USC-SIPI REPORT #405

Advanced Techniques for High Fidelity Video Coding

by

Qi Zhang

August 2010

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

ADVANCED TECHNIQUES FOR HIGH FIDELITY VIDEO CODING

by

Qi Zhang

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

August 2010

Table of Contents

List Of Tables

List Of Figures

Abstract

This research focuses on two advanced techniques for high-bit-rate video coding: 1) subpel motion estimation and 2) residual processing.

First, we study sup-pixel motion estimation for video coding. We analyze the characteristics of the sub-pel motion estimation error surface and propose an optimal subpel motion vector resolution estimation scheme that allows each block with different characteristics to maximize its RD gain through a flexible motion vector resolution. Furthermore, a direct subpel MV prediction scheme is proposed to estimate the optimal subpel position. The rate-distortion performance of the proposed motion prediction scheme is close to that of full search while it demands only about of 10% of the computational complexity of the full search.

Secondly, we investigate high-bit-rate video coding techniques for high definition video contents. We observed that under the requirements of high-bit-rate coding, there still left a large portion of uncompensated information in the prediction residual that represents similar signal characteristics of film grain noise. Due to small quantization step size used by high-bit-rate coding, these untreated small features render all existing coding schemes ineffective. To address this issue, a novel granular noise prediction and coding scheme

is proposed to provide a separate treatment for these residuals. A frequency domain-based prediction and coding scheme is proposed to enhance the coding performance. The proposed granular noise prediction and coding scheme outperforms H.264/AVC by an average of 10% bit rate saving.

Thirdly, we further investigate on the impact of high-bit-rate coding from the more fundamental signal characteristics point of view. A probability distribution analysis on DCT coefficients from the H.264/AVC codec under different bit rates is conducted to reveal that the prediction residual in the form of DCT coefficients have a near uniform distribution for all scanning positions. To further understand this phenomenon, a correlation based analysis was conducted to show that the different types of correlations existed in the video frame and the distribution of these correlations highly impact the coding efficiency. A significant amount of short and medium-range correlations due to the use of a fine quantization parameter cannot be easily removed by existing compensation techniques. Consequently, the video coding performance degrades rapidly as quality increases. A novel Multi-Order-Residual (MOR) coding scheme was proposed. The concept is based on the numerical analysis to extract different correlation through different phases. A different DCT-based compensation and coding scheme combined with an improved rate-distortion optimization process was proposed to target the higher-order signal characteristics. An additional pre-search coefficient optimization phase was proposed to further enhance compression performance. Experimental results show that the proposed MOR scheme outperforms H.264/AVC by an average of 16% bit rate savings.

Chapter 1

Introduction

1.1 Significance of the Research

Video coding has been extensively studied in the last three decades. It has been widely used in video storage and communication. Earlier video compression research has put a lot of emphasis on low bit rate coding due to the limited availability in storage and bandwidth. However, with the increased popularity of high definition (HD) video content and increased transmission bandwidth in recent years, more research focus has shifted from low-bit-rate video coding to high-bit-rate (or high fidelity) video coding.

We show the video quality as a function of the compression ratio in Fig. 1.1 for today's state-of-the-art video coding algorithms. For applications such as video conferencing and DVD, the current H.264/AVC video coding algorithm can achieve a compression ratio of 100 or higher. However, when the quality requirement is above 35dB (*e.g.* for high-fidelity video), the compression ratio drops rapidly. This observation motivates us to investigate ways to further improve coding efficiency of high-fidelity video. Studies in Chapter 4 reveal that some of the uncompensated fine structural information present

in HD video contents prevents the current compression algorithms from achieving good coding efficiency in the high-bit-rate range. Additional analysis in Chapter 5 shows that the different types of correlation existing in the video highly impacts the coding efficiency under high-bit-rate coding environments. With the market demand shifting more towards high-fidelity and high-definition video, a more effective video coding algorithm for such an application is highly desirable.

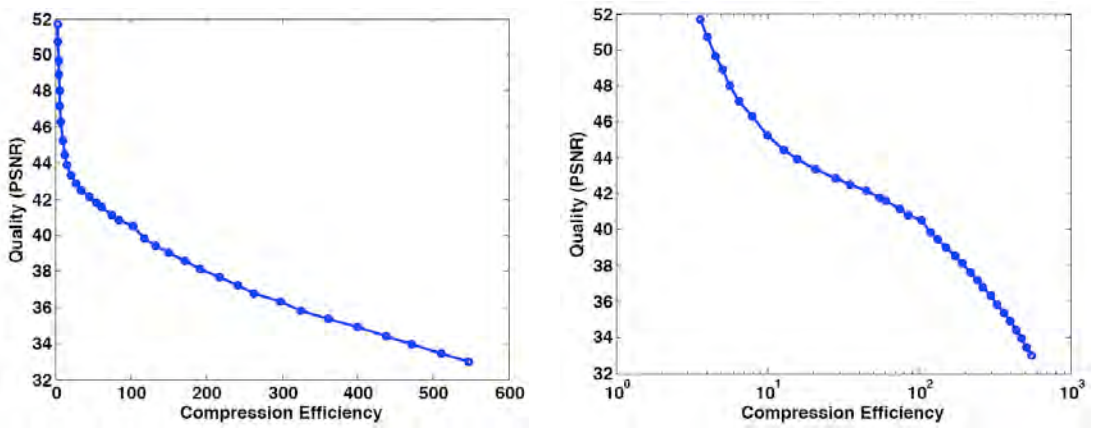
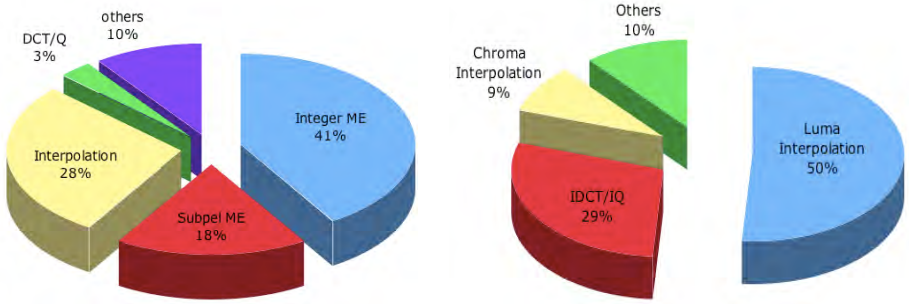


Figure 1.1: The video quality as a function of the compression ratio for the state-of-the-art video coding algorithms, where the compression ratio is in (a) the regular scale and (b) the logrithmic scale.

Subpel motion estimation (ME) provides another mechanism to achieve high fidelity video coding. However, the computational complexity of subpel ME is high. The complexity of a coding algorithm is typically measured in terms of the number of arithmetic operations (millions of instructions per second or MIPS), memory requirement, power consumption, chip area and the hardware cost. As compared to previous standards, H.264/AVC delivers the best coding performance at the cost of highest complexity. As shown in Fig. 1.2 which was reported in [40], integer and sub-pel motion estimation (ME) and interpolation are the two most time-consuming coding modules in the H.264/AVC

encoder while the luma interpolation is the most demanding module in the H.264/AVC decoder.



(a) (b)

Figure 1.2: The complexity profiling of H.264/AVC (a) encoder and (b) decoder [40].

In addition to computation complexity, HD video coding also imposes a lot of stress on frame memory allocation and access. For example, to encode a HD frame of size 1920×1080 with 5 reference frames, the interpolation module alone would require $1920 \times 1080 \times 5 \times 16 = 158MB$ of the frame memory just to store the pixel values on each integer position. Another $158 \times 4 = 632MB$ of frame memory for 1/2 pel values and another $632 \times 4 = 2528MB$ of frame memory for 1/4 pel values. That is, a total of more than 3GB memory is needed to save the interpolated values at the subpel position. Hence, there has been a large amount of efforts on the speed-up of subpel ME and interpolation, including instruction level optimization, fast search and fast interpolation algorithms. However, in spite of previous work as reviewed in Sec. 1.2, we see good opportunities for further performance improvement.

1.2 Review of Previous Work

Some previous work that is closely related to our research is reviewed in this section.

1.2.1 Lossless Coding

JPEG 2000 and M-JPEG2000 have been chosen by the Digital Cinema Initiative (DCI) [3] as the lossless video coding standard. In H.264/AVC, the 4x4 integer DCT and quantization processes contain shift operations, thus causing rounding errors. To meet the lossless coding requirement for intra coding in H.264, a Differential Pulse Code Modulation (DPCM)-based prediction scheme is first applied and prediction errors are then fed into the entropy coder in [48], where the transform and quantization processes are skipped. Although this new lossless coding scheme is more efficient than the M-JPEG2000 lossless standard, its coding efficiency is still slightly worse than that of the JPEG-LS standard [45].

1.2.2 Fast ME in Transform Domain

One ME approach is to estimate the cross-correlation of two macroblocks in the frequency domain [37]. The frequency spectrum of the input are normalized to give a phase correlation. However, the correlation performed by DFT-based methods is in circular (rather than linear) fashion. Hence, the correlation function could be inaccurate due to the edge effect. To reduce the problem of edge artifacts, Kuglin and Hines [9] proposed the zero padding method at the expense of higher computation complexity. Another ME approach as studied in [41] is to use a transform whose size is much larger than the maximum displacement under consideration. This approach is able to limit the amount of introduced errors, yet it is more suitable for global motion estimation rather than block-based local motion estimation. The third approach is to use the complex lapped transform (CLT) in the cross correlation calculation [49]. This technique is based on lapped orthogonal

transform (LOT), where its basis functions are overlapped and windowed by a smooth function with a shape like a half cosine. It introduces less block edge artifacts as compared to LOT in the spatial domain. The latest effort of the frequency domain prediction was proposed for intra prediction in VC-1 [3], where DC and AC components are separated and predicted from their left and top neighbor frequency components.

1.2.3 Fast Subpel ME

There has been extensive research on complexity reduction for subpel ME. In general, fast sub-pel ME schemes fall into two categories: 1) search complexity reduction and 2) interpolation complexity reduction. Fast search schemes lower subpel search complexity by reducing the number of sub-pixel search points under the assumption that the subpel error surface is monotonic (or parabolic) [36]. Lee *et al.* [26] proposed a subpel ME scheme that tests the four most promising half-pixel locations out of the eight. Thus, the complexity is halved. The surrounding eight integer positions are used to decide which half-pixel locations are selected. Yin *et al.* [36] proposed a similar method that used fewer integer positions which are determined by a thresholding method. The resultant fast search can reduce the complexity by 85% with the PSNR degradation of around 0.1 dB. A center-biased fractional pel search (CBFPS) algorithm for fractional-pel ME in H.264/AVC was proposed in [50] based on the characteristics of multi-prediction modes, multi-reference frames. However, CBFPS is only applied to smaller blocks while full fractional ME search is still adopted for larger blocks such as 16×16, 16×8, and 8×16. As a result, the speedup of CBFPS is somehow limited. To overcome this issue, an improved sub-pel search method was proposed in [47], which includes a simple and efficient sub-pel

skipping method based on statistical analysis and an immediate-stop technique based on the minimum cost. However, the total computation, as compared to the full subpel search method, is decreased only by 30%, because all candidate blocks still need to be interpolated to obtain the fractional-pixel resolution. Thus, the interpolation module is still the major bottleneck in terms of computation and memory access latency.

Fast interpolation schemes reduces the computational complexity by reducing the interpolation complexity instead of the search complexity. By establishing a subpel error surface based on a mathematical model, subpel ME errors on each sub-pixel position can be calculated from the model, thus eliminating the need of interpolation computation [25, 35]. Usually, model parameters can be found from errors at the nine integer MV locations and the optimal sub-pel MV location can be solved directly or iteratively. However, there is one major drawback with this approach; namely, the accuracy of subpel prediction can be heavily impacted if the model cannot characterize the error surface accurately. There may exist some discrepency between the actual and the modeling error surfaces as a result of the local image texture pattern.

1.3 Contribution of the Research

In this research, we propose a fast subpel MV prediction scheme in Chapter 3, and residual prediction and coding schemes in Chapters 4, and 5 respectively. Major contributions are summarized as below.

- Contributions in fast subpel ME (Chapter 3)

- We conduct an in-depth analysis on the problem of previous optimal subpel MV resolution estimation algorithms. Basically, they are based on input block texture characteristics. As the input block subject to motion compensation, quantization and noise factors, they are not accurate for some cases. Then, we propose an optimal MV resolution estimation scheme that can provide a more accurate estimation.

- The proposed optimal MV resolution estimation allows blocks with different characteristics to maximize its RD gain through a flexible MV resolution while significantly reduce complexity. Based on how well the error surface is conditioned, two different optimal MV prediction schemes are proposed respectively. The rate-distortion performance of the proposed optimal MV prediction excels that of full search with an average of 90% complexity reduction.

• Contributions in granular noise prediction and coding (Chapter 4)

- We conduct an mode distribution analysis on the residual image from current H.264 codec under the requirements of high-bit rate coding. The study shows that there are still a large amount of uncompensated fine features in form of granular noise left in the residual that causes the coding efficiency to degrade significantly. The design of proposed GNPC system allows the system to be easily integrated with any traditional video codec. No modification is required to the existing codec. Decoder could discard the transmitted noise frame if the decoding time frame is not sufficient. This would not affect the decoding of the consecutive incoming content frames, as they are coded independently.

- We propose an granular noise prediction and coding scheme that resembles the film grain extraction process to extract these fine features in the residual. A frequency-domain based prediction and compensation scheme was further proposed for granular noise data. By correlating the same frequency bands between different blocks, we could maximize the possibility between target GN block and candidate blocks that might contain similar low frequency components but different high frequency components to be considered as candidate reference blocks and vice versa. The prediction between the same frequency bands avoids the complication of sparse matrix multiplication for reconstruction as required in earlier ME in frequency domain.
- By quantizing the input DCT block before the prediction module, there will be no additional computation requirement to perform the quantization and inverse quantization during the rate-distortion optimization phase. Hence, complexity can be significantly reduced. Furthermore, the proposed coding scheme is more friendly for the rate control purpose. As quantization is done prior to the GNPC, the distortion/PSNR of each block/frame can be known ahead of time without going through the entire RDO process.
- Experimental results demonstrate the effectiveness of proposed frequency-domain based GNPC scheme with an average bit rate reduction of 10%.

- Contributions in multi-order residual coding (Chapter 5)
  - To understand the impact of high bit rate coding, we study the DCT coefficient distribution and show that, as the video quality requirement increases, the

distribution of DCT coefficients is close to an uniform one. This explains the poor performance of traditional image/video codecs in the high bit rate region.

– We conduct a correlation analysis on input image frames, which reveals that there exist different types of correlations in the image, which has a significant impact on coding efficiency. To address this problem, we adopt different coding schemes to remove different types of correlations in image frames, which is called the multi-order residual (MOR) prediction and coding system.

– We study the characteristics of the extracted medium and long correlations in the higher-order residuals. Since these MOR data have a small dynamic range with a flat distribution at every scanning position in the block, the traditional MCP with RDO in the pixel domain may not be effective. This observation motivates us to adopt a frequency-domain based prediction for MOR data.

– By quantizing the input DCT block before the prediction module, the RDO phase can have a direct evaluation of prediction results without going through the DCT, quantization, inverse DCT and inverse quantization for each search position. Hence, the complexity can be reduced significantly.

– The effectiveness of the proposed MOR coding scheme is demonstrated by experiments, which outperforms the state-of-the-art H.264/AVC codec by 30-50% in the bit rate saving.

1.4 Organization of the Dissertation

The rest of this dissertation is organized as follows. Related research background on lossless compression, intra coding, film grain noise sythesis, and subpel motion prediction is reviewed in Chapter 2. A direct subpel motion vector prediction is proposed in Chapter 3. The granular noise prediction and coding scheme (GNPC) is presented in Chapter 4. The multi-order residual (MOR) prediction and coding scheme is investigated in Chapter 5. Finally, concluding remarks and future work are given in Chapter 6.

Chapter 2

Research Background

Earlier video coding algorithms have focused on low bit rate coding due to the limited storage space and transmission bandwidth. However, due to increased popularity of high definition video and availability of the broadband networking infrastructure, the research focus has gradually shifted from low-bit-rate coding to high-bit-rate (or high fidelity) coding. The latter includes lossless and near lossless video coding.

High definition (HD) video programs have several unique characteristics worth special attention. First, as compared with standard definition TV (SDTV), more detail textures are recorded at a much higher fidelity range to create a more involving experience to the audience. Second, HD video has a higher spatial resolution. The well-known resolution is 1920x1080 progressive (or 1080p). The latest released HD-DVD and Blu-ray disc both support 1080p. Even higher resolutions have been considered. For example, Digital Cinema Initiatives (DCI) [1] recommends the 4K by 2K camera.

As compared to previous video coding standards, the latest H.264/AVC standard [46] can provide about 50% bit rate saving for the same video quality. However, H.264/AVC was initially developed for the low bit rate applications, and most of its experiments were

conducted on low resolution QCIF and CIF sequences. As the spatial resolution increases, H.264/AVC reaches a performance bottleneck. Thus, one of the long term objectives set by the Joint Video Team (JVT) is to develop a new generation video coding standard that would keep abreast with this significantly increased demand on high definition and high fidelity video coding.

In this chapter, we briefly review some background in the areas of motion compensated prediction, frequency-domain motion estimation (ME), noise synthesis/coding and fast subpel ME.

2.1 Motion Compensated Prediction

The main module in a compression system is prediction, which exploits the spatial and temporal redundancy between pixels to achieve bit rate saving. A classic prediction schemes can be divided into two distinct phases: *1) modeling* and *2) coding*. In the modeling phase, the encoder gathers the statistics about the input data and builds up a probabilistic model. A prediction model is formed to make inference on the coming sample by assigning a conditional probability distribution to it. The prediction error is then sent to the coding phase with some level of quantization, where either arithmetic coding or Huffman coding is used as the lossless entropy coder. Since the entropy coder is well developed, the most critical design choice is hence the algorithm in the modeling phase.

The structure of a typical prediction scheme can be stated as follows.

1. A prediction phase is carried out to determine the prediction, $\hat{x}_{i+1}$, of the input data sample, x_{i+1}, based on a finite set of previously coded data $x_1, x_2, ..., x_i$.

2. The prediction error, $\epsilon_{i+1} = x_{i+1} - \hat{x}_{i+1}$, is computed.

3. A context decision rule is developed to determine a context c_{i+1} in which x_{i+1} occurs. This context is usually another function of elements in the previously coded data set.

4. A probabilistic model is derived for the prediction error ϵ_{i+1} based on context c_{i+1}.

The prediction error is then entropy coded based on the incoming symbol probability distribution. Since the decoder follows the same set of rules while decoding, the same prediction, context and probability distribution can be repeated at the decoder and, hence, the original input data sample can be reconstructed completely without any error.

Thus, the key to an efficient coding scheme lies in the capability of the prediction scheme to minimize the prediction error. In the following, we will present several prediction schemes.

2.1.1 Block-based Motion Compensated Prediction

Block-based motion compensated prediction was mainly used to explore temporal similarities and hence were widely adopted for inter-frame coding [21]. It is initially designed based on the concept of block matching as shown in Fig. 2.1. It assumes that there is a very small displacements (d_x, d_y) between the consecutive frames. Thus, the frame to frame difference $FD(x, y)$ can be approximated mathematically as:

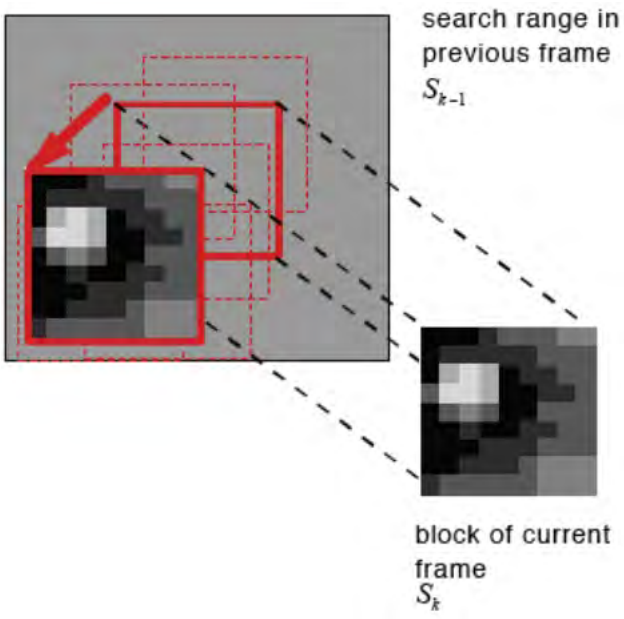


Figure 2.1: Block matching process

$$FD(x,y) \approx -\frac{\partial S(x,y)}{\partial x}d_x - \frac{\partial S(x,y)}{\partial y}d_y, \tag{2.1}$$

For practical implementation, the block matching process is proposed as follows. It first subdivide an input image into squared block and find a displacement vector for each block. Within the given search range, a best "match" is found based on minimizing a given error measure criteria [28].

Some of the popular error measurement matrix include sum of squared error (SSD) in Eq. (2.2), sum of absolute difference (SAD) in Eq. (2.4) and sum of absolute transformed difference (SATD) in Eq. (5.2).

$$SSE(d_x, d_y) = \sum \left[S_k(x,y) - S_{k-1}(x+d_x, y+d_y)\right]^2, \tag{2.2}$$

$$SAD(d_x, d_y) = \sum \left|S_k(x,y) - S_{k-1}(x+d_x, y+d_y)\right|. \tag{2.3}$$

$$SATD(d_x, d_y) = \sum \left|T\left[S_k(x,y) - S_{k-1}(x+d_x, y+d_y)\right]\right|. \tag{2.4}$$

The T used in Eq.(5.2) is usually Hadmard transform for simplicity. The general understanding is that SATD usually offers the best performance as it is more close to the true prediction error that is being encoded by the entropy coder. While SSE and SAD offers very similar performance, with SAD has the lowest computation requirements.

As motion compensated prediction is the most computationally complex module in the encoder, there have been extensive research done to speed up the search while maintain a good search quality [31, 11, 33, 14].

2.2 Frequncy-domain MCP

Another type of MCP is to conduct the MCP in frequency domain rather than in the pixel domain. Earlier MCP scheme is to estimate the cross-correlation function in the frequency domain [37]. The frequency spectrum of the input can be normalized to give a phase correlation. However, the correlation performed by a DFT-based method corresponds to a circular convolution rather than a linear one, and the correlation function could be

affected by the edge effect. To reduce the edge artifact, Kuglin and Hines [9] proposed to use zero padding to the input data sequence at the cost of higher complexity. Another technique is to use a transform size which is much larger than the maximum displacement considered [41]. This approach can limit the error size, but it is more suited for global ME rather than block-based ME. A third technique is to use the complex lapped transform (CLT) to perform the cross correlation in the frequency domain [49]. Since the basis functions are overlapped and windowed by a smooth function that shapes like a half cosine, it introduces less block edge artifacts as compared to the LOT in the spatial domain.

