H.264/AVC," *Innovative Computing, Information and Control, International Conference on*, vol. 4179, pp. 442-453, 2006.

[34] T. Takahashi, M.Sugano, and S. Sakazawa, "Fully-zoomable streaming method for three screen services (in Japanese)," *The Institute of Image Information and Television Engineers (ITE) Winter Annual Convention 2010*, November 2011.

[35] A. Sinha and A. P. Chandrakasan, "Dynamic voltage scheduling using adaptive filtering of workload traces," *VLSI Design, 2001. Fourteenth International Conference on*, pp. 211-226, January 2001.

[36] T. Burd and R. Broderson, "Processor design for portable systems," *Journal of VLSI Signal Processing Systems*, vol. 13, pp. 203–222, Auguest 1996.

[37] K. Govil, E. Chan, and H. Wasserman, "Comparing algorithms for dynamic speed-setting of a low power CPU," *in MobiCom '95 Proceedings of the 1st annual international conference on Mobile Computing and Networking*, pp. 13-25, 1995.

[38] D. Grunwald, P. Levis, K. I. Farkas, III C. B. Morrey, and M. Neufeld, "Policies for dynamic clock scheduling," *Proceedings of the 4th Symposium on Operating System Design & Implementation*, vol. 4, pp. 73-86, 2000.

[39] T. Pering, T. Burd, and R. Brodersen, "The simulation and evaluation of dynamic voltage scaling algorithms," *Low Power Electronics and Design, 1998. Proceedings. 1998 International Symposium on*, pp. 76-81, Auguest 1998.

[40] X. Jin and S. Goto, "Hilbert Transform based Workload Estimation for Low Power Surveillance Video Compression," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 31, no. 5, pp. 649-661, 2012.

[41] Y. Huang, V. A. Tran, and Y. Wang, "A workload prediction model for decoding MPEG video and its application to workload-scalable transcoding," *Proceedings of the 15th international conference on Multimedia (ACMMM)*, pp. 952-961, September 2007.

[42] C. Liu, X. Jin, T. Zhang, and S. Goto, "Encoder-unconstrained User Interactive Partial Decoding Scheme," *IEICE Trans. Fundamentals*, vol. E95-A, no. 8, p. in press, Augeust 2012.

[43] SuperH RISC engine Family Renesas Electronics. [Online]. http://www.renesas.com/products/mpumcu/superh/superh−landing.jsp

[44] M. Ito et al., "An 8640 MIPS SoC with independent power-off control of 8 CPUs and 8 RAMs by an automatic parallelizing compiler," *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International* , pp. 90-598, February 2008.

Publications

Academic Journals

[1] **Chen Liu**, Xin Jin, Tianruo Zhang, and Satoshi Goto, “Encoder-unconstrained User Interactive Partial Decoding Scheme,” IEICE Trans. Fundamentals, Vol.E95-A,No.8, pp.1288-1296, Aug. 2012.

[2] Tianruo Zhang, **Chen Liu**, Minghui Wang, and Satoshi Goto, "Multiple region-of-interest based H.264 encoder with a detection architecture in macroblock level pipelining," IEICE Trans. Electron, Vol.E94-C, No.4, Apr. 2011.

[3] Tianruo Zhang, Xin Jin, **Chen Liu**, Minghui Wang, and Satoshi Goto, "ROI Based Computational Complexity Reduction Scheme for H.264/AVC Encoder," The Journal of the Institute of Image Electronics Engineers of Japan, Vol. 40, No.2, pp.333-344, Mar. 2011.

[4] **Chen Liu**, Tianruo Zhang, Xin Jin, Minghui Wang and Satoshi Goto, “Fast Inter Mode Decision Algorithm Based on Residual Feature,” The Journal of the Institute of Image Electronics Engineers of Japan, Vol. 39, No.5, pp.663-671, Sep. 2010.