A frequency-domain ME technique was proposed by Chang and Messerschmitt in [12]. As shown in Fig. 2.2, motion search of 8x8 DCT blocks with respect to the previous frame is conducted in the DCT domain. The prediction error is quantized and entropy coded. This allows to skip the inverse DCT (IDCT) since ME is performed in the frequency domain. Since the coding loop of the spatial domain ME is modified, the memory requirement reduces as well [27]. However, these schemes are not widely used for some reasons. First, most previous video coding algorithms focus on low bit rate coding. With coarse quantization used in low bit rate coding, most DCT coefficients in a block are quantized to zero and there is little space for rate distortion improvement. Second, the proposed frequency-domain ME treats all frequency components equally, where all frequency components are compensated simultaneously with the same spatial offset. It is similar to that of motion compensation in the spatial domain except that frequency components are compensated directly rather than pixel values. Then, if the spatial domain cannot provide enough correlation, it is unlikely to get a better prediction in the

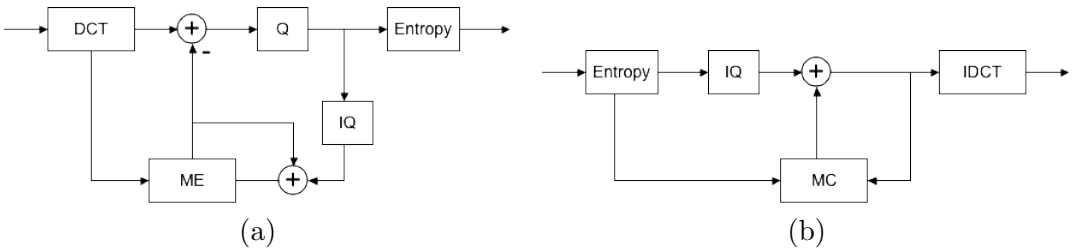


Figure 2.2: The block diagram of (a) the encoder and (b) the decoder with motion estimation and compensation in the DCT domain.

frequency domain. The latest effort of prediction in the frequency domain was proposed for intra prediction in VC-1. The DC and the AC components are predicted from their left and top neighboring frequency components.

2.2.1 H.264/AVC Intra Prediction

As intra frame coding does not have the luxury to explore temporal correlation, intra prediction is mainly designed to explore only spatial correlation. H.264 employes a unique line-based intra prediction scheme. The prediction is carried out on a marcoblock basis, but can be subdivided into smaller partitions such as 8x8 and 4x4 subblock sizes.

For intra 16x16 predictions, one of the four prediction modes can be chosen: horizontal, vertical, DC and plane modes. For intra 8x8 or 4x4 predictions, nine directions can

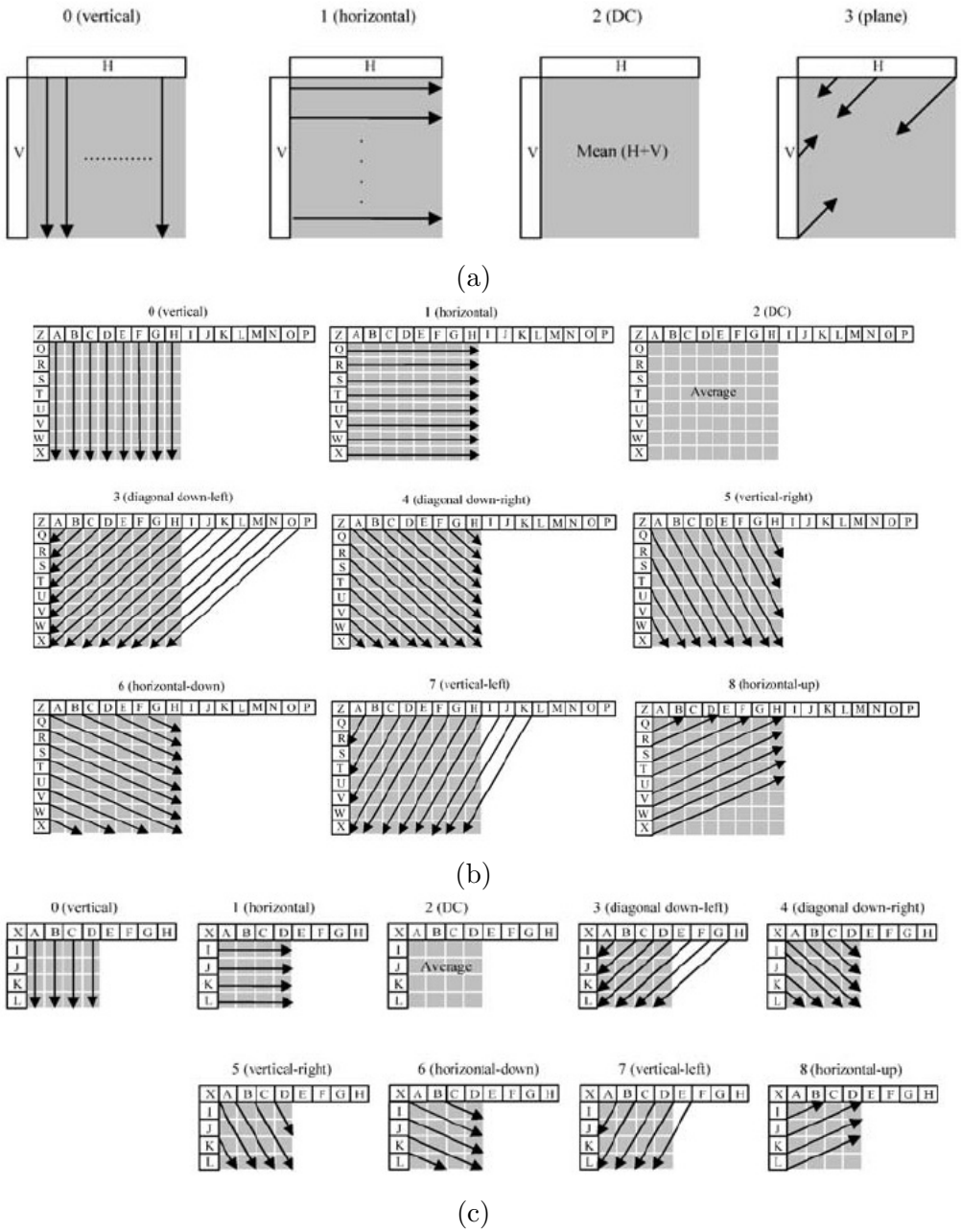


Figure 2.3: Neighboring pixel samples used in (a) Intra 16x16 (b) Intra 8x8 and (c) Intra 4x4 modes.

be applied. See Fig. 2.3. We may use the horizontal prediction mode as an example, where the prediction can be expressed as

$$r_0 = p_0 - q_0, \tag{2.5}$$

$$r_1 = p_1 - q_0, \tag{2.6}$$

$$r_2 = p_2 - q_0, \tag{2.7}$$

$$r_3 = p_3 - q_0, \tag{2.8}$$

The residual difference of $r_0, \cdots, r_3$ predicted from block boundary samples are sent to the decoder together with the mode information for correct reconstruction of the block. The only difference between lossless and lossy intra predictions is that the residual difference will go through DCT and quantization in lossy coding but these steps are skipped in lossless coding. An improved lossless intra prediction was proposed by Lee *et al.* [48] that changes the block-based prediction to a sample-based prediction. For example, for the horizontal prediction (mode 1), the residual difference of $r_0, \cdots, r_3$ are predicted using a sample-by-sample DPCM method. Mathematically, it can written as

$$r_0 = p_0 - q_0, \tag{2.9}$$

$$r_1 = p_1 - p_0, \tag{2.10}$$

$$r_2 = p_2 - p_1, \tag{2.11}$$

$$r_3 = p_3 - p_2. \tag{2.12}$$

The vertical mode (mode 0), mode 3 and mode 4 can be conducted in a similar fashion.

2.3 Film Grain Noise Compression

Film grain noise is related to the physical characteristics of the film and can be perceived as a random pattern following a general distribution statistics [8]. Film grain noise is not prominent in the low-resolution video format such as CIF and SD. However, the fine structure becomes more visible once the video resolution goes to HD. Film grain noise is one of key elements used by artists to relay emotion or cues so as to enhance the visual perception of the audience. Sometimes, the film grain size varies from frame to frame to provide different clues in time reference, etc. Here, we consider film grain noise as one type of granular noise. For lossless video coding, it is desirable to preserve the quality of granular noise without modifying the original intent of filmmakers. In addition, it is the requirement in the movie industry to preserve granular noise throughout the entire image and delivery chain.

Due to the random nature of granular noise, it is difficult to have an efficient energy compaction solution. Since film grain noise has a relatively larger energy level in the high frequency band, the block-based encoder in the current video coding standards is not efficient even in the DCT domain. Besides, it also degrades the performance of motion estimation. Thus, researches have been focusing on granular noise removal and synthesis. That is, granular noise is first removed in a pre-processing stage at the encoder using and then re-synthesized using a model and added back to the filtered frame at the decoder. Because only the noise model parameters are sent to the decoder instead of actual noise, the overall bit rate can be reduced significantly. Film grain coding has been considered in H.264/AVC [32].

Several algorithms on texture synthesis have been proposed and can be used for granular noise synthesis [16, 42]. Research on granular noise synthesis can be classified into three areas: 1) sample noise extraction, 2) granular noise database, and 3) model-based noise synthesis. Gomila and Kobilansky [10] proposed a sample-based approach that extracts a noise sample from a source signal and applies a transformation to it. Only one noise block is sent to the decoder in the SEI message. However, it could suffer from visible discontinuity and repetition. The granular noise database method employs a comprehensive granular noise database [22] that contains a pool of pre-defined granular noise values for the film type, exposure, aperture, etc. The film grain selection process follows a random fashion corresponding to the average luminance of the block, and a deblocking filter is used to blend in granular noise. This method allows the generation of realistic granular noise but requires both the encoder and the decoder to have access to the same granular noise database. The model-based noise synthesis approach extracts granular noise in a pre-processing stage, the extracted noise is analyze and a parametric model containing a small set of parameters is estimated and sent to the decoder. It provides an efficient coding method. However, the noise removal operation could potentially remove actual contents (*e.g.,* the explosion dust) as well.

2.4 Sub-pel Motion Estimation

A well performed motion vector (MV) search is critical to the efficiency of video coding because of its capability to reduce temporal redundancy between a sequence of frames [21,

33]. The motion estimation algorithm using the full search block matching algorithm (FS-BMA) is often used as performance benchmarking. The best integer MV is obtained under the assumption that all pixels within the same blocks have the same horizontal and vertical displacements in an integer unit. However, the best frame-to-frame block displacement of video contents may not coincide with the sampling grid. As a result, the integer MV cannot represent the desired displacement well, and a sub-pixel motion compensation scheme is more suitable. The importance of sub-pel accuracy in ME has been widely recognized. An increased subpel MV resolution will provide significant improvement on rate-distortion performance for some blocks as analyzed by Girod in [7].

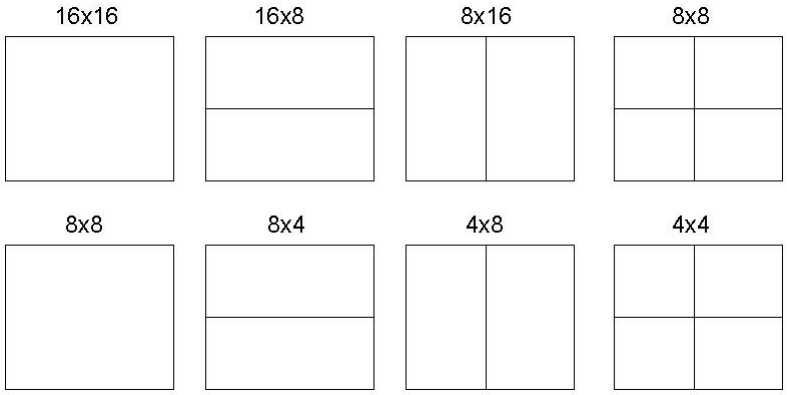


Figure 2.4: Illustration of different inter prediction block sizes in H.264/AVC.

To implement sub-pel motion search, either the reference frame has to be completely interpolated and stored in the memory or some blocks need to be repeatedly interpolated as subpel refinement is performed. The former requires a large storage space while the latter will significantly increase computational complexity. This problem becomes even more severe if a higher sub-pel resolution (such as the 1/8-pel) is used. Therefore,

although 1/8 pel motion estimation was proposed, only up to quarter-pel ME is standardized in H.264/AVC. In addition, subpel ME is performed together with the mode decision algorithm in H.264/AVC. When inter prediction is used in H.264/AVC, one 16x16 MB can be partitioned into one 16x16 block, two 16x8 or 8x16 blocks or four 8x8 blocks while each 8x8 block can be further partitioned into two 8x4 or 4x8 blocks or four 4x4 blocks as shown in Fig. 4.1. As a result, there are totally 19 modes to encode one 16x16 MB.

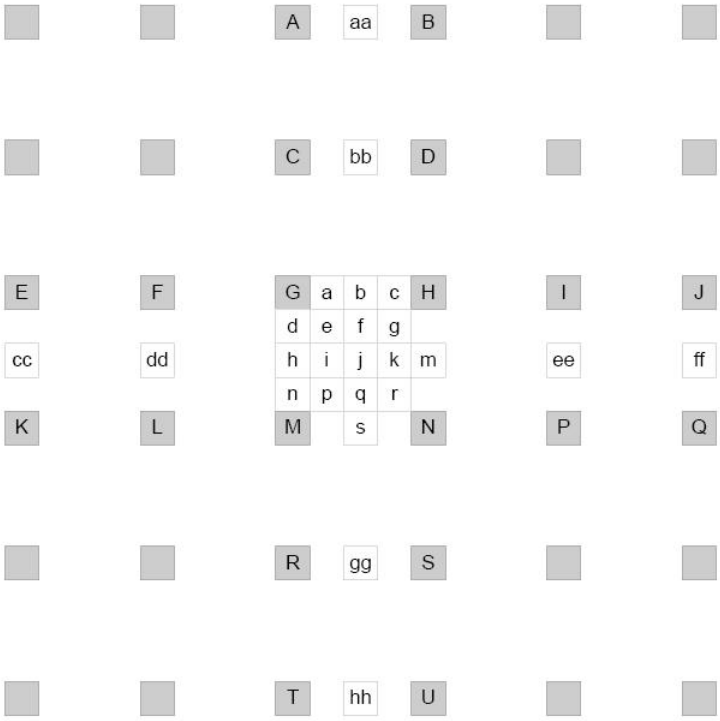


Figure 2.5: Interpolation filter for sub-pel accuracy motion compensation.

Moreover, each block whose size is larger than 8x8 can be predicted using different reference frames. For each mode, the MV can be of integer-, half- or quarter-pel resolution. The half-pel value is obtained by applying a one-dimension 6-tap FIR interpolation filter horizontally (the x-direction) or vertically (the y-direction). The quarter-pel value

is obtained by the average of two nearest half-pel values. For example, the half-pel value of the fractional sample b in Fig. 2.5 is obtained by applying 6-tap FIR interpolation filter to those pixels *E, F, G, H, I and J* as

$$b = \frac{E - 5F + 20G + 20H - 5I + J}{32} \tag{2.13}$$

Then, the quarter-pel value of the fractional sample a in Fig. 2.5 is given by

$$a = \frac{b + G}{2}. \tag{2.14}$$

2.5 Conclusion

In this chapter, we provided a review on the background that is relevant to the research presented in the following chapters The challenges and requirements with high fidelity video coding were presented. In Chapter 3, we would like to design an accurate subpel MV estimation scheme that has the ability to predict the optimal subpel MV position without exorting to the overly complex subpel interpolation. In Chapters 4 and 5, our goal is to design a high fidelity video coding system that has the ability to encode high definition video in the high bit rate range more efficiently.

Chapter 3

Direct Subpel Motion Estimation Techniques

3.1 Introduction

A well performed motion vector (MV) search is critical to the efficiency of video coding because of its capability to reduce temporal redundancy between frames of a sequence. The importance of subpel accuracy in motion estimation (ME) has been widely recognized [21]. An increased subpel MV resolution will provide significant improvement on the rate-distortion (R-D) performance for some blocks as analyzed by Girod [7]. Traditionally, to implement subpel MV search, either the reference frame is completely interpolated and stored in the memory or some blocks are repeatedly interpolated as the subpel refinement process is performed. The former requires a large storage space while the latter will have higher computational complexity. This problem becomes more severe if a higher subpel resolution is adopted. For example, with 1/8-pel MV resolution, the computational complexity and memory requirements involved in the motion estimation and interpolation are very high [46, 40]. For this reason, although 1/8 pel motion estimation was proposed for H.264/AVC, only up to quarter-pel ME is standardized in H.264/AVC.

There have been extensive research on reducing the complexity of subpel motion estimation (ME). In general, fast subpel ME schemes fall into two categories: 1) reducing the search complexity and 2) reducing the interpolation complexity. Fast search schemes lower the subpel search complexity by reducing the number of search points on each subpel position based on the assumption that the subpel error surface is often concave [36, 50, 47, 26]. However, as they are search-based, each subpel position still needs to be interpolated ahead of time, which could be a major bottleneck in performance speedup. Fast interpolation schemes address this issue by reducing the interpolation complexity. By establishing a subpel error surface with a mathematical model, the subpel ME error at each subpel position can be extrapolated from the model, thus eliminating the need of heavy interpolation computation [25, 24, 13, 35]. However, the performance of these schemes is highly dependent on model accuracy. In addition, fast interpolation is conducted for one resolution at a time. That is, one has to perform search for the optimal subpel position among extrapolated subpel ME errors at all subpels of a given resolution before moving to the next subpel resolution.

Although it is common to find the optimal MV position by fitting a local error surface using integer-pel MVs, the characteristics of the error surface have not been thoroughly studied in the past. In this work, we use the condition number of the Hessian matrix of the error surface to characterize its shape in a local region. Specifically, we characterize an error surface by its condition number, which is defined as the ratio of the largest and the smallest eigenvalues of the 2×2 Hessian matrix (or the ratio of the long and the short axes of its 2D elliptic contour). To reduce the complexity, we propose an approximate

condition number in the implementation. After the error shape analysis, we study direct techniques for the optimal resolution estimation and position prediction of subpel MVs.

It is known in the literature [7], [39] that the optimal MV resolution should be adaptive to the characteristics of the underlying video. However, there is no practical algorithm that estimates the optimal subpel MV resolution on a block-to-block basis. Ribas-Corbera and Neuhoff [39] proposed a texture-based estimation scheme to determine the optimal MV resolution for different blocks. Their method only considers the characteristics of the input block without leveraging integer search results. By exploiting the result of the error surface analysis, we propose a block-based subpel MV resolution estimation scheme that allows blocks of different characteristics to maximize their rate-distortion (R-D) gain by choosing the optimal subpel MV resolution adaptively. Fast subpel MV prediction has been studied by researchers before, *e.g.,* Suh and Jeong [25, 24], Cho *et al.* [13] and Hill *et al.* [35]. However, there has been no rigorous study on the accuracy of predicted subpel MVs. We propose two MV prediction schemes for well-conditioned and ill-conditioned blocks, respectively. All proposed techniques are called direct methods, since no iteration is involved in optimal subpel MV resolution estimation and position prediction. Experimental results are given to show the excellent R-D performance of the proposed sub-pel MV prediction schemes.

In this work, we first conduct an analysis on the existing subpel MV estimation model to reveal its weakness in Sec. 3.2. Then, we propose a block-based optimal subpel MV resolution estimation scheme in Sec. 3.3. Based on how well the error surface is conditioned, two optimal MV prediction schemes are presented in Sec. 3.4. A subpel MV prediction scheme is proposed for ill-conditioned blocks to estimate the optimal subpel

position in one step (namely, without refining the resolution by half at a time as done in [25, 24, 13, 35]) in Sec. 3.4.1. This direct prediction scheme is further extended to provide accurate prediction for well-conditioned blocks in Sec. 3.4.2 [51]. Experimental results are provided to demonstrate the effectiveness of the proposed schemes in Sec. 3.5. Finally, concluding remarks and future research directions are given in Sec. 3.6.

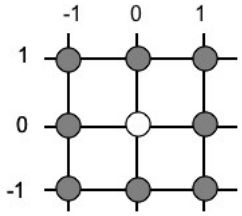


Figure 3.1: Illustration of a square window of dimension $-1 < \Delta x, \Delta y < 1$ centered around the optimal integer-pel MV position indicated by the central empty circle.

3.2 Characterization of Local Error Surface

Subpel ME is usually conducted after the optimal integer MV is obtained through integer motion estimation. It is typically assumed that the optimal subpel MV should reside within a square window of dimension $-1 \leq x, y \leq 1$ centered around the optimal integer position as shown in Fig. 3.1. Then, we can define a subpel motion estimation error surface over this window using a common error measure known as the sum of squared differences (SSD)

$$E(\Delta x, \Delta y) \quad = \quad \sum \Big[s(x,y) - c(x_0 + \Delta x, y_0 + \Delta y) \Big]^2, \tag{3.1}$$

where (x_0, y_0) is the location of the optimal integer pel, $s(x, y)$ is the target block and $c(x_0 + \Delta x, y_0 + \Delta y)$ is a reference block with $-1 < \Delta x, \Delta y < 1$.

3.2.1 Problem with Traditional Surface Modeling

Several surface models have been used as to approximate the SSD error surface. Examples include the 9-term, 6-term and 5-term error models, denoted by E_9, E_6 and E_5, respectively. Mathematically, they can be written as [25], [34]:

$$\begin{aligned} E_9(\Delta x, \Delta y) &= a\Delta x^2\Delta y^2 + b\Delta x^2\Delta y + c\Delta x\Delta y^2 \\ &\quad + d\Delta x\Delta y + e\Delta x^2 + f\Delta x + g\Delta y^2 \\ &\quad + h\Delta y + i, \end{aligned} \tag{3.2}$$

$$\begin{aligned} E_6(\Delta x, \Delta y) &= a\Delta x^2 + b\Delta x\Delta y + c\Delta y^2 + d\Delta x \\ &\quad + e\Delta y + f, \end{aligned} \tag{3.3}$$

$$E_5(\Delta x, \Delta y) = a\Delta x^2 + b\Delta y^2 + c\Delta x + d\Delta y + e. \tag{3.4}$$

Coefficients a, b, $\cdots$ in above are model parameters and they are calculated based on the measured prediction error at the specified nine integer positions [25], [34]. Note that a contour of surface model E_6 corresponds to a rotated 2D ellipse while that of surface model E_5 corresponds to a simple ellipse whose axes aligned well with the x- and the y-axes. Simply speaking, the ratio of the long and the short axes of these ellipses defines the condition number of an error surface. The ratio is small (or large) for a well-conditioned (or ill-conditioned) surface.

Usually, one estimates model parameters (*i.e.,* cofficients in these models) based on errors in the nine integer MV locations given in Fig. 3.1 and solve for the optimal subpel MV location directly (for models E_5 and E_6) or iteratively (for model E_9). However, depending on the local image texture pattern, there may exist great discrepency between the actual error surface and the approximated ones provided by models E_9 and E_6 sometimes. We show the 3D plot and the 2D contour plot of the actual error surface and the 2D contour plots of models E_9 and E_6 for well-conditioned and ill-conditioned error surfaces in Figs. 3.2 and 3.3, respectively. One problem with previous model-based fast interpolation schemes [25, 35]) is that the minimum of the subpel error surface predicated by models may fall outside the defined square window as shown in Fig. 3.3 (c). More recently, methods were proposed to reduce the 2-D model into 1-D models in [24, 13], where the minimum search along the X and the Y axes is done independently.

To overcome the problem that the minimum of the subpel error surface will fall outside the defined square window, Cho *et al.* [13] added a step of selective interpolation with a hope that the error surface could be well behaved in a smaller window of size $-0.5 \leq x, y \leq 0.5$. Their scheme demands additional MV error computation at eight new half-pel locations. Although they can get the optimal half-pel MV among these evaluated locations, the resulting subpel MV may deviate from the true one significantly as shown in Fig. 3.3 (c). Besides, there is no easy way to determine the behavior of the error surface at a finer resolution. Hill *et al.* [35] derived a surface model from E_6 for the quarter-pel MV resolution and adopted a fallback scheme by performing the actual interpolation if the quality of MV estimation is poor. However, no statistical analysis

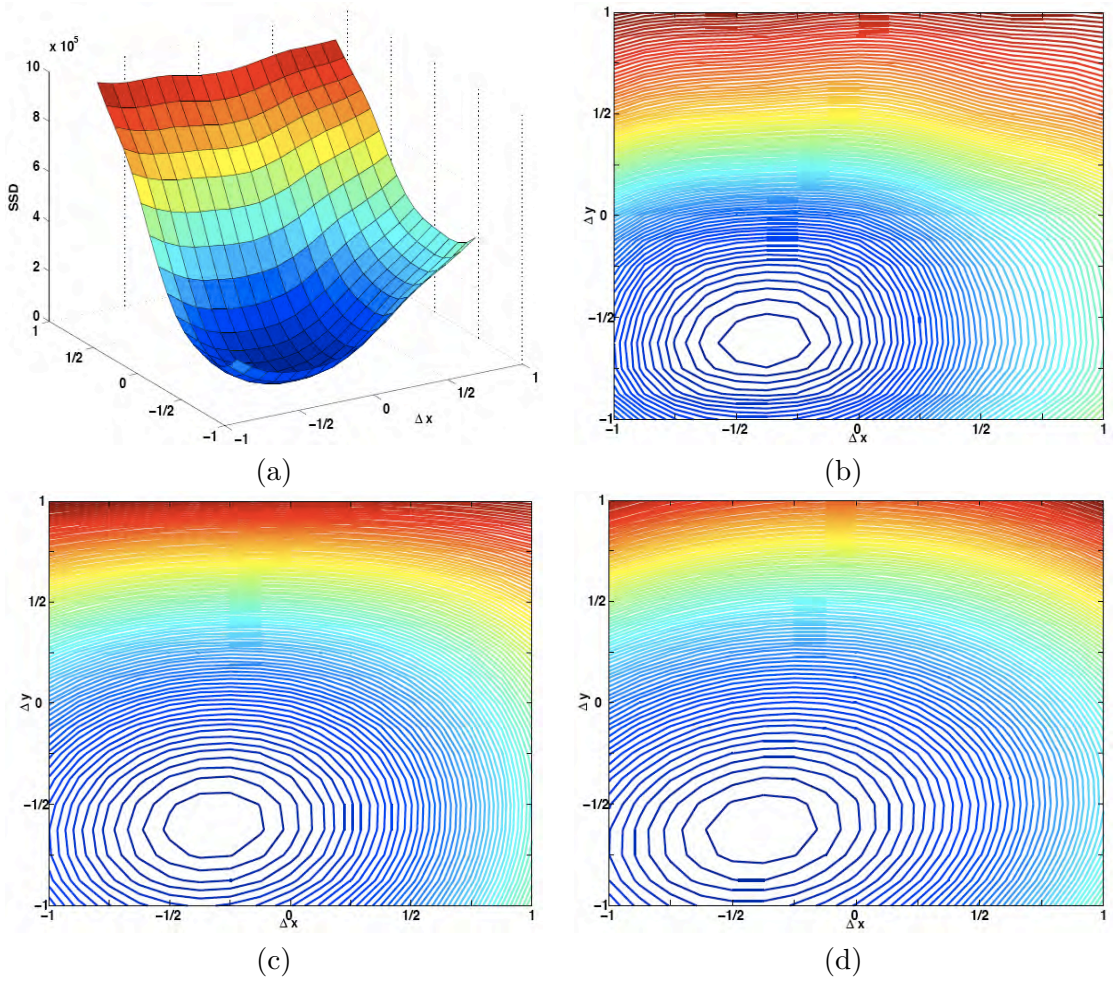


Figure 3.2: Illustration of error surfaces for a well-conditioned block: (a) the 3D plot of the actual error surface; and the 2D contour plots of (b) the actual error surface, (c) error surface model E_9, and (d) error surface model E_6.

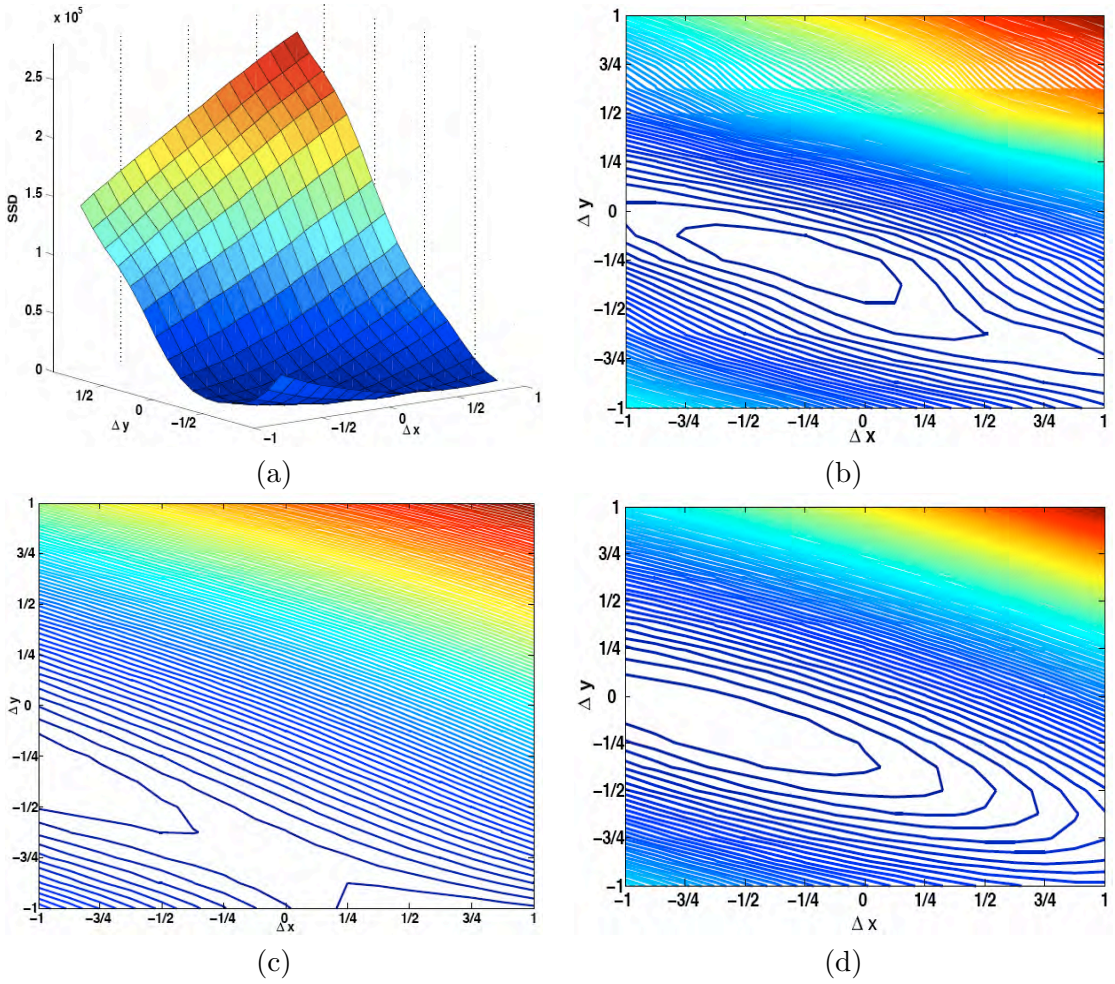


Figure 3.3: Illustration of error surfaces for an ill-conditioned block: (a) the 3D plot of the actual error surface; and the 2D contour plots of (b) the actual error surface, (c) error surface model E_9, and (d) error surface model E_6.

on the relationship between the model quality and the accuracy of subpel MV resolution prediction has been conducted before.

By following previous work, we assume that the local error surface is a convex function (*i.e.* a uni-modal error surface). We claim that prediction accuracy actually depends on whether it has a narrow valley with a certain orientation at the bottom of the error surface. This is visually apparent by comparing Figs. 3.2(b) and 3.3(b). Mathematically, the local error surface can be characterized by the second-order derivatives of the center pixel, known as the Hessian matrix [6] of that point. The eigenvalues of the Hessian matrix are called *principal curvatures*. The condition number is the ratio of its largest and smallest eigenvalues of the Hessian matrix (or, geometrically, the ratio of the long and the short axes of its 2D elliptic contour). For a well-conditioned block, its error surface is circularly symmetric. As the condition number increases, it becomes ill-conditioned gradually. To solve the coefficients of error surface models E_5, E_6 and E_9, we need to solve a linear system of equations via matrix inversion. If the matrix is ill-conditioned, one cannot get the model coefficients robustly. This explains why the model-based approach fails to predict the optimal subpel motion location accurately.

3.2.2 Condition Number Estimation

In this subsection, we focus on the problem of estimating the condition number of the error surface in a local window consisting of 3×3 pixels based on the nine sampled points. Here, we consider four slices of the error function; namely, the intersection of the error surface and four planes:

- the horizontal (or the 0-dgree) slice with $\Delta y = 0$ as the intersecting plane;

- the vertical (or the 90-degree) slice with $\Delta x = 0$ as the intersecting plane;

- the 45-degree slice with $\Delta x = \Delta y$ as the intersecting plane;

- the 135-degree slice with $\Delta x = -\Delta y$ as the intersecting plane.

The four intersection curves are shown in Figs. 3.4 (a) and (b). Fig. 3.4 (a) correspond to a well-conditioned block case where the four curves have similar curvatures. Fig. 3.4 (b) correspond to an ill-conditioned block case, where the curvatures spread over a wider range.

Based on the above observation, we can derive a simple test to check how well a block is conditioned. That is, we can define the following four paramters:

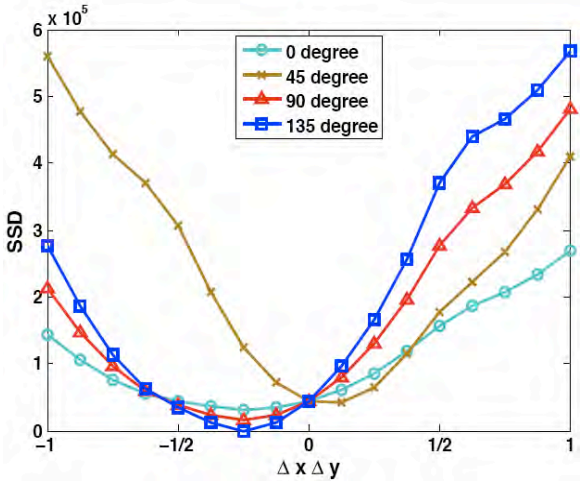
$$\alpha_0 = |e(-1,0) + e(1,0) - 2e(0,0)|, \tag{3.5}$$

$$\alpha_{45} = |e(1,1) + e(-1,-1) - 2e(0,0)|, \tag{3.6}$$

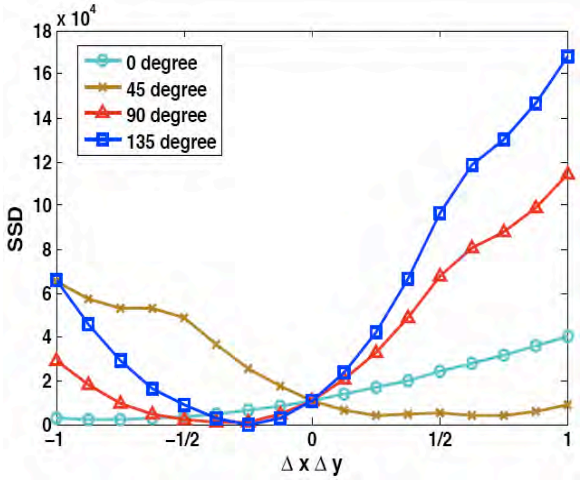
$$\alpha_{90} = |e(0,-1) + e(0,1) - 2e(0,0)|, \tag{3.7}$$

$$\alpha_{135} = |e(-1,1) + e(1,-1) - 2e(0,0)|, \tag{3.8}$$

where $e(\Delta x, \Delta y)$ is the measured integer-pel error at $(\Delta x, \Delta y)$. They correspond to the 1D discrete Laplacian along the 0-, 45-, 90- and 135-degree directions, respectively. Generally speaking, a larger (or smaller) value of α implies a more rapidly-changing (or slowly-changing) error surface along the corresponding direction. The maximum and the



(a)



(b)

Figure 3.4: The error curves passing through the origin along the 0-, 45-, 90- and 135-degree directions for (a) a well-conditioned block, and (b) an ill-conditioned block.

minimum of these four parameters are denoted by $\alpha_{\max}$ and $\alpha_{\min}$, respectively. Then, we can compute an approximated condition number of the Hessian matrix in this region via

$$C = \frac{\alpha_{\max}}{\alpha_{\min}}. \tag{3.9}$$

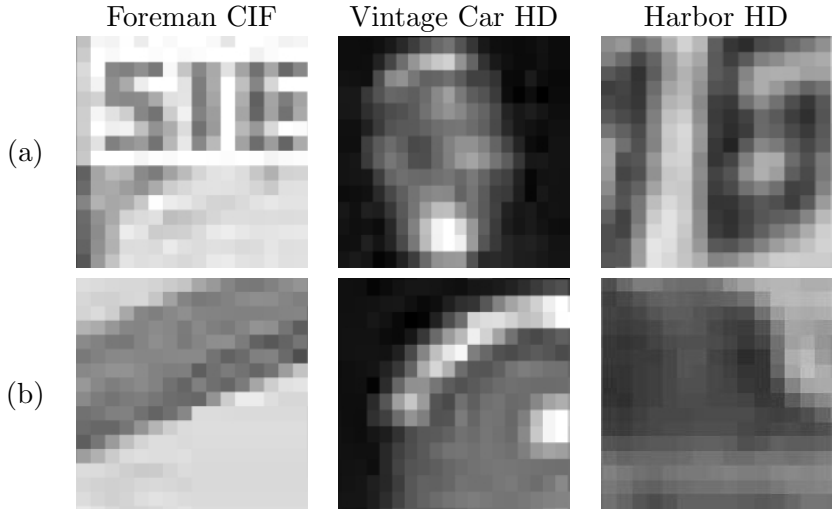


Figure 3.5: (a) Block examples that are likely to have well-conditioned error surfaces; (b) block examples that are likely to have ill-conditioned error surfaces. Blocks are taken from sample sequences of Foreman CIF, Vintage Car HD and Harbor HD.

In Fig. 3.5, we show some representative regions from three test sequences that yield well-conditioned or ill-conditioned error surfaces after the sub-pixel motion estimation process. As shown in the Foreman, Vintage Car, the Harbor examples in Fig. 3.5, we see that it is likely to get well-conditioned cases for regions with certain symmetry and ill-conditioned cases for regions with angled textures. However, the characteristics of the local error surface is ultimately determined by the temporal relationship of two adjacent frames. In other words, we are not able to make robust decision based on the texture pattern of a single frame. This also explains why previous work [39], which determines

the subpel MV accuracy based on the input block texture only, does not yield satisfactory results.

3.2.3 Deviation from Flatness

The optimal subpel MV resolution is related to the curvature of the error surface. For a flat error surface, the cost of the increased MV resolution tend to impact the overall R-D performance negatively. On the other hand, for a steep error surface, a finer subpel resolution is advantageous as it would result in an additional R-D gain. To capture the curvature information of the error surface, we may consider the following two simple measures:

$$D_f = \sqrt{\alpha_{\max}^2 + \alpha_{\min}^2}. \tag{3.10}$$

For simpler computation, we can approximate this parameter as:

$$D_f \approx \left|\alpha_{\max}\right| + \left|\alpha_{\min}\right|. \tag{3.11}$$

In this work, we adopt the measure defined in Eq. (3.11) and call it the *deviation from flatness* for its greater simplicity. The optimal subpel MV position prediction is related to the bottom shape of the error function. We will elaborate this in the following section.

3.3 Optimal Subpel MV Resolution Estimation

Girod [7] pointed out that optimal subpel MV resolution has critical impact on coding efficiency and not all blocks need the same MV resolution to obtain the best coding performance. Some blocks can benefit from higher MV resolution while others cannot. Thus, it is desirable to have adaptive MV resolution for optimal coding efficiency rather than fixed MV resolution. To study the problem of optimal subpel MV resolution, Girod [7] provided an analytical framework that estimates the difference-frame energy with an optimal subpel MV resolution, which is expressed as a function of the probability distribution of MV accuracy, the Fourier transform of the frame and the power spectral density of inter-frame noise. This framework is however not easy to implement in practice. Ribas-Corbera and Neuhoff [39] extended Girod's framework and developed a scheme to estimate the optimal MV resolution for a block using its texture. However, their method is still too complex for actual implementation and not accurate enough for prediction on a block-to-block basis. In this section, we propose an optimal subpel MV resolution estimation scheme. It is related to the characterization of the subpel error surface features with parameter D_f as given in Eq. (3.11). This method is not only easy to compute on a block-to-block basis but also effective in enhancing the R-D performance.

In Figs. 3.6 (a), we show the histogram of D_f for a collection of video sequences while Figs. 3.6 (b)-(e) depict the probability for the optimal subpel MV resolution to be of integer, half-, quarter-, eighth-pel accuracy as a function of D_f with quantization parameter $QP = 20$. In computing these probabilities, the optimal MV resolution selection

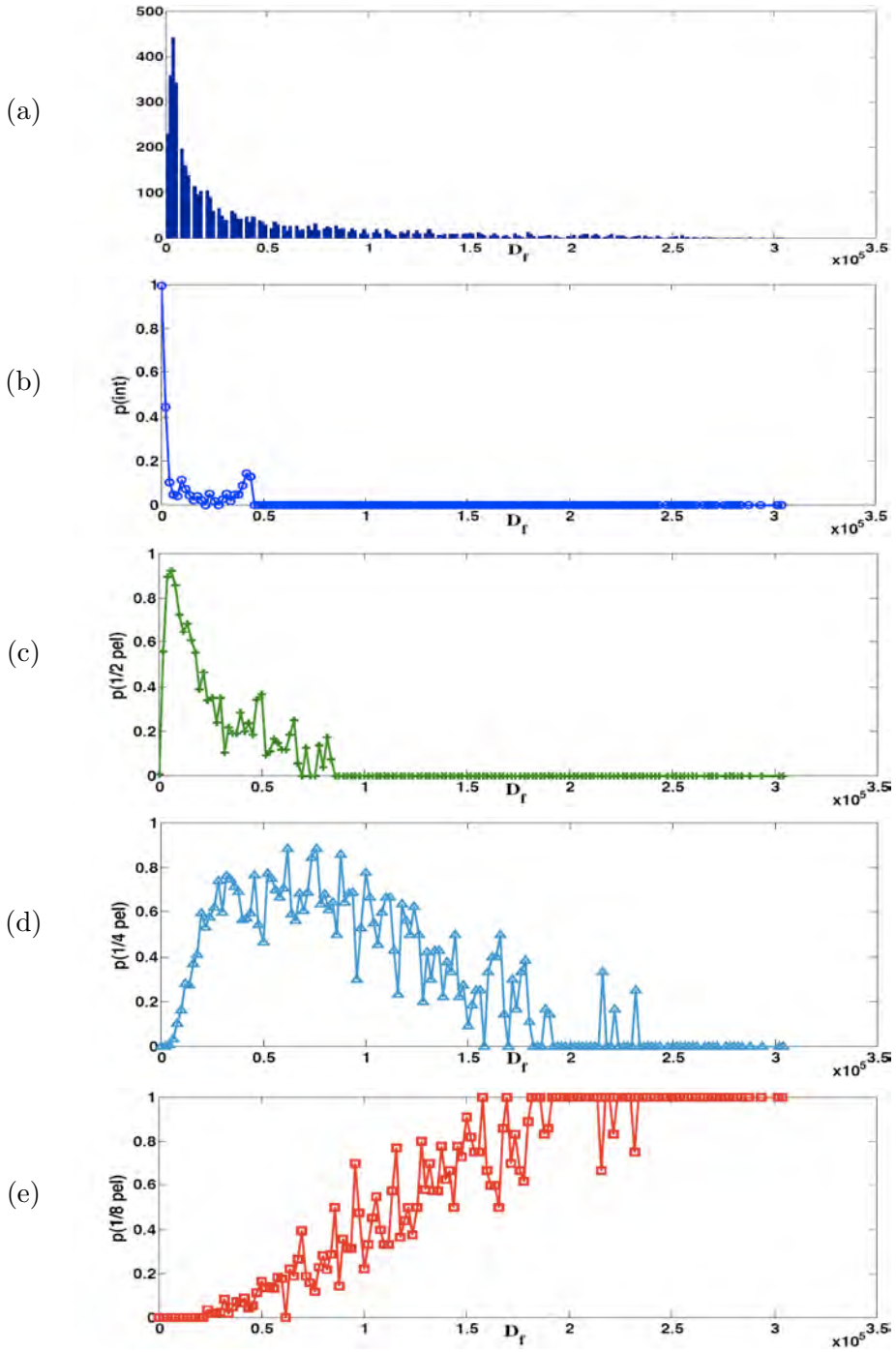


Figure 3.6: (a) The histogram of D_f at QP=20 and (b)-(e) the probability distributions for the optimal MV resolution at integer-pel, 1/2-pel, 1/4-pel and 1/8-pel for a set of test video sequences.

process is similar to the H.264/AVC Rate-Distortion Optimization (RDO) procedure [44] using the following Lagrangian cost function:

$$J(s, c|\lambda_{motion}) = SSD(s, c) + \lambda_{motion} \cdot R(s, c), \tag{3.12}$$

where $R(s, c)$ is the number of bits associated with the coding of the prediction error and MV, s is the source block texture, c is the reference block texture, and λ_{motion} is the Lagrangian multipler which is set to $\sqrt{0.85 \cdot 2^{QP/3}}$. Under this RDO framework, the distortion model does not consider the quantization effect on the prediction error since the optimal MV resolution selection is performed with a given QP value.