[5] Minghui Wang, Tianruo Zhang, **Chen Liu** and Satoshi Goto, "Region-of-Interest based Preprocessing for H.264/AVC Encoding," The Journal of the Institute of Image Electronics Engineers of Japan, Vol. 39, No. 5, pp.682-691, Sep. 2010.

International Conference

[1] **Chen Liu**, Xin Jin and Satoshi Goto, “Envelope Detection Based Workload Prediction for Partial Decoding Scheme,” 2012 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Kaohsiung, Taiwan, Dec. 2-5, 2012 (Accepted).

[2] Minghui Wang, **Chen Liu**, Tianruo Zhang, Xin Jin, Satoshi Goto, “Region of Interest Oriented Fast mode Decision for Depth map Coding in DIBR,” 2011 IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA 2011), pp. 177-180, Mar. 2011.

[3] Tianruo Zhang, Minghui Wang, **Chen Liu**, and Satoshi Goto, "Complexity Reduction Algorithm for Region-of-Interest based H.264 encoding," Asia-Pacific Conference on Circuits and Systems (APCCAS), pp. 336-339, Kuala Lumpur, Dec. 2010.

[4] **Chen Liu**, Xin Jin, Tianruo Zhang, Minghui Wang and Satoshi Goto, “PARTIAL DECODING SCHEME FOR H.264/AVC DECODER,” 2010 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2010), Chengdu, China, pp.345-348, Dec. 2010.

[5] **Chen Liu**, Xin Jin, Tianruo Zhang, Minghui Wang and Satoshi Goto, "Interactive partial video decoding for viewing resolution adaptation," International SoC Design Conference (ISOCC 2010), Incheon, Korea, pp.244-247, Nov. 22-23, 2010.

[6] Tianruo Zhang, Xin Jin, **Chen Liu**, Minghui Wang, and Satoshi Goto, "ROI based Complexity Reduction Algorithm For H.264 Encoder," International SoC Conference (ISOCC), pp. 236-239, Nov. 2010.

[7] **Chen Liu**, Tianruo Zhang, Xin Jin, Minghui Wang and Satoshi Goto, "Residual Analysis Based Fast Inter Mode Decision for H.264/AVC," 2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2009), Kanazawa, Japan, TP1-B-3, pp.339-342, Dec. 2009.

[8] Tianruo Zhang, **Chen Liu**, Minghui Wang, Satoshi Goto, "Region-of-Interest based H.264 Encoder for Videophone with a Hardware Macroblock Level Face Detector," 2009 IEEE International Workshop on Multimedia Signal Processing (MMSP'09), Rio de Janeiro, Brazil, pp.1-6, Oct. 2009.

[9] Tianruo Zhang, Minghui Wang, **Chen Liu**, and Satoshi Goto, "VLSI architecture of a low complexity face detection algorithm for real-time video encoding," the IEEE 8th international conference on ASIC (ASICON2009), Changsha, China, pp. 147-150, Oct. 2009.

[10] Minghui Wang, Tianruo Zhang, **Chen Liu**, Satoshi Goto, "Region-of-Interest Based Dynamical Parameter Allocation for H.264/AVC Encoder," 27th Picture Coding Symposium (PCS 2009), Chicago, Illinois, USA, pp. 1-4, May, 2009.

[11] **Chen Liu**, Tianruo Zhang, Xin Jin, Satoshi Goto, "An Inter Mode Decision Method with Motion Correlation Adaptation for H.264/AVC," The 5th

International Colloquium on Signal Processing & Its Applications (CSPA 2009), Kuala Lumpur, Malaysia, pp.264-268, Mar. 2009.

[12] Minghui Wang, Tianruo Zhang, **Chen Liu**, Satoshi Goto, "Region-of-Interest Based H.264 Encoding Parameter Allocation for Low Power Video Communication," The 5th International Colloquium on Signal Processing & Its Applications 2009 (CSPA 2009), Kuala Lumpur, Malaysia, pp. 235-239, Mar. 2009.