For a very flat error surface whose D_f value is extremely small, additional subpel MV accuracy does not bring a sufficient performance gain to justify the rate overhead. Thus, the probability of selecting the integer pel as its optimal MV resolution is nearly 100%. However, as D_f increases, the probability of selecting 1/2 pel resolution as its optimal MV becomes dominant. The quarter-pel MV resolution is important when D_f exceeds 25,000. The switch between quarter- and eighth-pel resolution is more gradual as shown in Fig. 3.6 3.6(d) and (e). Actually, the probability of selecting quarter-pel or eight-pel is similar for a range of D_f values. When the error surface is very steep (corresponding to a large D_f value), the chance of selecting eight-pel MV becomes very high. Since there are few blocks that has a D_f value over 150,000 as shown in Fig. 3.6 (a), there is no advantage to go to finer MV resolution such as the 1/16 pel. Although QP is chosen in Fig. 3.6, the same observation holds for different QP values. In other words, only the D_f

value is critical to the subpel MV resolution estimation. This conclusion is much simpler than that given in [7] and [39].

Based on the above discussion, we can have a simple estimation scheme for the optimal subpel MV resolution of local block b, denoted by $\phi_b(MV)$, as

$$\phi_{i,j}(MV) = \begin{cases} 1 & \text{if } D_f(i,j) \leq \tau_1, \\ \frac{1}{2} & \text{if } \tau_1 < D_f(i,j) \leq \tau_{1/2}, \\ \frac{1}{4} & \text{if } \tau_{1/2} < D_f(i,j) \leq \tau_{1/4}, \\ \frac{1}{8} & \text{if } \tau_{1/4} < D_f(i,j) \leq \tau_{1/8}, \end{cases} \tag{3.13}$$

where $D_f(i,j)$ is the deviation from flatness measure at pixel (i,j) that has the smallest integer MV value (*i.e.* the central pixel in Fig. 3.1 and τ_i, $i = 1, 1/2, 1/4$ are proper threshold values). We do not observe an advantage to go to a subpel of less than 1/8 so that we choose 1/8 as the finest resolution as shown in above.

If $D_f \leq \tau_1$, only the best integer position is coded. No further subpel MV prediction is needed since the error surface in this block is too flat for subpel MV to improve the coding gain. Generally speaking, thresholds τ_1, $\tau_{1/2}$ and $\tau_{1/4}$ can be selected via statistical analysis. Sometimes, we may set them to higher (or lower) values to trade the quality for lower (or higher) computational complexity.

3.4 Direct Subpel MV Position Prediction

In this section, we propose two direct methods for subpel MV position prediction depending on the condition number of a local block. The ill-conditioned and well-conditioned blocks are considered in Secs. 3.4.1 and 3.4.2, respectively.

First, we show the distribution of the condition number of all blocks from a set of test video sequences in Fig. 3.7 (a). They include four CIF sequences (*i.e.,* Container, Football, Coastguard and Tempete), one HD sequence at 1280× 720 resolution (*i.e.,* Sheriff) and one HD sequence at 1920×1080 resolution (*i.e.,* Station2). We did not use the same sequences in Sec. 3.5 to show the robustness of the training process in determining the well- and ill-conditioned cases.

The performance of the E_9 model by Suh and Jeong [25], called the SJ E_9 model in short, is evaluated in Fig. 3.7 (b), where the average Euclidean distance between the predicted and the actual subpel positions is plotted as a function of the condition number. This prediction error distance can be written mathematically as

$$\varepsilon_s = \sqrt{(\Delta x_a - \Delta x_s)^2 + (\Delta y_a - \Delta y_s)^2}, \tag{3.14}$$

where $(\Delta x_s, \Delta y_s)$ and $(\Delta x_a, \Delta y_a)$ are the predicted and the actual subpel MV positions, respectively.

We see that, for a well-conditioned block with $C \leq 4$, the prediction error distance, ε_s, generated by the SJ E_9 model is small enough for accurate quarter-pel MV resolution. However, as the condition number increases, the average prediction error becomes larger. The average prediction error goes beyond the quarter-pel resolution for blocks with $C > 4$,

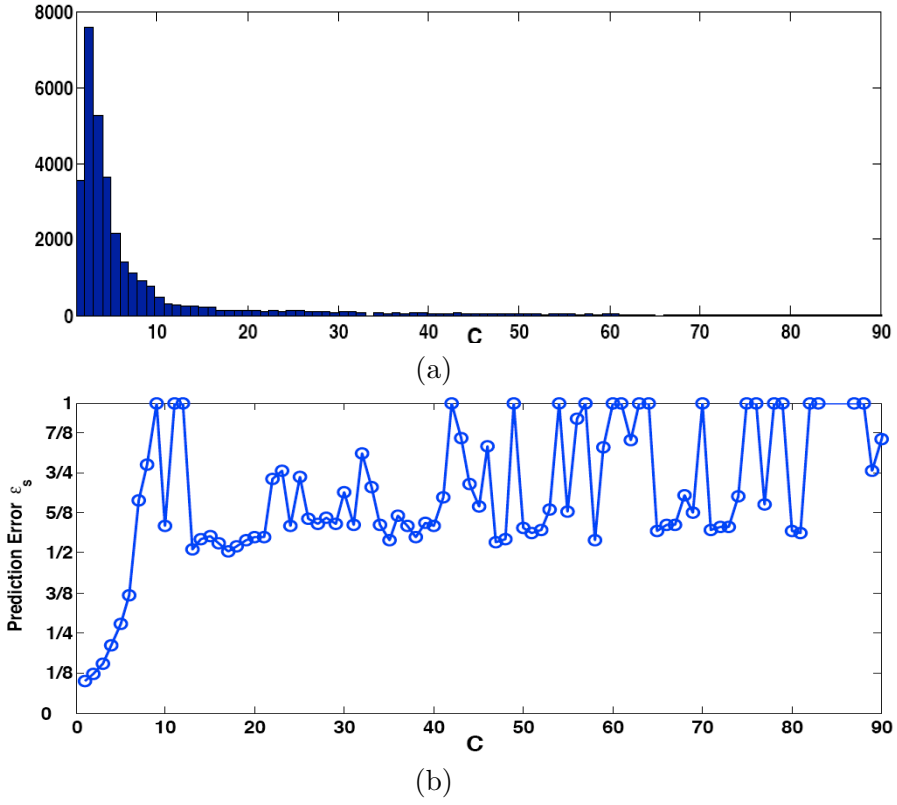


Figure 3.7: (a) The histogram of condition numbers and (b) the prediction error distance ε_s as a function of the condition number using the SJ E_9 model .

and exceeds the half-pel resolution for blocks with $C > 10$. As the condition number continues to increase, the prediction error increases to the maximum allowed by any subpel search method (*i.e.,* the integer-pel resolution), and the SJ method fails completely.

For a typical video stream, a large percentage of blocks falls in the well-conditioned block group as shown in Fig. 3.7 (a). Generally speaking, about 60% of blocks are in the well-conditioned group. For the remaining blocks, if only the half-pel resolution is needed, the SJ E_9 model can cover additional 20% of blocks. For a higher subpel resolution such as the 1/8 pel, the SJ E_9 model only applies to a very small percentage of blocks, *i.e.* only blocks with $C = 1$.

3.4.1 Ill-Conditioned Blocks

In this subsection, we focus on blocks whose error surface is ill-conditioned and propose a direct subpel MV position prediction scheme. The basic idea is to decompose a 2D optimization problem into two 1D optimization problems. Without loss of generality, we assume $\alpha_{90} > \alpha_0$ in the following discussion. Under this condition, we know that the error surface changes more rapidly along the axis of Δy than that of Δx.

Our algorithm consists of the following two steps as illustrated in Figs. 3.8 (a) and (b), respectively.

- Step 1: For a fixed value of $\Delta x(=-1,0,1)$, we use three values $e(\Delta x,1)$, $e(\Delta x,0)$ and $e(\Delta x,-1)$ to fit a quadratic function. Mathematically, they are in form of

$$e(-1,\Delta y) = A_{-1}\Delta y^2 + B_{-1}\Delta y + C_{-1}, \tag{3.15}$$

$$e(0,\Delta y) = A_0\Delta y^2 + B_0\Delta y + C_0, \tag{3.16}$$

$$e(1,\Delta y) = A_1\Delta y^2 + B_1\Delta y + C_1. \tag{3.17}$$

The global minima of Eqs. (3.15)-(3.17), denoted by $(-1,\Delta y^*_{-1})$, $(0,\Delta y^*_0)$ and $(1,\Delta y^*_1)$, can be determined analytically as

$$\Delta y^*_{-1} = \frac{1}{2}\frac{e(-1,1)-e(-1,-1)}{e(-1,1)+e(-1,-1)-2e(-1,0)}, \tag{3.18}$$

$$\Delta y^*_0 = \frac{1}{2}\frac{e(0,1)-e(0,-1)}{e(0,1)+e(0,-1)-2e(0,0)}, \tag{3.19}$$

$$\Delta y^*_1 = \frac{1}{2}\frac{e(1,1)-e(1,-1)}{e(1,1)+e(1,-1)-2e(1,0)}. \tag{3.20}$$

Based on Eqs. (3.18)-(3.20) and (3.15)-(3.17), we can compute the minimal error value. This process is shown in Fig. 3.8(a).

- Step 2: When the approximate condition number $\alpha_{\max}/\alpha_{\min}$ is larger, it is observed that the three minima, $(-1,\Delta y^*_{-1})$, $(0,\Delta y^*_0)$ and $(1,\Delta y^*_1)$, tend to have a co-linear relationship as illustrated in Fig. 3.8(b). Then, we can examine the plane that passes through these three points and fit another quadratic function

$$e_f(t) = A_f t^2 + B_f t + C_f, \tag{3.21}$$

which goes through their corresponding error values. Coefficients A_f, B_f and C_f can be solved and the minimum of Eq. (3.21) gives the optimal MV position at $(\Delta x_{opt}, \Delta y_{opt})$.

Although the predicted optimal MV position can take any real value in Δx_{opt} and Δy_{opt}, their values should be quantized to the optimal MV resolution, which is estimated using the technique presented in Sec. 3.3. There exist four possible candidates around $(\Delta x_{opt}, \Delta y_{opt})$ at a supported subpel resolution. A simple quantization scheme is to select its nearest neighbor among these four.

3.4.2 Well-Conditioned Blocks

We extend the direct subpel MV position prediction to blocks with a well-conditioned error surface in this section. One distinct error surface characteristics associated with ill-conditioned blocks is that the surface has a narrow valley with a certain orientation. Hence, for direct subpel MV prediction, there exists only one axis that can produce two or three minima to form $e_f(t)$ as shown in Fig. 3.8. On the other hand, for well-conditioned blocks, Step 1 should be repeated for both x- and y-axis. Thus, we can modify the process as follows.

- Step 1-x: It is the same as Step 1 in Sec. 3.4.1.

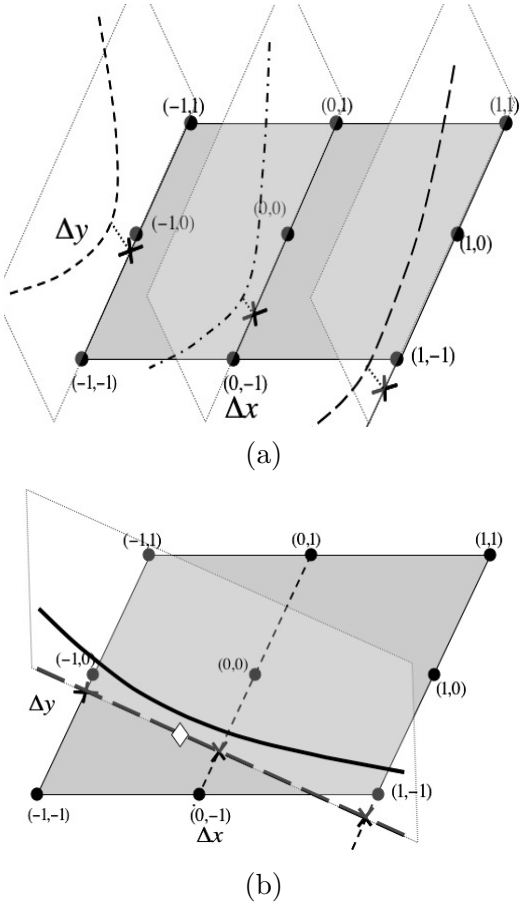


Figure 3.8: Illustration of the optimal subpel MV position prediction for ill-conditioned blocks: (a) Step 1: finding the minma in three vertical planes using quadratic curve fitting and (b) Step 2: connecting the three minima found in Step 1 and finding the optimal subpel MV position with another quadratic curve fitting.

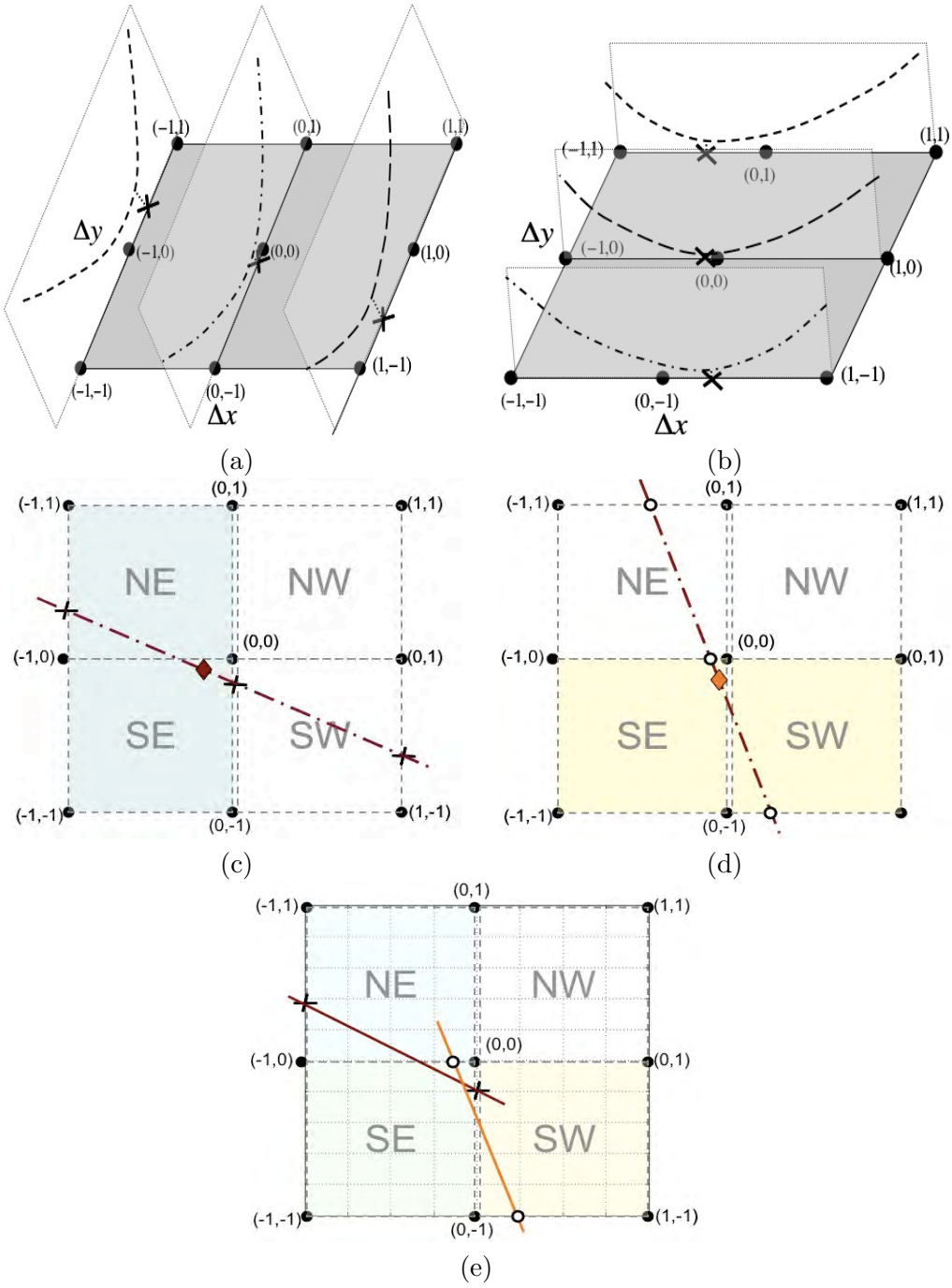


Figure 3.9: Illustration of the optimal subpel MV position prediction for a well conditioned block.

- Step 1-y: For given $\Delta y = (-1, 0, 1)$ we use $e(1, \Delta y)$, $e(0, \Delta y)$ and $e(-1, \Delta y)$ to fit a quadratic function. Mathematically, they are in form of

$$e(\Delta x, -1) = D_{-1}\Delta y^2 + E_{-1}\Delta y + F_{-1}, \tag{3.22}$$

$$e(\Delta x, 0) = D_0\Delta y^2 + E_0\Delta y + F_0, \tag{3.23}$$

$$e(\Delta x, 1) = D_1\Delta y^2 + E_1\Delta y + F_1. \tag{3.24}$$

The global minima of Eqs. (3.22)-(3.24), denoted by $(\Delta x^*_{-1}, -1)$, $(\Delta x^*_0, 0)$ and $(\Delta x^*_1, 1)$, can be determined analytically as

$$\Delta x^*_{-1} = \frac{1}{2}\frac{e(1,1) - e(1,-1)}{e(-1,-1) + e(1,-1) - 2e(0,-1)}, \tag{3.25}$$

$$\Delta x^*_0 = \frac{1}{2}\frac{e(1,0) - e(-1,0)}{e(-1,0) + e(1,0) - 2e(0,0)}, \tag{3.26}$$

$$\Delta x^*_1 = \frac{1}{2}\frac{e(1,1) - e(-1,1)}{e(-1,1) + e(1,1) - 2e(0,1)}. \tag{3.27}$$

- Step 2-x: It follows the same process as Step 2 in Sec. 3.4.1, which will produce a vertical-oriented optimal MV position at $(\Delta x_v, \Delta y_v)$. Based on Eqs. (3.25)-(3.27), we can compute the minimal error value using Eqs. (3.22)-(3.24).

- Step 2-y: We examine the plane that passes through these three points obtained in Step 1-y and fit them with another quadratic function

$$e_h(t) = D_h t^2 + E_h t + F_h, \tag{3.28}$$

which goes through their corresponding error values. Coefficients D_h, E_h and F_h can be solved and the minimum of Eq. (3.28) gives the horizontal-oriented optimal MV position at $(\Delta x_h, \Delta y_h)$. We then divide the $(-1,1)\times(-1,1)$ window into north-east (NE), north-west (NW), south-east (SE), and south-west (SW) four sectors. Then, $(\Delta x_v, \Delta y_v)$ obtained in Step 2-x would identify the east or the west sector of the actual optimal MV horizontally, and $(\Delta x_h, \Delta y_h)$ would identify the south and the north sector of the actual optimal MV vertically. This process is shown in Fig. 3.9 (c) and (d).

- Step 3: Based on the coordinates of $(\Delta x_v, \Delta y_v)$, two closest integer positions along the vertical direction can be selected to form one line. The same process can be done in the horizontal direction based on the coordinates of $(\Delta x_h, \Delta y_h)$ to form another line. These two lines are denoted by

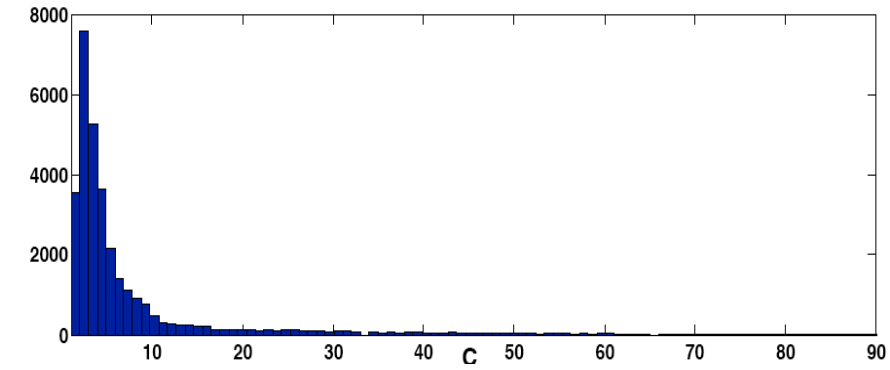
$$v = m\Delta x + n, \tag{3.29}$$

$$h = p\Delta y + q. \tag{3.30}$$

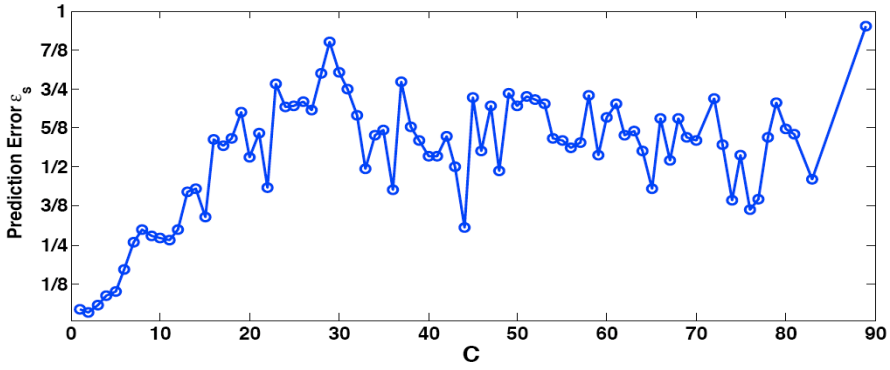
Finally, The optimal subpel MV position is the intersection point of these two lines as illustrated in Fig. 3.9 (e).

3.4.3 Performance Evaluation

The performance of the proposed subpel MV position prediction method is shown in Fig. 3.10 (b). As compared to SJ's E_9 method in Fig. 3.7 (b), we see that more than 90% of the blocks can achieve an average prediction error smaller than 1/4 pel resolution using



(a)



(b)

Figure 3.10: (a) The histogram of the condition number, and (b) the prediction error distance ε_s as a function of the condition number using the proposed prediction method as described in Secs. 3.4.1 and 3.4.2.

the proposed method. In addition, we see from the error histogram that as the condition number exceeds a certain value, the prediction error would be larger than the required

subpel MV resolution. Thus, we modify the optimal subpel MV estimation scheme given in Eq. (3.13) as follows:

$$\phi(MV) = \begin{cases} 1 & \text{if } D_f \leq \tau_1, \\ \frac{1}{2} & \text{if } \tau_1 < D_f \leq \tau_{1/2}, \\ \frac{1}{4} & \text{if } \tau_{1/2} < D_f \leq \tau_{1/4}, \\ \frac{1}{8} & \text{if } \tau_{1/4} < D_f \\ \text{Disable prediction} & \text{if } C > \tau_C, \end{cases} \tag{3.31}$$

where the threshold value, τ_C, can be obtained statistically.

3.5 Experimental Results

As the existing H.264/AVC reference codec does not support subpel MV resolution higher than quarter-pel MV, we modified reference codec JM12.1 [2] slightly to accommodate the 1/8-pel MV resolution for optimal subpel MV search in this section. A total of eight video sequences were tested: four of the CIF resolution (*i.e.*, Foreman, Mobile, Stefan, and Flower garden @352x288) and four of the HD resolution (*i.e.,* City corridor @1280x720 HD, Night @1280x720 HD, Blue sky @1920x1080 HD and Vintage car @1920x1080 HD). We adopted a window of size 32×32 for full integer MV search with one reference frame. The rate-distortion (R-D) optimization was employed in the MV search process. Each GOP consisted of 15 frames. The CAVLC was chosen as the entropy coder. The thresholds in Eq. (3.13) for the optimal MV resolution selection were set experimentally.

The experiments were run on a Macbook with Intel core 2 duo at 2.2GHz. We have implemented a state-of-the-art fast subpel MV estimation algorithm proposed in [24], which is an extension of the E_9 model in [25], for performance benchmarking. It is called the SCJ method, and we use SCJ 1/4pel and SCJ 1/8pel to denote the results for its application to the quarter-pel and the eighth-pel cases. We conducted experiments with the following two test settings.

- Test Setting 1

  We compare the R-D performance between H.264 full quarter-pel MV search (denoted by H.264 1/4pel), the proposed subpel MV position prediction scheme using the same quarter-pel MV resolution without optimal MV resolution (denoted by ZDK-I), and the SCJ 1/4pel method.

- Test Setting 2

  We compare the performance between H.264 full subpel MV search with the eighth-pel MV resolution (denoted by H.264 1/8pel), the proposed MV prediction scheme with the optimal MV resolution estimation method enabled (denoted by ZDK-II), and the SCJ 1/8pel method.

First, we examine the complexity saving of the direct subpel MV position prediction. Here, the complexity saving factor is defined as

$$S = \left\{1 - \frac{C_{proposed}}{C_{full}}\right\} \times 100(\%), \tag{3.32}$$

Table 3.1: The complexity saving $S(\%)$ of the proposed ZDK-I, ZDK-II and the SCJ method with respect to H.264 full search.

		Proposed Methods		SCJ Method	
Resolution	Sequences	ZDK-I	ZDK-II	SCJ 1/4pel	SCJ 1/8pel
352x288	Foreman	82.28	99.35	83.34	84.54
	Mobile	87.85	99.78	84.82	81.46
	Stefan	86.54	99.61	82.48	85.76
	Flower garden	87.91	99.50	86.35	87.25
Average		**86.15**	**99.56**	**84.25**	**84.75**
1280x720	City corridor	84.26	99.46	82.11	84.38
	Night	83.24	99.58	83.49	85.59
1920x1080	Blue sky	85.16	99.81	80.32	82.96
	Vintage car	89.24	99.67	82.77	82.05
Average		**85.48**	**99.63**	**82.17**	**83.75**

where $C_{proposed}$ and C_{full} denote the computational time required for the proposed subpel MV prediction and the full subpel search processes, respectively. For the latter, it includes the time required to interpolate the reference frame.

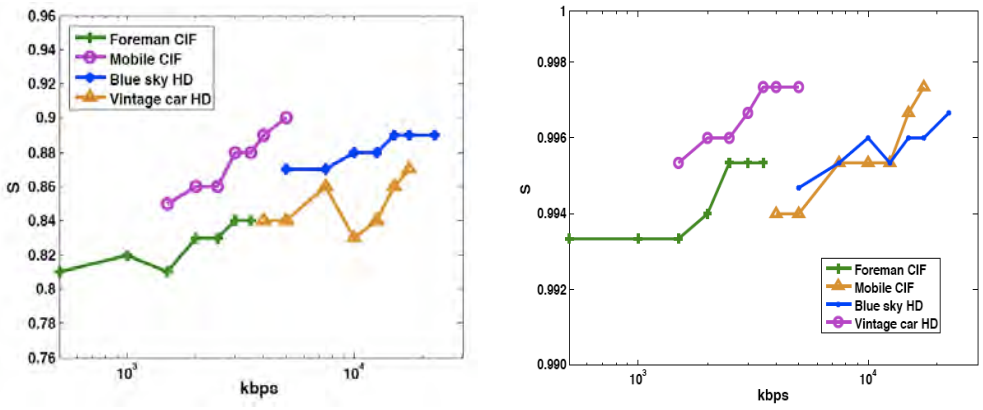


Figure 3.11: The complexity saving as a function of the coding bit rate with (a) ZDK-I and (b) ZDK-II for four sample sequences.

The complexity saving factors for ZDK-I, ZDK-II and the SCJ method are shown in Table 3.1. We see that ZDK-I, ZDK-II and the SCJ method all offer a significant amount

of complexity saving. With the help of the optimal subpel MV resolution, ZDK-II can provide additional 13-15% complexity saving. Furthermore, the complexity saving for ZDK-I and ZDK-II are shown as a function of the bit rate in Fig. 3.11. We depict the results of Foreman and Mobile CIF sequences at a lower bit rate range and those of Blue sky and Vintage car at a higher bit rate range. Generally speaking, the complexity saving is stable for a range of bit rates.

The R-D performance of various subpel MV search schemes is compared in Figs. 3.12-3.15. We use two subpel MV search schemes for performance benchmarking. They are: 1) the integer-pel MV and 2) the subpel MV with a fixed resolution (of quarter-pel or eighth-pel).

The results of eight test sequences with ZDK-I are shown in Figs. 3.12 and 3.13. For performance evaluation, we provide the rate reduction comparison in Table 3.2 and Table 3.3 based on the method described in [18]. We see that the proposed ZDK-I has very small rate increase (around 5%) as compared with the full H.264 1/4pel search. In contrast, the SCJ 1/4pel method has a larger rate increase (around 15 to 20%).

The results of eight test sequences with ZDK-II are shown in Figs. 3.14 and 3.15. We see that, with the optimal MV resolution estimation enabled, the proposed ZDK-II scheme can achieve almost the same R-D performance as the H.264 1/8pel scheme with a complexity saving factor of 99.6%.

Table 3.2: Coding efficiency comparison of the proposed ZDK-I scheme and the SCJ method with respect to H.264 with quarter pel resolution.

		ZDK-I vs. H.264 1/4pel	SCJ 1/4pel vs. H.264 1/4pel
Sequence	Resolution	ΔBit Rate (%)	ΔBit Rate (%)
Foreman	352x288	4.59	14.93
Mobile	352x288	5.89	16.37
Stefan	352x288	4.73	16.17
Flowergarden	352x288	2.37	15.84
Average		**4.40**	**15.83**
City Corridor	1280x720	4.40	16.54
Night	1280x720	5.01	14.39
Vintage Car	1920x1080	5.78	16.70
Blue Sky	1920x1080	5.60	15.02
Average		**5.20**	**15.66**

Table 3.3: Coding efficiency comparison of the proposed ZDK-I scheme and the SCJ method with respect to H.264 with eighth pel resolution.

		ZDK-II vs. H.264 1/8pel	SCJ 1/8pel vs. H.264 1/8 pel
Sequence	Resolution	ΔBit Rate (%)	ΔBit Rate (%)
Foreman	352x288	3.14	17.22
Mobile	352x288	4.28	20.43
Stefan	352x288	3.21	19.06
Flowergarden	352x288	3.56	18.69
Average		**3.54**	**18.85**
City Corridor	1280x720	3.46	17.79
Night	1280x720	4.31	19.05
Vintage Car	1920x1080	4.09	19.46
Blue Sky	1920x1080	4.40	20.83
Average		**4.07**	**19.28**

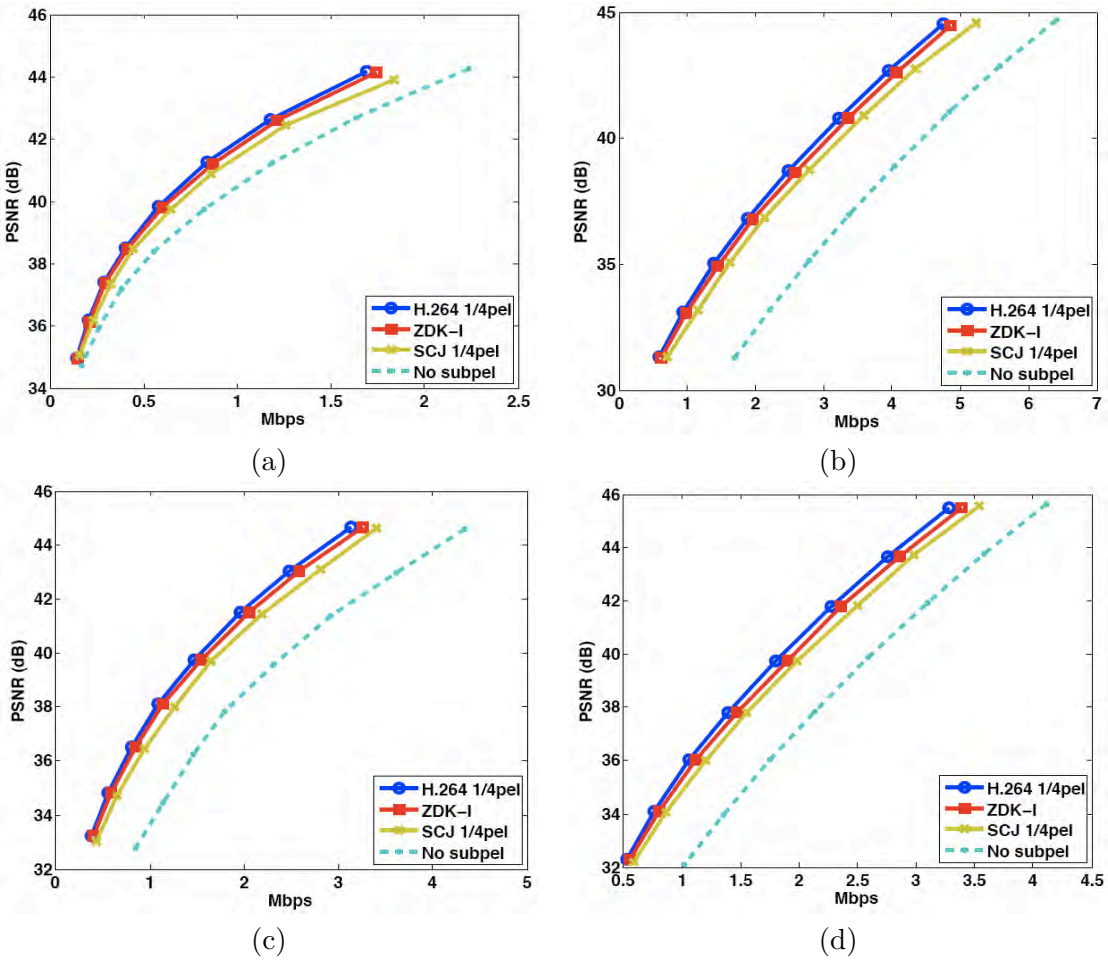


Figure 3.12: The R-D performance of ZDK-I and two benchmark methods for four CIF sequences: (a) Foreman, (b) Mobile, (c) Stefan, and (d) Flower garden.

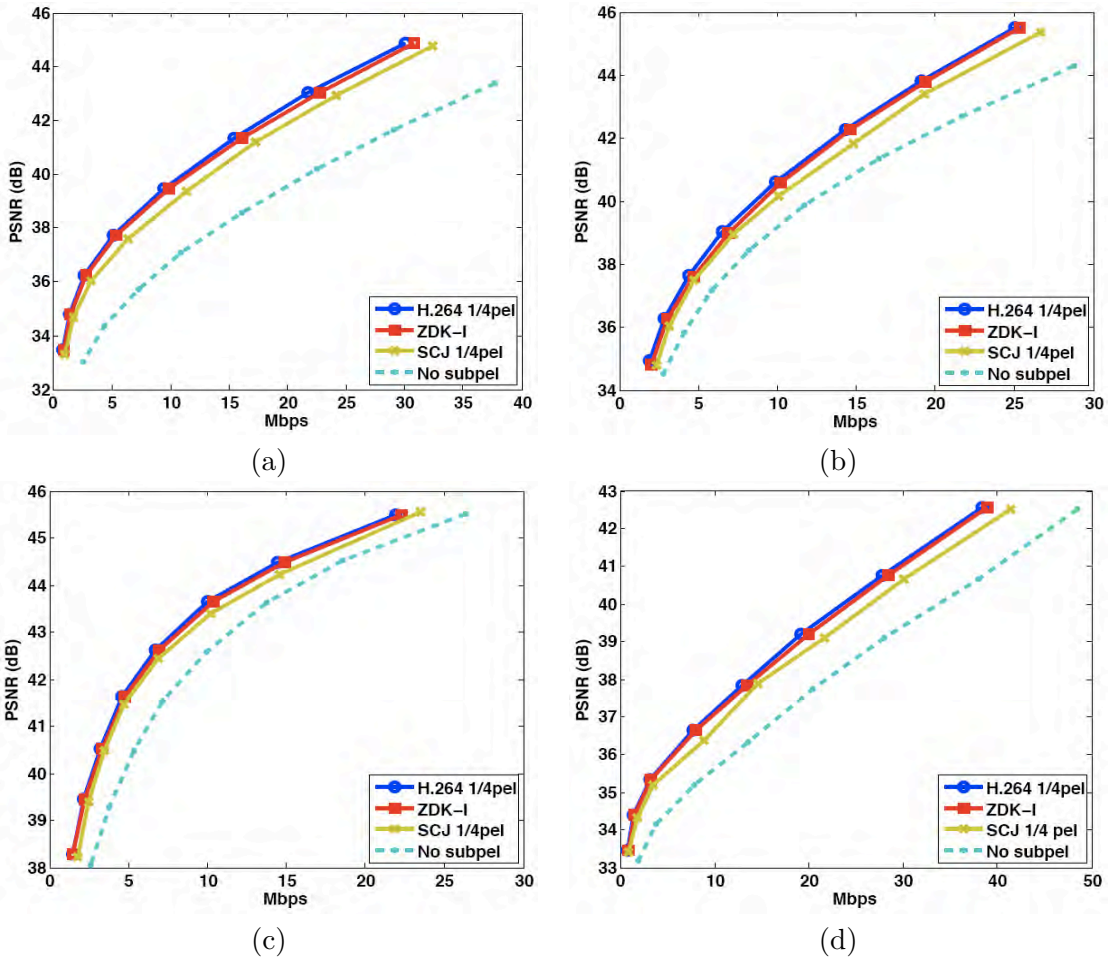


Figure 3.13: The R-D performance of ZDK-I and two benchmark methods for four HD sequences: (a) City Corridor, (b) Night, (c) Blue sky, and (d) Vintage car.

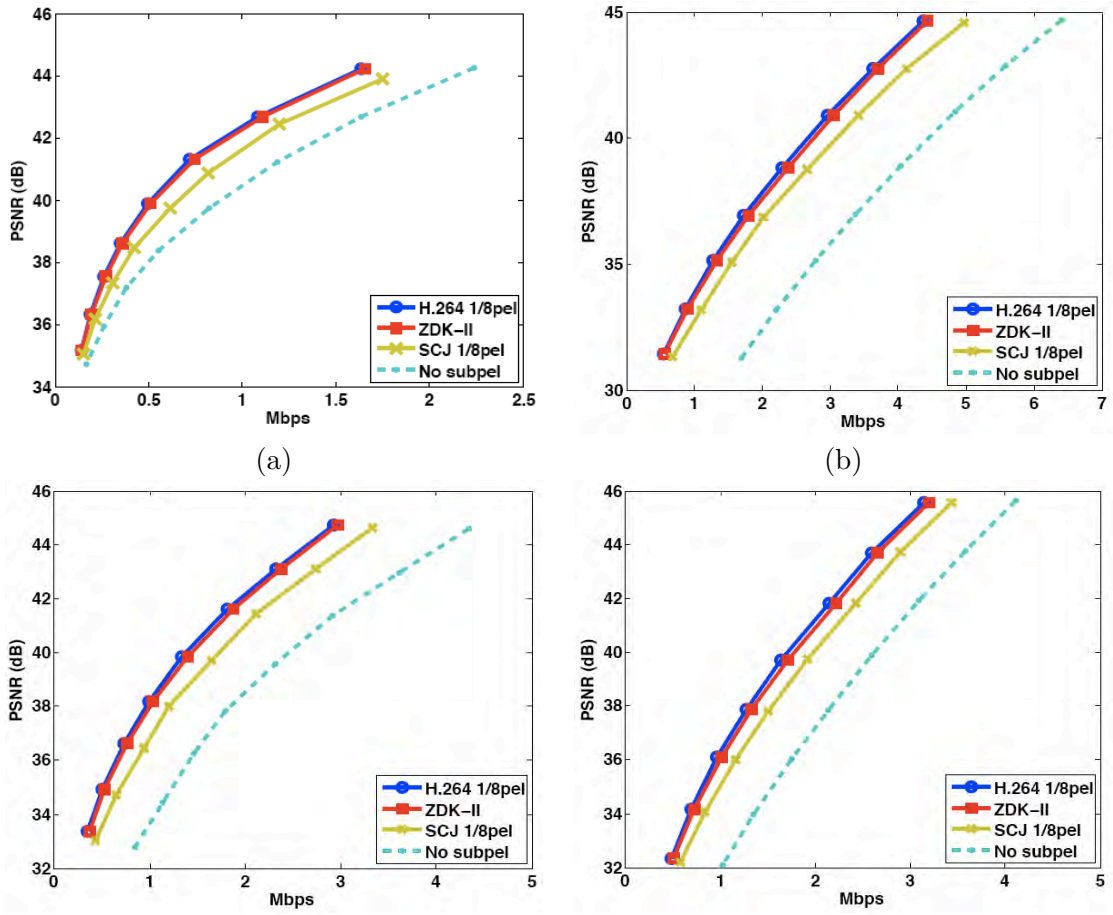


Figure 3.14: The R-D performance of ZDK-II and two benchmark methods for four CIF sequences: (a) Foreman, (b) Mobile, (c) Stefan, and (d) Flower garden.

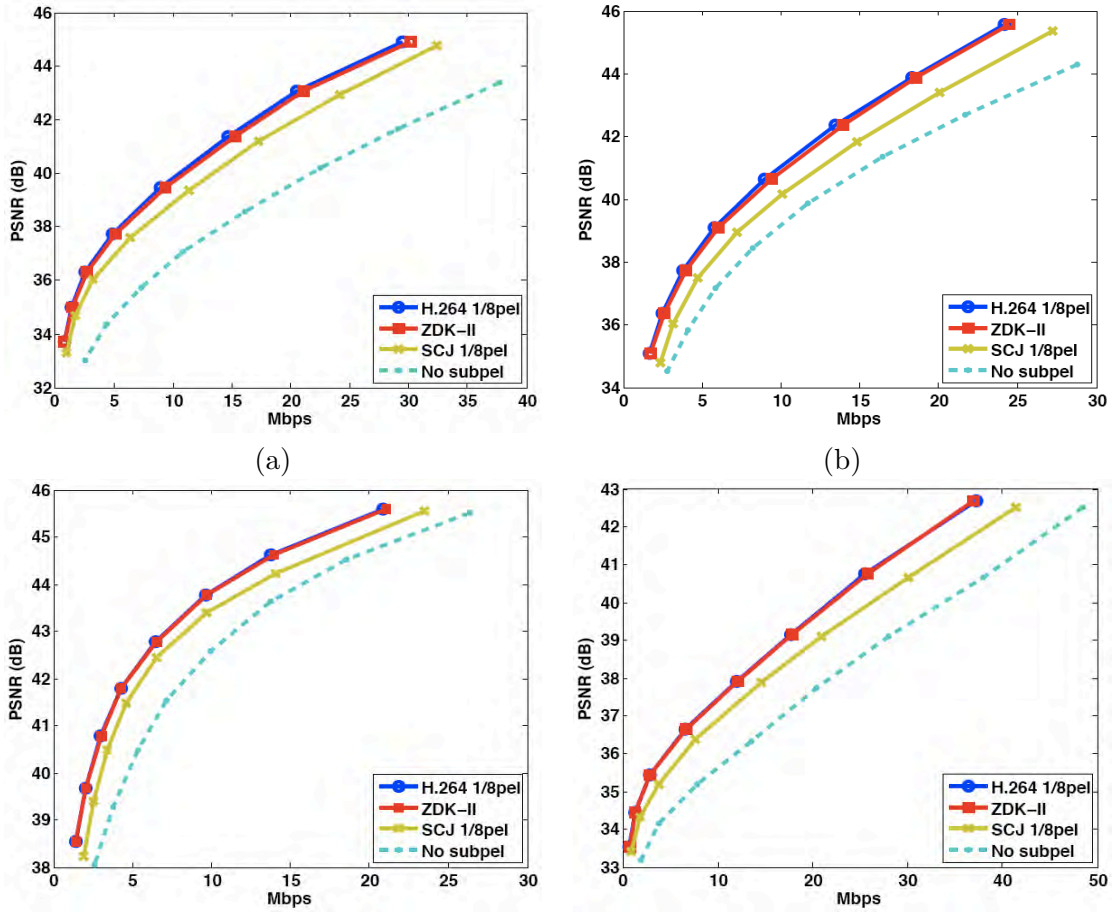


Figure 3.15: The R-D performance of ZDK-II and two benchmark methods for four HD sequences: (a) City Corridor, (b) Night, (c) Blue sky, and (d) Vintage car.

3.6 Conclusion

The behavior of the subpel MV error surface was studied and two parameters were proposed to chacterize the error surface; namely, the condition number and the deviation from flatness. These two parameters can be easily computed based on the prediction residuals at nine integer MV values centered at the minimum integer MV location. Then, an optimal MV resolution estimation scheme was derived, which allows each block to

select an optimal MV resolution adaptively based on the deviation from flatness parameter. Furthermore, two direct subpel MV position prediction schemes were described for ill- and well-conditioned blocks, respectively. It was shown by experimental results that the R-D performance of the proposed ZDK-II scheme is comparable with that of the full subpel MV search at a much lower computational complexity.

Several extensions of our current work can be explored in the future. For example, we adopt fixed thresholds on the condition number and the deviation from flatness parameters for all test sequences based on an off-line training process in this work. It may be worthwhile to investigate adaptive thresholding based on the properties of the underlying video sequences to achieve better R-D performance. Furthermore, we may consider the framework of rate-distortion-complexity (RDC) optimization and adjust threshold values accordingly to find a good balance between complexity and the R-D performance.

Chapter 4

Granular Noise Prediction and Coding Techniques

4.1 Introduction

Video compression has been extensively studied for the last two decades. Earlier research has primarily focused on low-bit-rate coding due to the limited storage space and bandwidth. Recently, research focus has shifted to high-bit-rate video due to increased popularity of high definition (HD) video and availability of broadband network infra-structure in recent years. HD video offers higher spatial resolution as well as enhanced quality (which means a higher PSNR range). To meet these requirements, the H.264/AVC standard has included the Fidelity Range Extension (FR-Ext) in its high profile to support 4k/2k contents [43].

High definition (HD) video sequences have several unique characteristics as compared to video sequences of lower resolution. First, its content has higher fidelity with more detail texture recorded to create an more involving experience to the audience. Promoted by the advancement in storage and transmission technologies, the market is gearing towards a high bit rate, high quality content coding. To satisfy these unique requirements,

H.264/AVC incorporated the Fidelity Range Extension (FR-Ext) in its high profile to support 4K/2K contents [43].

Second, HD video sequences are typically of very high resolution. The current well-known resolution is about 1920x1080 progressive. Even higher resolutions are being introduced into the market. For example, Digital Cinema Initiatives (DCI) specified the use of 2k/4k cameras [1], the latest released HD-DVD and Blu-ray disc both supports resolution of 1080p. Because of the above-mentioned unique characteristics, the main challenges associated with HD content are storage and bandwidth requirements for compression and streaming over IP networks. Compared to earlier standards, the latest compression standard H.264/AVC is able to provide a nearly 50 percent rate saving with the same PSNR requirements [46]. However, H.264/AVC initially was proposed to target at low bit rate coding environment and most of the experiments are conducted on low resolution QCIF and CIF sequences as well. Hence, as the spatial resolution increases, H.264/AVC reaches a performance bottleneck. Therefore, in JVT meetings, one of the long term objectives is to develop a new generation video coding standard that would keep abreast with this significantly increased demand for storage and transmission bandwidth.

In summary, the coding efficiency from the existing coding schemes are limited once high fidelity is required. This phenomenon indicates that there exists some unique features causing inefficiency in high-bit-rate coding environments. In this paper, we first provide a systematic analysis on the unique characteristics of this feature that we identify as granular noise and its impact on high bit rate, high fidelity video coding in Section 4.2. The analysis emphasizes the importance to treat granular noise different and separately.

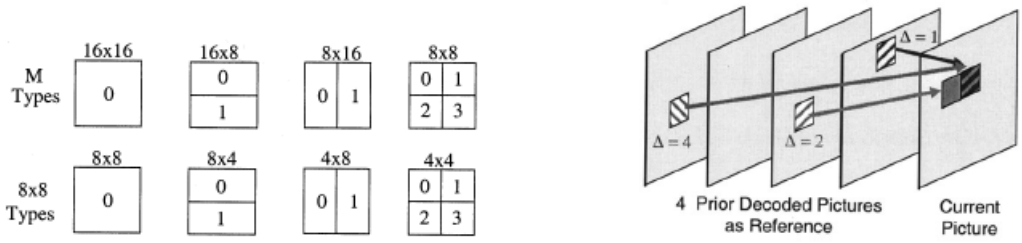
The rest of this chapter is organized as follows. In Sec. 4.3, a new granular noise prediction and coding scheme is proposed. This is an extension from our earlier proposed residual image prediction and coding (RIPC) for lossless coding [52]. This GNPC is further extended to incorporate a frequency-domain based prediction scheme to enhance the coding performance Sec.4.4. Experimental results are given to demonstrate the effectiveness of the proposed GNPC scheme in Sec. 4.5. Finally, concluding remarks are given in Sec. 4.6.

4.2 Impact of Graunular Noise on High Fidelity Coding

Due to the increased resolution in HD video, the texture complexity of a block is often simpler than that of SD video. Generally speaking, if an image has a higher correlation among pixels, its entropy will be lower and it is possible to achieve a higher compression ratio. However, we do not observe such a coding gain in existing video coding standards. One of the main reasons is the existence of granular noise in HD video. We will elaborate on this topic in this section.

4.2.1 Observations

H.264/AVC was initially proposed for low bit rate coding. It has several unique features such as the use of sophisticated multiple frame reference motion search, quarter-pel motion compensation, multiple mode selection, rate-distortion optimization (RDO) techniques. Its coding performance outperforms previous MPEG standards by a significant margin.



(a) (b)

Figure 4.1: The marcoblock partition modes and (b) B-frame prediction.

For P frame prediction, a marcoblock can be divided into smaller partitions as shown in Fig. 4.1. H.264/AVC supports luma block partitions of 16x16, 16x8, 8x18 or 8x8. One additional syntax element can be assigned to each 8x8 partition to indicate if it will be further divided into smaller sub-partitions of 8x4, 4x8 or 4x4. The motion prediction for each block is performed by searching a displacement in the reference frame. H.264/AVC also supports multi-frame motion compensated prediction, where more than one of previously coded frames can be used as the reference frame for inter prediction. For B frame prediction, more reference frames are incorporated so that a marcoblock in a B frame can use a weighted average of two distinct motion compensated prediction values to construct the prediction. In the B frame prediction, four different types of inter prediction can be used: list 0 (first list of reference pictures), list 1 (second list of reference pictures), bi-predictive and direct prediction. Being similar to the P frame prediction, the same marcoblock partitions as indicated in Fig. 4.1 are used.

Besides inter prediction modes, a SKIP mode is also introduced for extremely efficient coding. In this SKIP mode, all residual DCT coefficients of the block are quantized to zero so that neither quantized residuals nor the motion vector (or the reference index) is encoded. Only one bit is used to signal this SKIP mode. For a large area with no change

or motion, a large number of blocks can be coded efficiently by this SKIP mode using a small number of bits. The statistics given in Table 4.1 [23] show that the SKIP mode in H.264/AVC is very effective for a large QP value, where a large majority of blocks are quantized to zero and encoded as the SKIP mode.

Table 4.1: Mode distribution of blocks at QP=28 for several test CIF sequences.

Sequence	SKIP mode	Other inter modes	Intra mode
Container	98.133%	1.847%	0.020%
Foreman	53.949%	45.648%	0.403%
Mobile	54.534%	45.447%	0.019%
News	86.139%	13.861%	0.000%
Tempete	62.877%	36.541%	0.609%
Average	71.126%	28.669%	0.205%

Equipped with powerful inter-prediction tools, H.264/AVC can provide efficient coding performance for video sequences of lower resolution (*e.g.* QCIF, CIF and SD video) with a medium to coarse quantization stepsize. In contrast, the mode distribution for the coding of HD video sequences is very different as shown in Table 4.2. Due to the small quantization stepsize, the SKIP mode is rarely selected. Furthermore, intra modes are preferred over inter modes for most macroblocks. This is especially true for Riverbed and Rush Hour sequences.

Table 4.2: Mode distribution of macroblocks for HD sequences with QP=8.

Sequence	SKIP mode	Other inter modes	Intra mode
Riverbed	0%	0.086%	99.914%
Blue sky	0.064%	35.794%	64.142%
Rush hour	0%	6.642%	93.358%
Station2	0%	31.366%	68.634%
Pedestrian	8.012%	11.517%	80.472%
Average	**1.615%**	**17.081%**	**81.304%**

However, this scenario changes dramatically in high bit rate video coding. The percentages of modes used in the coding of HD sequences of resolution 1920x1080 are shown in Table 4.2. The SKIP mode is rarely used due to the small quantization stepsize. Furthermore, most macroblocks choose intra modes over inter modes as its optimal prediction mode. The percentages may go higher than 90% for Riverbed and Rush hour sequences. Then, the inter frames are coded nearly in the same fashion as an I frame.

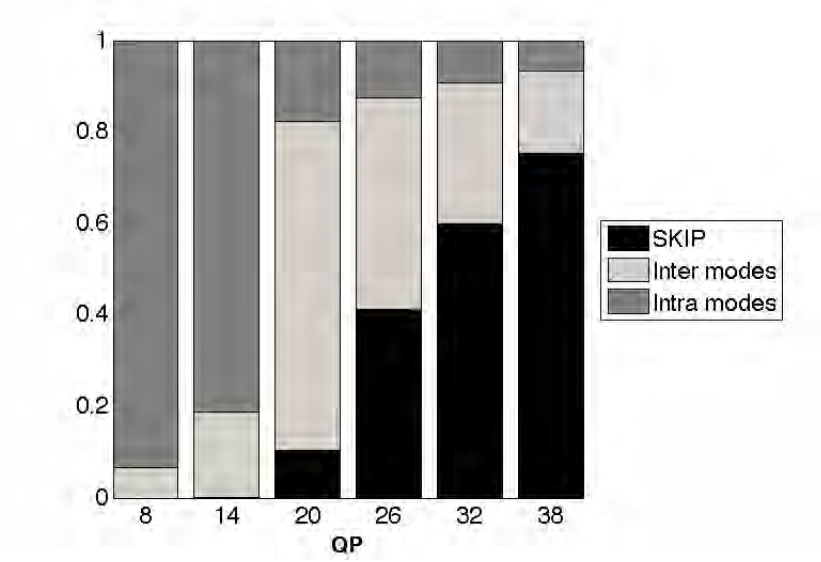


Figure 4.2: The mode distribution of H.264/AVC for Rush Hour HD sequence at various QP values.

4.2.2 Analysis

To understand the shift from the inter modes to the intra modes, we perform a detailed analysis on mode distributions with respect to a wide range of QP values for the Rush Hour sequence. We show the mode distribution for QP equal to 8, 14, 20, 26, 32 and 38 in Fig. 4.2. For the high-bit-rate coding with QP=8 or 14, intra modes are clearly

the dominant choice with its occurrence frequency higher more than 80%. However, for low and medium QP values, the occurrence of intra modes dropped significantly to 20% or lower. As the quantization stepsize becomes larger, more macroblocks can take advantages of the SKIP mode, which becomes dominant when QP becomes 32 and 38. To conclude, for the coding of HD video, the efficiency of MCP is not obvious until the QP value is close to 20 or higher, which corresponds to a medium-bit-rate coding setting. Otherwise, intra modes are the dominant choice for a smaller QP.

As the QP value becomes smaller, more details and fine textures appear in a macroblock, and the number of quantized zero DCT coefficients decreases. Generally speaking, smaller partitions have a higher probability to find a match yet they demand more header bits as the overhead. The reduced residual bits can be offset by the increased overhead bits. To obtain the best trade-off mathematically, the Lagrangian cost function is commonly used, which is also known as the Rate-Distortion Optimization (RDO) process [44] in the H.264/AVC mode evaluation.

The Lagrangian function is expressed as

$$\min\{J(blk_i|QP,m)\},$$

$$J(blk_i|QP,m) = D(blk_i|QP,m) + \lambda_m \cdot R(blk_i|QP,m), \tag{4.1}$$

where $D(blk_i|QP,m)$, $R(blk_i|QP,m)$ and λ_m are the distortion, the bit rate and the Lagrangian multiplier of block blk_i for a given coding mode m and quantization parameter

(QP), respectively. In H.264/AVC, λ_m can be expressed as a function of the quantization parameter QP:

$$\lambda_m = \alpha_m \cdot 2^{QP/3}. \tag{4.2}$$

Thus, for a given QP, the total number of bits associated with a marcoblock coded with intra mode m_I or inter mode m_P can be calculated as

$$R_{total}(m_I|QP) = R_{hdr}(m_I|QP) + R_{coef}(m_I|QP), \tag{4.3}$$

and

$$R_{total}(m_P|QP) = R_{hdr}(m_P|QP) + R_{coef}(m_P|QP), \tag{4.4}$$

respectively. Hence, the cost difference between intra and inter modes can be expressed as

$$\begin{aligned} J(m_P) - J(m_I) &= [D(m_P) - D(m_I)] \\ &+ \lambda_P \cdot [R_{hdr}(m_P|QP) + R_{coef}(m_P|QP)] \\ &- \lambda_I \cdot [R_{hdr}(m_I|QP) + R_{coef}(m_I|QP)]. \end{aligned} \tag{4.5}$$

Since the inter prediction can find a better match than the intra prediction for the CIF video, the distortion for inter modes $D(m_I)$ is usually less than that of intra mode $D(m_P)$. Therefore, in spite of the advantage of lower header bits associated with intra modes, the Lagrangian optimization process is still in favor of inter modes. For the HD

video, quantization is restricted to a small value for the high fidelity requirement. Based on high bit rate coding theory in [19], the distortion can be expressed as

$$D(m_I|QP) \cong D(m_P|QP) = \frac{\Delta^2}{12}. \tag{4.6}$$

where Δ is a small quantization stepsize. By substituting (4.6) into (4.5), we get the cost difference as

$$\begin{aligned} J(m_P) - J(m_I) &= \lambda_P \cdot R_{hdr}(m_P|QP) - \lambda_I \cdot R_{hdr}(m_I|QP) \\ &+ \lambda_P \cdot R_{coef}(m_P|QP) - \lambda_I \cdot R_{coef}(m_P|QP), \end{aligned} \tag{4.7}$$

which mainly depends on the total number of bits required to encode this marcoblock.

Generally speaking, the same α is used for both intra and inter modes, which results in $\lambda_I = \lambda_P$. Thus, if MC cannot find a good match due to external noise, texture, motion blur, etc, the inter prediction can be worse than the intra prediction. As a result, the overall cost will be decided based on the difference of header bits used in intra and inter modes. This gives an advantage to the intra prediction since it demands no bits for the reference frame and the motion vector. The above discussion studied why a majority of blocks choose intra modes over inter modes at a fine QP value for HD video.

To summarize that, there exists some fine information in a frame due to an increased resolution, which cannot be well compensated, due to the existence of film grain noise [32] and tiny surface variation in HD video. However, they do not show up in lower resolution

video since they are averaged out in a lowpass filtering process. In the following, we will propose a novel coding scheme for HD video with granular noise in Sec. 4.3.

4.3 Overview of GNPC Coding Framework

Film grain noise is a type of random optical texture from processed film. It is linked to the physical characteristics of the film and is perceived as a random pattern and normally follows a gGNPCeneral distribution statistics [8]. Film grain noise is not prominent in standard definition television format and is even less perceivable in smaller formats such as CIF or QCIF. However, these fine surface variations become much more visible once the resolution is increased to HD. In addition, film grain noise is usually one of the key elements used by artists to relay emotion or provide cue that enhance the visual perception of the scene to the audience. Sometimes, film grain size varies from frames to frames to provide different clues as to time reference and etc. Therefore, for lossless or high fidelity video coding, it is desirable to preserve the quality of the film grain noise without modifying the original intent of filmmakers. In addition, it has become the requirement in the motion picture industry to preserve film grain throughout the entire image and delivery chain.

Due to the random nature of film grain noise, it is very difficult to have efficient energy compaction solution. Therefore, conventional researches have been focusing on film grain synthesis. Film grain noise is first removed during a pre-processing stage at the encoder using a filter, then re-synthesized at the decoder end and added back to the filtered frame. As film grain is known to follow a near Gaussian distribution and therefore, instead of

coding the noise block by block, a good approximation model is composed based on the extracted film grain noise features and sent to the decoder. With the received model parameter, the decoder is able to re-synthesized noise, then added back to the decoded frames in the post-processing stage. Because only the noise model parameters are sent to the decoder instead of the real noise, the overall bit rate can be reduced significantly. Many successful algorithms on texture synthesis have been proposed and can be used on film grain noise synthesis [16, 42].

To reduce the computational complexity, Gomila and Kobilansky [10] proposed a sample based approach using a noise sample extracted from source signal and apply different transformation on it. Only one noise block is sent to the decoder in the SEI message. However, these approaches general involve duplication part of the original grain source and could suffer from visible discontinuity and repetition. Another type of film grain noise synthesis method employees the use of a comprehensive film grain database [22]. This film grain database contains a pool of pre-defined film grain values. The film grain selection process follows a random fashion corresponding to the average luminance of the block and a deblocking filter is applied to blend in the film grain. This method allows generation of realistic film grain but requires both ends to have access to the same film grain database.

In this work, we consider this type of film grain noise and other surface variations as granular noise. We attempt to exploit the spatial and temporal correlation of granular noise by another level of prediction so as to lower redundancy furthermore. In this section, we propose a new lossless granular noise prediction and coding (GNPC) scheme targeting

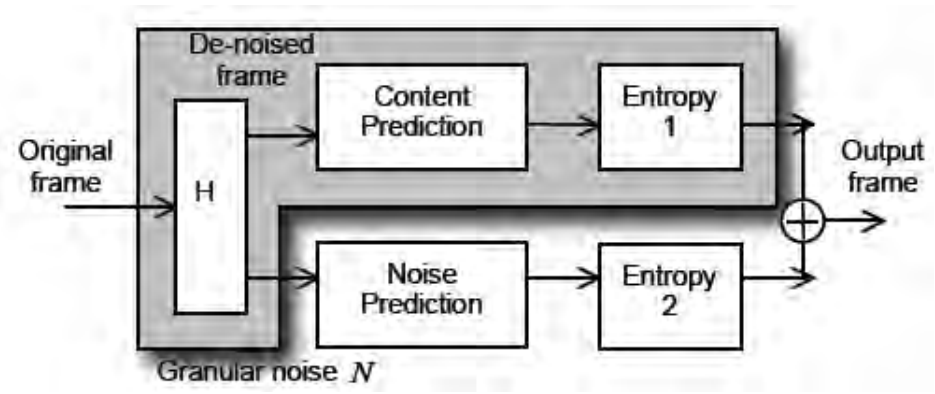


Figure 4.3: The block diagram of the proposed granular noise extraction process.

HD video/image contents. The overview of the coding system architecture is shown in the block-diagram in Fig. 4.3.

The input frame is decomposed into two parts via a de-noising technique. Then, these two parts can be coded independently. They are integrated again in the decoder end. Thus, there are two key questions in this design; namely, 1) the development of a good decomposition scheme; and 2) the design of an effective residual image prediction and coding scheme. They will be addressed below.Two different prediction schemes are performed for contents and granular noise, and their residuals are entropy coded differently in our proposed coding system.

There are many ways to extract granular noise. As shown in Fig.4.3. Here, we use a H.264/AVC based video coding process as the noise filtering process. There are several advantages associated with this proposed scheme. First, it can be easily integrated with any traditional video codec. No modification is required to the existing codec. Decoder could discard the transmitted noise frame if the decoding time frame is not sufficient. This would not affect the decoding of the consecutive incoming content frames, as they

are coded independently. Second, current lossless and lossy systems are designed with completely different prediction schemes, which are inherently mutually exclusive. Therefore, to have a system that can produce both lossless and lossy results, the hardware designers need to have two independent modules to achieve that. In contrast, with the proposed scheme, one system can achieve both lossless and lossy goals by only turn off the residual image prediction module. Third, the proposed system can achieve scalability with minimal modification (*e.g.* to quantize prediction errors of the residual image before entropy coding). Fourth, encoder no longer needs to encode different versions of bitstreams if the decoders ranges from mobile device to HDTV sets. Encoder only needs to encode once and it depends on the decoder to decide which granular noise layers are not needed. This could potentially save the storage and streaming bandwidth significantly.

For example, consider to encode a video program with lossy H.264/AVC. Then, for an input frame F, we first encode it with the H.264 encoder with a medium coarse QP. Then, the difference between reconstructed frame F' and original frame F is the extracted noise frame denoted by N. There is a tradeoff between bits assigned to the coding of F' and the coding of N, depending on the selection of QP.

4.4 Granular Noise Prediction and Coding in Frequency Domain

We exploit the frequency correlation of granular noise by another prediction so as to remove redundancy furthermore. In this section, we first review the frequency domain

prediction techniques and then discuss the proposed GNPC in frequency domain coding method.

4.4.1 Review of Frequency Domain Prediction Techniques

Most prevalent motion estimation (ME) and motion compensation (MC) algorithms used in image and video compression areas are based on block matching techniques in the spatial domain. An alternative ME scheme is to estimate the cross-correlation function in the frequency domain [37]. The frequency spectrum of the input can be normalized to give a phase correlation. However, the correlation performed by a DFT-based method corresponds to a circular convolution rather than a linear one, and the correlation function could be affected by the edge effect. To reduce the edge artifact, Kuglin and Hines [9] proposed to use zero padding to the input data sequence at the cost of higher complexity. Another technique is to use a transform size which is much larger than the maximum displacement considered [41]. This approach can limit the error size, but it is more suited for global ME rather than block-based ME. A third technique is to use the complex lapped transform (CLT) to perform the cross correlation in the frequency domain [49]. Since the basis functions are overlapped and windowed by a smooth function that shapes like a half cosine, it introduces less block edge artifacts as compared to the LOT in the spatial domain. The latest effort of prediction in the frequency domain was proposed for intra prediction in VC-1. The DC and the AC components are predicted from their left and top neighboring frequency components.

However, the above-mentioned schemes are not widely used for several reasons. First, most previous video compression algorithms focus on low bit rate coding. With efficient

spatial motion estimation and coarse quantization, most block DCT coefficients are quantized to zero. Hence, there is little room left for the rate-distortion improvement with the frequency domain prediction. Second, all frequency components are compensated simultaneously with the same spatial offsets, which is similar to the motion compensation in the spatial domain except that frequency components are compensated rather than pixel values. Then, if there is little correlation in the spatial domain, it is difficult to get a better prediction in the frequency domain.

4.4.2 Granular Noise Prediction in Frequency Domain

As studied in previous section, granular noise consists of high frequency components. Hence, if the quantization step size is too fine to reduce them to zero, the coding performance suffers since the entropy coder is optimized with respect to long runs of zeros. In the new scheme, a target block first goes through the DCT and quantization and, then, the quantized DCT block is subject to the following two prediction modes.

- The full_mode

  The target DCT block is predicted by its candidate DCT blocks with their AC and DC components completely aligned.

- The par_mode

  The target DCT block is further partitioned into four frequency partitions of size mxm, where $M = 2m$. From low to high frequencies, each partition is named as np_0, np_1, np_2 and np_3, respectively, as shown Fig. 4.4(a).

An additional mode, call the zero_mode, is proposed to further improve coding efficiency. If the residual of the prediction in the full_mode results in all zero coefficients within a DCT block, this block is coded as the zero_mode, where no prediction residual will be coded.

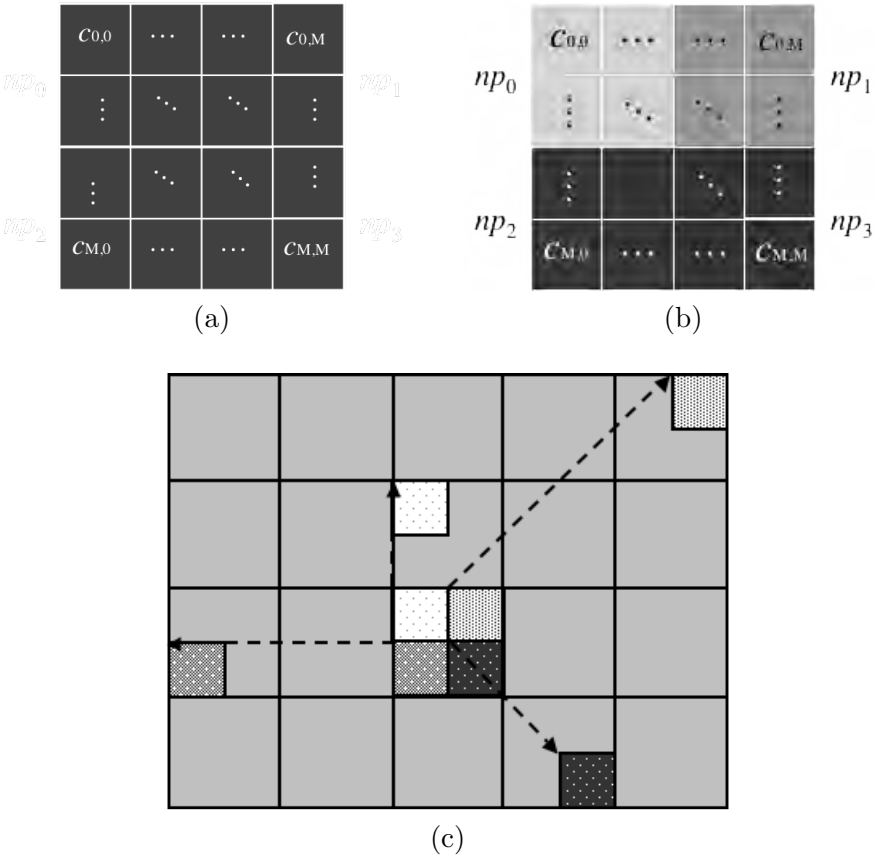
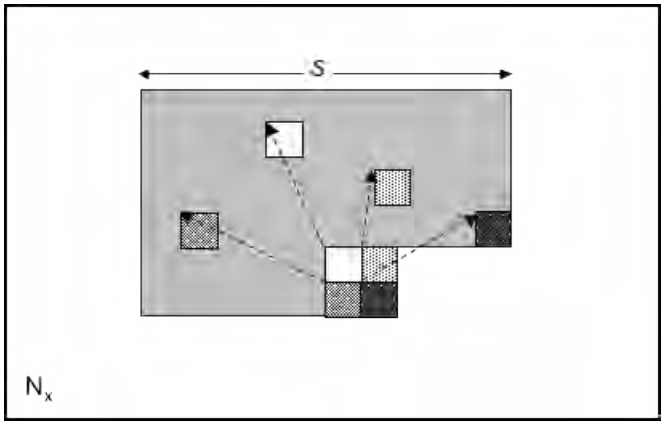
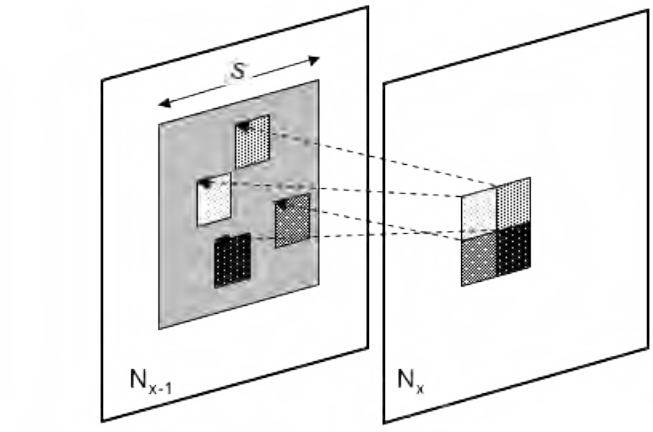


Figure 4.4: Granular noise block in frequency domain partition (a) full mode, (b) par mode and (c) prediction alignment for par mode.

Each partition np will be independently predicted from its own corresponding partition of candidate blocks. Hence, the predicted DCT block could be composed by partitions from different reference blocks. As prediction errors are the differences in the frequency domain, they do not demand any additional DCT or quantization operations and can be



(a)



(b)

Figure 4.5: The DCT-domain based granular noise prediction for (a) intra noise frame and (b) inter noise frame with search range S.

sent directly to the entropy encoder. The granular noise prediction for intra noise frame and inter noise frame are illustrated as in Fig. 4.5. When the par_mode is chosen for a transformed and quantized granular noise block, there are four displacement vectors pointing to four predicted partitions of candidate blocks in the reference frame so that the number of overhead bits could be higher. On the other hand, since frequency bands should be well aligned between the target and the reference blocks, the unit of search stride should be the same of the block width (or height).

4.4.3 Rate-Distortion Optimization

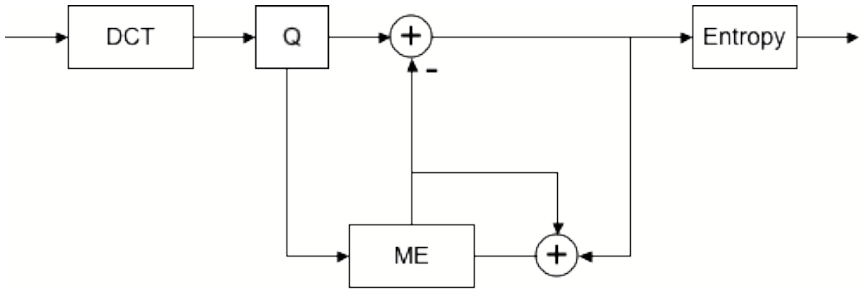
To select the best mode for the coding of a GN block, we can employ the Lagrangian Rate-Distortion Optimization (RDO) technique as

$$\min\{J(blk_i|m)\},$$
$$J(blk_i|m) = R_{hdr}(blk_i|m) + R_{coef}(blk_i|m). \tag{4.8}$$

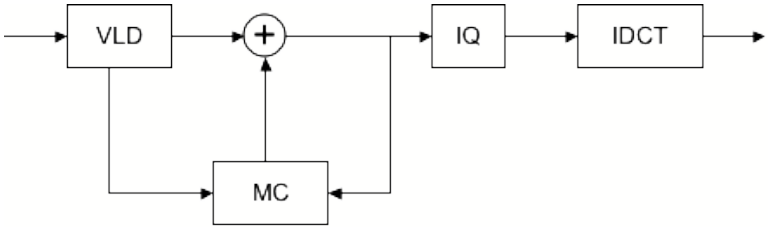
As the prediction is performed on the already DCT transformed block, the residual is in fact in the form of quantized and transformed DCT coefficients. As a result, the distortion can be taken out from the Lagrangian optimization formula and the entire process can be simplified to a rate optimization process. Based on the rate estimation given in [29], the total number of bits needed to code prediction error R_{coef} is modeled as

$$R_{coef} = \gamma \cdot \frac{SATD(QP)}{Q^p}, \tag{4.9}$$

where γ is a model parameter and p is a frame type dependent value. In our case, we use $p = 1.0$. usually represents the transformed coefficients of the prediction error within a block. The total number of bits needed to encode the GN block is a sum of the prediction error R_{coef} and header bits R_{hdr}. A simple block code of bits $log_2 S$ are assigned to the pair of displacement vectors. We will explain more details in the next subsection. This simplified RDO process helps to reduce the computational complexity in rate control for the video streaming application. The block diagram of the high fidelity GNPC scheme is shown in Fig. 4.6.



(a) Encoder



(b) Decoder

Figure 4.6: The block diagram of (a) the encoder and (b) the decoder of the proposed GNPC scheme for high fidelity video coding.

There are a few advantages with the proposed GNPC in frequency domain scheme. First, by correlating a sub-band of the target DCT block with those of different DCT blocks, we can enhance the matching probability and reduce the energy of prediction

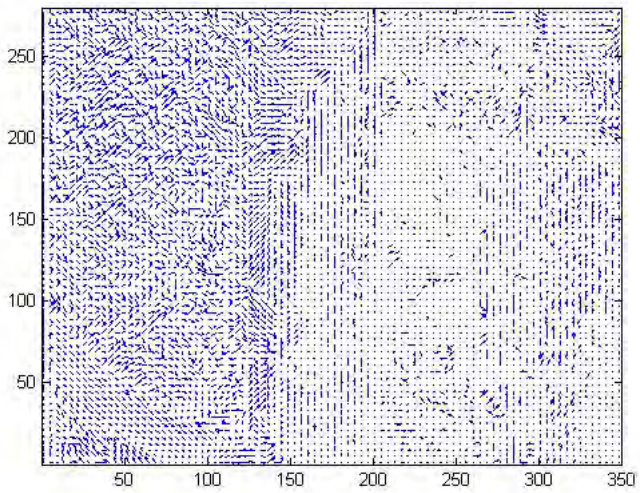
errors. Second, by quantizing the input DCT block before the prediction process, there will be no additional computation needed in the rate-distortion optimization phase and complexity can be significantly reduced. Third, the rate of the proposed scheme can be adjusted more easily in video streaming applications. As quantization is done prior to prediction, the distortion/PSNR of each block/frame can be known ahead of time without going through the entire RDO process[29].

4.4.4 Translational Index Mapping

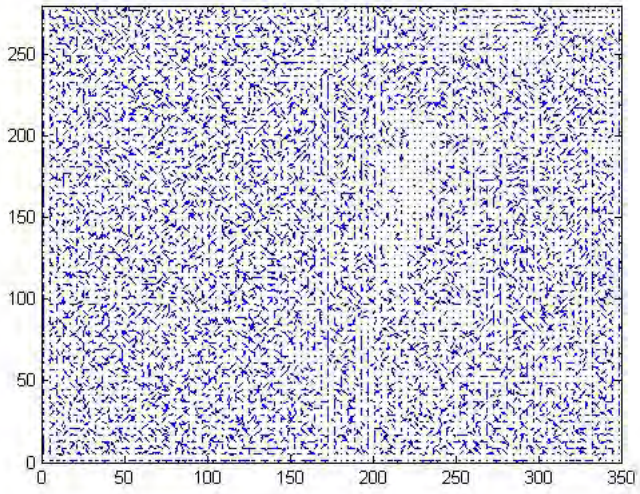
In the proposed frequency domain-based GNPC scheme, when the par_mode is chosen for a transformed and quantized GN block, there are four displacement vectors pointing to four predicted partitions of candidate blocks in the reference frame so that the number of overhead bits could be higher.

The proposed frequency-domain GNPC scheme has to encode translational vector pairs $(\delta_x, \delta_y)^T$ within the DCT domain to indicate the best match location. For the full_mode, one prediction error block ϵ^T plus one set of $(\delta_x, \delta_y)^T$ are needed. For the par_mode, one prediction error ϵ^T together with four sets of $(\delta_x, \delta_y)^T$ are needed since each partition has its own unique $(\delta_x, \delta_y)^T$ in the par_mode. Thus, the translational vector cost will be higher if the par_mode is chosen as the best mode. In addition, due to the random nature in the frequency domain-based prediction, classic DPCM-based translational vector prediction does not bring efficiency into the coding of the translational vectors. See Fig.4.7.

Hence, to limit the overhead cost from the use of four pairs of translational vectors, one translational index Δ^T is used to replace each pair of translational vectors. Each unit



(a)



(b)

Figure 4.7: The translational vector maps for (a) the content layer and (b) the granular noise layer for Rush Hour frame at a resolution of 352x288.

of Δ^T is equivalent to a certain distance in both horizontal and vertical directions. The translational indexing method is developed by modifying the circular zonal techniques [4]. Each zone is color differentiated as shown in Fig. 4.8.

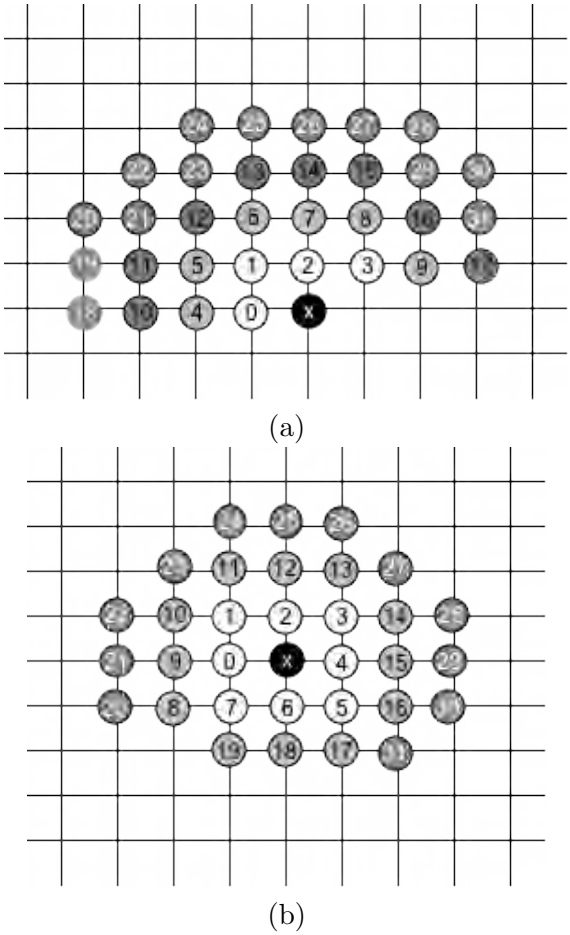


Figure 4.8: Illustration of the translational indexing scheme for (a) the intra GNPC frame and (b) the inter frame in frequency domain based GNPC.

As for the intra GN prediction, the zone is not completely circular because of the uncoded blocks in the zone and the indexing scheme represents a half zone and the indexing always starts from the left side of the target block with its equivalent $\delta_x = 0$ to

maximize the spatial correlation between the target block and the candidate blocks [15].

Table 4.3: Experimental setup for H.264/AVC and the GNPC scheme.

Parameter	H.264/AVC	GNPC in frequency domain		
		Content	Intra	Inter
Profile	High	baseline	n/a	n/a
QP	4,8,10,12,14,16,18	25(fixed)	4,8,10,12,14,16,18	
GOP	15	15	n/a	15
# of reference frame	5	1	na	1
RDO	Full complexity	Low complexity	Fast	Fast
Subpel ME	1/4 pel	na	na	na
Search range	64	32	32	32
Deblocking filter	Enabled	Disabled	N/A	N/A
Entropy	CAVLC	CAVLC	CAVLC	CAVLC

4.5 Experimental Results

In the experiment, we conducted experiments to compare the performance of H.264/AVC [2] and the proposed GNPC scheme for high fidelity video coding. Only the luminance channel is compared in this experiment. Four HD YUV sequences were used in the experiments, namely Rush hour(@1920x1080), Blue sky(@1920x1080), Sunflower(@1920x1080) and Vintage car(@1920x1080).

The results were averaged over 10 frames for each sequence. For H.264/AVC, we chose the high complexity RDO process, multiple reference frames, the quarter pel motion estimation, and the deblocking filter option. In contrast, for the high fidelity GNPC scheme, we set the encoder to the lowest complexity such as low complexity RDO, 1

reference frame, no B frame, no subpel search and no deblocking filter. More details of the experimental setup are given in Table 4.3.

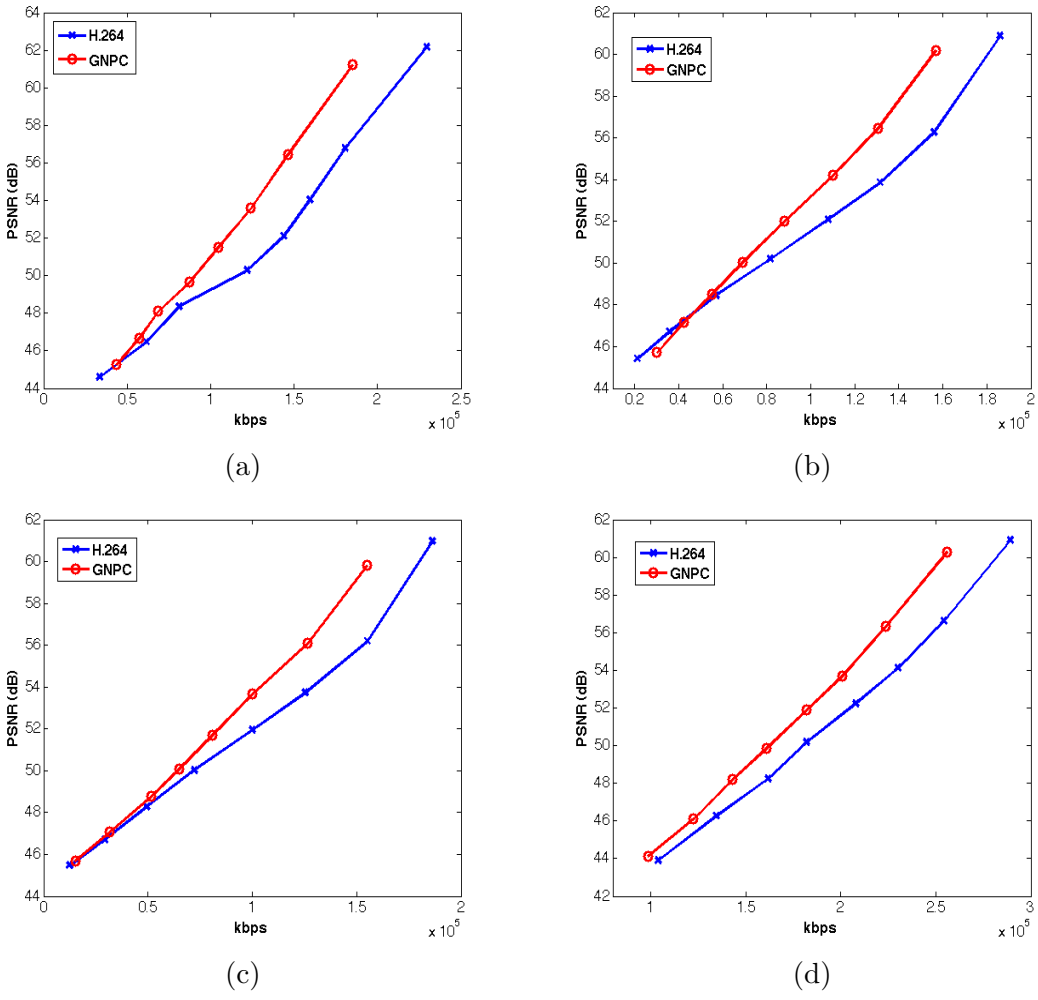


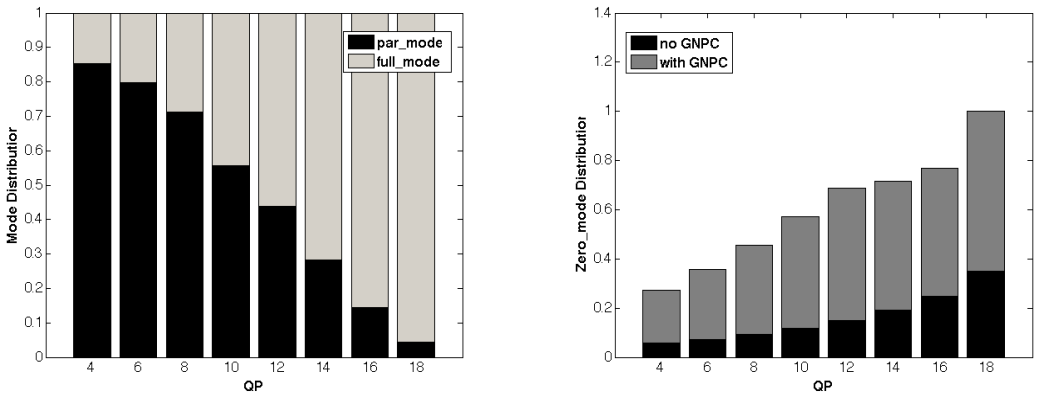
Figure 4.9: Rate-Distortion curves for HD video sequences (a) Rush Hour, (b) Blue Sky, (c) Sunflower and (d) Vintage Car.

The rate and distortion curves are shown in Figs. 4.9. The Δrate and Δdistortion difference are presented in Table 5.1. They are calculated based on the formula given in [18]. It was observed in [10] that a coarser quantization (with $QP > 18$) could reduce granular noise to the minimal. Here, we used a quantization stepsize range from 4 to

18 in the context of near-lossless coding. Results in Figs. 4.9 to **??** confirm a similar trend; namely, granular noise is gradually suppressed by the quantization effect. The performance of the proposed GNPC scheme and H.264/AVC converges for video sequences of simplex content at lower bit rates. The proposed GNPC provides a higher coding gain for highly complex sequences such as Vintage Car.

Table 4.4: Coding efficiency comparison between H.264 and GNPC in the frequency domain.

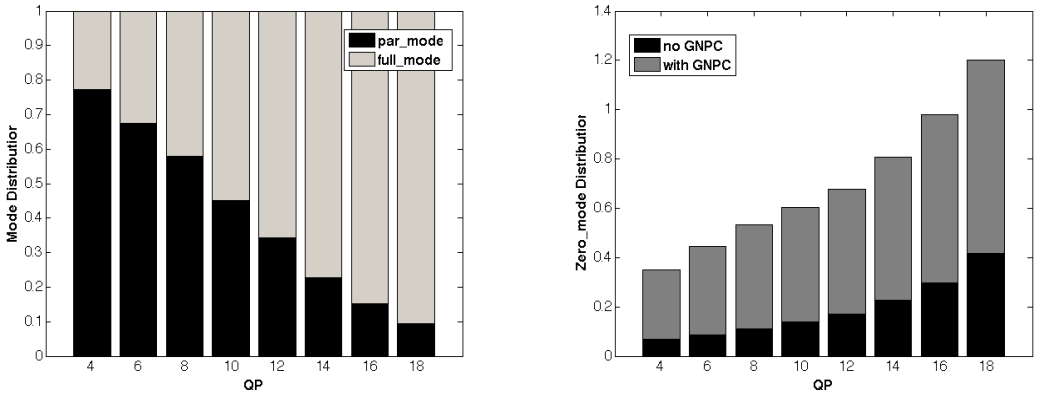
Resolution	Sequence	ΔBit Rate (%)
1920x1080	Rush Hour	-11.85
	Blue Sky	-7.20
	Sunflower	-11.54
	Vintage Car	-9.73
Average		**-10.08**



(a) (b)

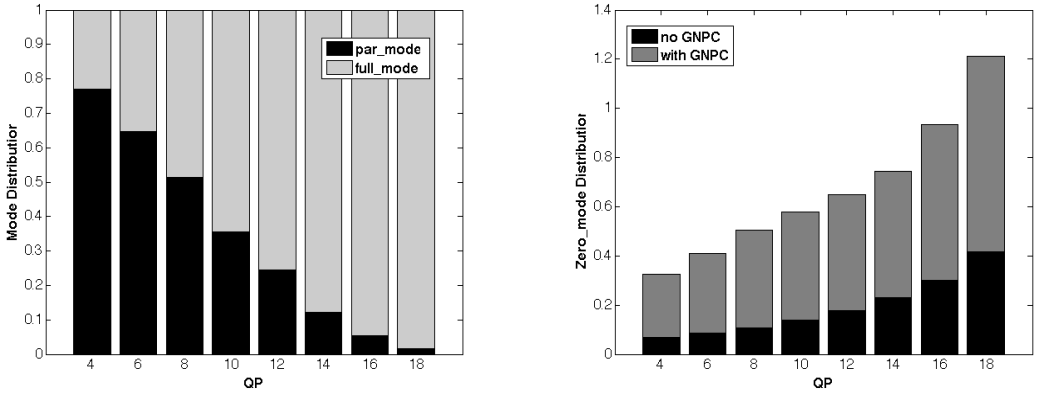
Figure 4.10: The mode distribution chart for the Rush Hour sequence with (a)full mode vs par mode and (b)zero mode distribution with and without GNPC in the frequency domain.

The mode distribution charts in subfigure (a) of Figs. 4.10-4.13 show that there are more blocks adopting the *par_mode* in the GNPC scheme if the quantization stepsize



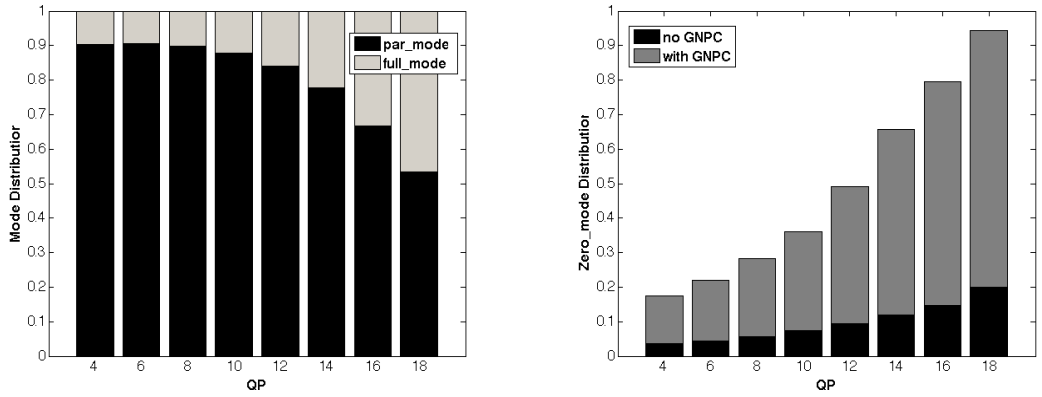
(a) (b)

Figure 4.11: The mode distribution chart for the Blue Sky sequence with (a)full mode vs par mode and (b)zero mode distribution with and without GNPC in the frequency domain.



(a) (b)

Figure 4.12: The mode distribution graph for the Sunflower sequence with (a)full mode vs par mode and (b)zero mode distribution with and without GNPC in the frequency domain.



(a) (b)

Figure 4.13: The mode distribution chart for the Vintage Car sequence with (a)full mode vs par mode and (b)zero mode distribution with and without GNPC in the frequency domain.

becomes smaller. By the *zero_mode* distribution charts in subfigure (b) of Figs. 4.10-4.13, we count the number of blocks that meet the criterion of the *zero_mode*. Note that this *zero_mode* has no counterpart in H.264/AVC. Thus, we only signify it as *no GNPC*. These charts show a larger percentage of blocks exploits the efficiency from the *zero_mode* enabled by the near-lossless GNPC. The above experimental results clearly demonstrate the effectiveness of the proposed GNPC scheme with an average bit rate gain of 10%.

4.6 Conclusion

In this chapter, we first conducted an analysis on existing video codecs to show that they are effective for high-bit-rate coding. We pointed out that the existence of granular noise could be the main reason for the coding inefficiency of existing coding techniques, and proposed a granular noise prediction and coding scheme. Furthermore, we proposed a novel prediction scheme for the GN coding based on the frequency domain prediction. The resultant scheme allows efficient prediction and coding at a low computational complexity. The proposed GNPC scheme can outperform H.264/AVC by an average of 10% bit rate reduction in the high-fidelity coding case.

Chapter 5

Multi-Order Residual (MOR) Coding

5.1 Introduction

Due to the constraint of communication and computational resources, traditional video compression algorithms have focused on the low bit rate video coding. The state-of-the-art video coding standard, H.264/AVC, was initially developed to target at the low-to-medium bit rate applications. However, with the popularity of high definition video such as high definition TV (HDTV) and the blue-ray disk (BD) in recent years, an effective high bit rate coding scheme becomes more and more important. The need to store/transmit high resolution video with high fidelity imposes a great challenge on the video coding technology.

To address this requirement, H.264/AVC has the Fidelity Range Extension (FR-Ext) in its high profile to support the coding of 2k/4k contents [43]. However, as explained in the last chapter, due to the existence of uncompensated fine structured features in the prediction residual, most prediction/compensation techniques used in H.264/AVC were ineffective in the high bit rate region. This results in coding efficiency degradation as

the quality requirement increases. It was claimed in previous work [10, 32] that these fine features exhibit a behavior similar to film grain noise. Some coding schemes were developed based on the idea of film grain noise synthesis [16, 10, 42, 8, 32]. Although these methods can improve the compression ratio, they are not widely adopted by the industry or the coding community due to the significant loss in the objective quality measure.

To address this problem, we introduced a coding scheme called the granular noise prediction and coding (GNPC) was proposed in Chapter 4.

As the analysis conducted in previous chapter was mainly based on the existing codec behavior, in this Chapter, we would like to provide a more thorough investigation on the target signal characteristics. We hence further investigate the impact of the high-bit-rate requirement on coding efficiency from two angles. First, we study the distribution of prediction residuals in form of DCT coefficients. Second, we conduct a correlation analysis on different video scenes to understand the long-, medium- and short-range correlations in the input video frame. Based on the analysis, we propose a new coding approach called the multi-order-residual (MOR) coding in Sec. 5.3 [54]. This MOR is a generalized and improved scheme based on our earlier proposed SOR scheme [53]. As compared with the previously proposed GNPC method, the MOR approach extract different types of correlation from the first-order prediction residuals in multiple stages based on the concept from numerical analysis. It is worthwhile to point out that the MOR approach is different from quality-scalable video coding since it allows different coding methods used in different stages (including different prediction, transform and entropy coding techniques) to achieve better overall coding efficiency. Experimental results are provided

in Sec. 5.4 to demonstrate the effectiveness of the proposed MOR coding approach. Concluding remarks and future directions are given in Sec. 5.5.

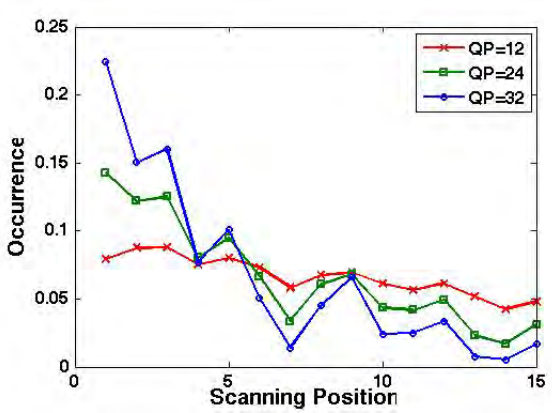
5.2 Signal Analysis for High-bit-rate Video Coding

In this section, we will examine how the original signal characteristics impacts on the codec design for high-bit-rate coding requirements.

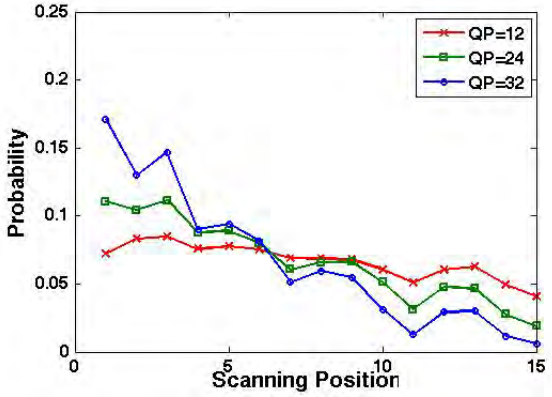
5.2.1 Distribution of DCT Coefficients

Generally speaking, a coding scheme can be classified into two distinct phases: *modeling* and *coding*. In the modeling phase, the spatial and temporal redundancy of the input video data is removed via transform and/or prediction and the statistics about the prediction residual is then gathered to form a probabilistic model [38]. It is one of the most fundamental pieces in data compression. In earlier research, the Gaussian distribution is often used to describe the distribution of AC coefficients [38]. However, it was soon found that the Laplacian distribution is more suitable to describe the signal statistics when the Kologorov-Smirnov goodness-of-fit test is used [17, 19]. Recent studies on the coding of standard definition video with H.264/AVC also reveals that the AC coefficients distribution is Laplacian-like and their probability distribution is skewed with a large zero peak after a large or medium quantization step-size is used [30]. This property is utilized to design the zigzag scanning order and entropy coding modules.

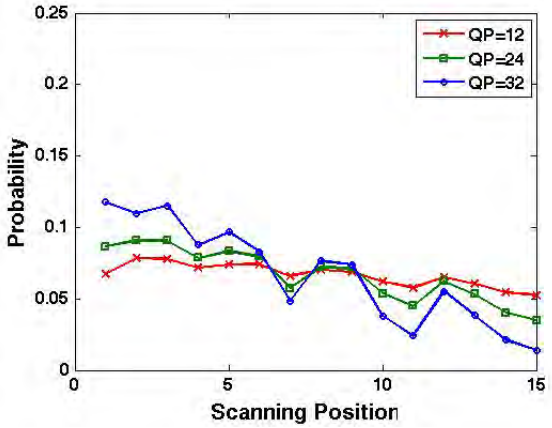
In this section, we take another look at the DCT coefficients distribution under a higher fidelity requirement since this requirement is accomplished through the use of finer quantization step sizes in today's video codec. We first analyze the effect of the



(a)



(b)



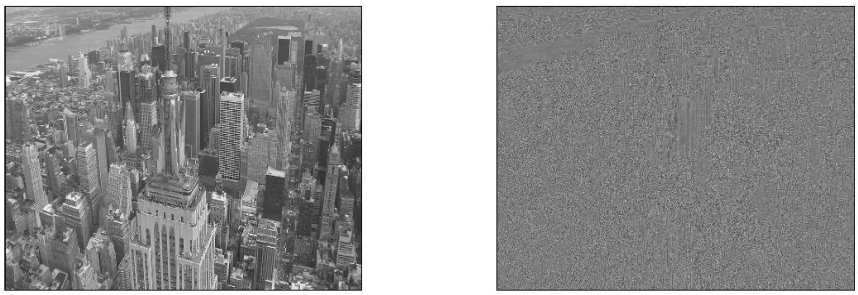
(c)

Figure 5.1: The probability distribution of non-zero DCT coefficients at each scanning position for (a)Jet (b)City Corridor and (c)Preakness frames.

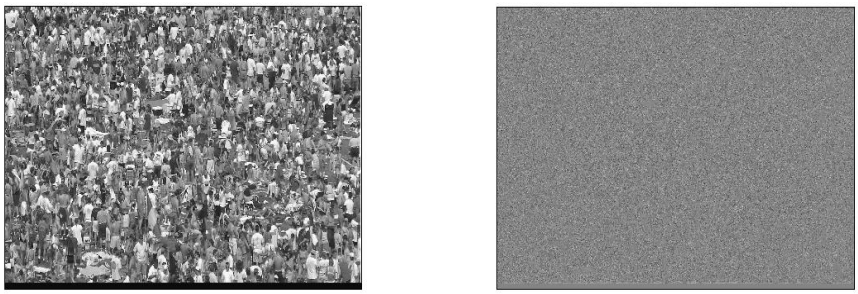
quantization parameters (QPs) for a block of size 4 by 4 on DCT coefficients distributions in H.264/AVC. In Fig. 5.1 we show the probability distribution of nonzero DCT coefficients of prediction residuals after the motion compensated prediction (MCP) process as a function of the scanning position with different QPs for three different types of video frames. position 1 is the DC coefficient, and position 2 through 16 are zigzag-scanned AC coefficients. We see that when a coarser quantization stepsize is used (*e.g.*, QP=32), higher frequencies (*e.g.*, with the position higher than 10) are all quantized to zero with most nonzero coefficients concentrated in the lower frequency region (scanning position smaller than 5) for all three sequences. This distribution is consistent with the Laplacian distribution assumption with a large zero peak, which indicates that the MCP process is effective with respect to coarse QPs.

The prediction residual has been properly processed to allow the following coding modules to encode the remaining nonzero coefficients. To be specific, the zigzag scanning process can be used to compact most zeros with a "end-of-block" symbol and the entropy coding module is effective in the coding of non-zero coefficients. However, we further observe that as the QP becomes finer to meet the higher fidelity requirement, the previously skewed distribution becomes increasingly uniform for each scanning position. In the case of QP=12, all three sequences have a close to uniform distribution of nonzero coefficients. In this case, we see that the Laplacian distribution with a large-zero peak is no longer a suitable model under high-bit rate coding. Most coefficients are non-zero and their probabilities become similar. This is a clear indication that the existing MCP process can no longer offer effective prediction for high-bit-rate video coding.

(a)



(b)

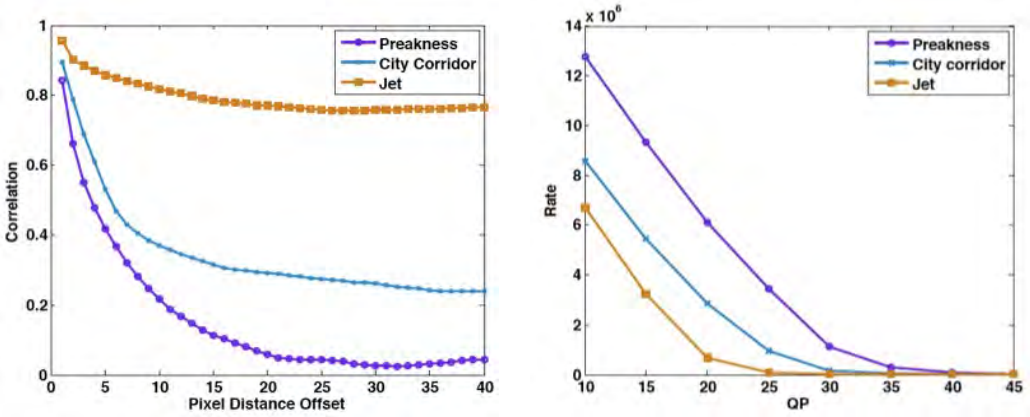


(c)

Figure 5.2: (a) A sample frame from the Jet sequence, and its prediction residual difference (b) a sample frame from the City Corridor sequence, and its prediction residual difference and (c) a sample frame from the Preakness sequence, and its prediction residual difference at 1280x720 resolution with $QP_1 = 10$ and $QP_2 = 30$.

5.2.2 Correlation Analysis

To further analyze this change in nonzero coefficients distribution, we examine the prediction residuals generated by H.264/AVC MCP. We take one sample frame from each of the three sequences we used in Fig. 5.1 and encode them with two different quantization parameters as shown in Fig. 5.2. We see that not only the residual difference images contain some untreated features very small sizes but also the amount of these untreated small structural features are directly related to the complexity of the input video frame.



(a) (b)

Figure 5.3: (a) The correlation analysis for scenes with different complexities and (b) the relationship between the bit rate and quantization.

We perform a correlation analysis on the input video signal in the context of high-bit rate coding based on the idea in [5]. Again, we use the Jet, City Corridor and Preakness sequences as examples. Note that the Jet sequence contains a scene of an airfield which is mainly still background with little detail. As shown in Fig. 5.3(a), the correlation analysis for such a low complexity scene reveals that the correlation remains very strong (> 0.9) even when the pixel distance offset has been increased steadily to 40 pixels. In other words, its frame mainly consists of long-range correlation. For a typical frame of

the City Corridor, which has medium complexity, we see that the correlation drops below 0.4 when the offset distance is greater than 8 pixels. This shows that the City Corridor frame contains a larger percentage of medium to short range correlations. For a Preakness frame, which is a highly complex scene, we observe that the overall correlation diminishes very quickly, which indicates that the preakness frame is dominated by the short range correlation. Recall the residual differences observed in Fig. 5.2. We see that different correlations inside a frame impacts the amount of fine structural features that cannot be well compensated by the current codec.

We further plot the rate-QP curves in Fig. 5.3(b) for three exemplary sequences to understand the impact of the correlation analysis on the video codec. When the QP is coarse (say, $QP > 35$), all three frames can be coded effectively. For a medium value of QP (say, $20 < QP < 35$), the Jet sequence can still be effectively encoded while the coding bit rate of the Preakness increases very quickly. In the high-bit-rate range with $QP < 20$, we see a huge rate increase in all three sequences. This observation can be explained as follows. The traditional MCP process is designed to remove the long-range correlation via block-based prediction with a search window. The neglected medium- and short-range correlations do not play an important role due to the use of a coarser QP. Thus, the overall coding efficiency is high in the low-bit-rate coding application. As the coding bit rate increases and the QP becomes finer, the quantization can no longer remove the medium and short-range correlations effectively. Thus, the overall coding gain drops significantly even for the Jet sequence of low complexity. The above analysis indicates a need for a new codec design that can efficiently remove the long-range correlation as well

as the medium- and the short-range correlations since the latter has a great impact on the high-bit-rate video coding.

5.3 Multi-Order-Residual (MOR) Prediction and Coding

Based on the analysis in Sec. 5.2, we propose a coding system that removes different correlations in the input sequence with multiple residual layers. Hence, it is called the multi-order residual (MOR) prediction and coding scheme.

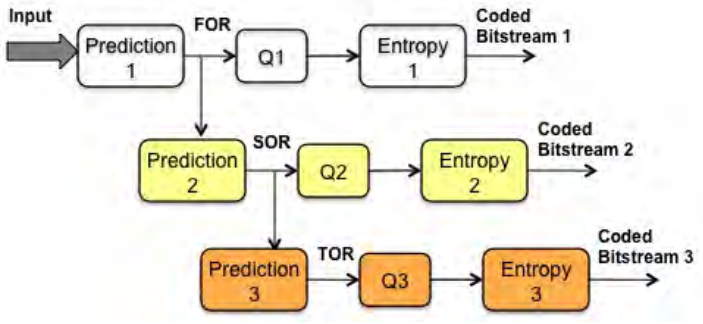


Figure 5.4: Overview of the Multi-Order-Residual (MOR) coding scheme.

5.3.1 Overview of MOR Coding System

The MOR coding system is motivated by the multi-order differencing operation in numerical analysis. In our current context, the long-range correlation of an input image sequence is treated in the first stage and the prediction residuals are called the first-order residuals (FOR). The medium-range correlation and the short-range correlation remain in the FOR image, which can be removed in the second and the third stages using different coding schemes. The prediction residuals in the second stage are considered uncompensated medium-range correlation and are termed as the second-order residuals

(SOR). Similarly, the prediction residuals in the third stage are considered short-range correlations and are called the third-order residuals (TOR). An overview of the MOR coding system is shown in Fig. 5.4.

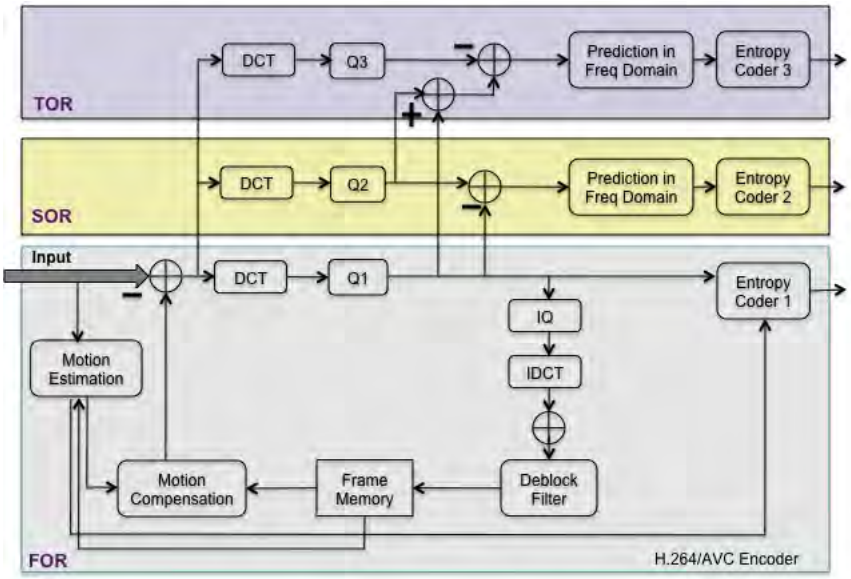


Figure 5.5: The block diagram of the proposed MOR coding scheme.

In the following subsections, we will discuss specific design choices in the three coding stages as shown in Fig. 5.5.

- For the FOR coding, since H.264/AVC is highly effective in removing the long-range correlation, it is employed to encode the FOR image with a coarser quantization value denoted by Q_1.

- Second- and third-order residuals mainly consist of the medium-range correlation, and the traditional H.264/AVC MCP process no longer provides an efficient solution. Thus, the higher residual images are transformed and quantized with a finer quantization value denoted by Q_2 and Q_3 at the second and the third stages,

respectively. In other words, higher-order residuals can be coded using this new framework. A new MCP process in the frequency domain is proposed in Sec. 5.3.3 to remove the medium- and the short-range correlations in the higher-order residuals.

5.3.2 Goals of MOR Prediction

In Chapter 4 we proposed a frequency domain-based prediction and compensation for granular noise data. It is based on the concept that as the GN data have the similar characteristics of film grain noise. Therefore, it will be mainly manifested on the high frequency bands of a transformed DCT block. To directly compensated these high frequency residuals, we propose to perform a prediction and compensation phase in the frequency domain.

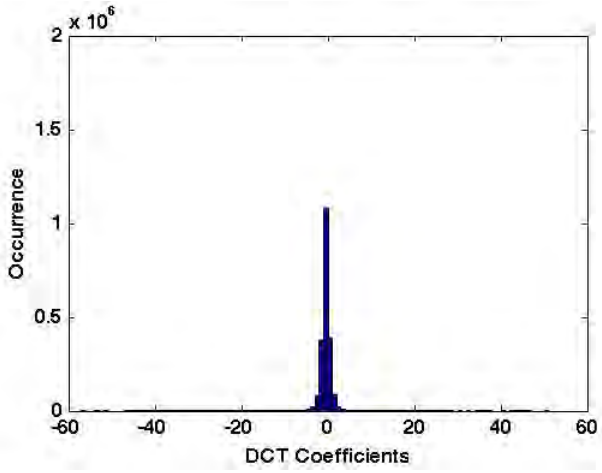


Figure 5.6: A typical histogram of prediction residuals in the DCT domain.

Here, we propose a different frequency domain-based prediction technique to predict signals with MC and SC. The idea is not to replace pixel domain motion compensation,

but to introduce a more effective prediction scheme that can achieve a good prediction results without incurring a high increase in computation complexity. In order to illustrate the MOR predictor design purpose, we first show a histogram for residual signals in the form of DCT coefficients after applying traditional prediction in Fig. 5.6.

After the MCP of H.264/AVC, the dynamic range of the prediction residual in form of DCT coefficients is much smaller (*i.e.* from -60 to 60) compared to the original pixel value range (from 0 to 255). The evaluation process used in H.264/AVC is to minimize the following cost function:

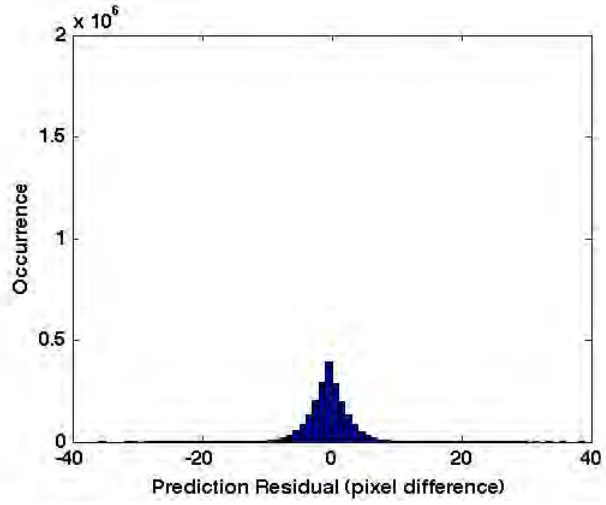
$$J(s,c|\lambda_m) \quad = \quad D(s,c) + \lambda_m \cdot R(s,c), \tag{5.1}$$

where s and c are the original and reconstructed blocks, $D(s,c)$ is the distortion and $R(s,c)$ is the number of bits required to encode the residual and the overhead. The distortion, $D(s,c)$, is obtained by calculating the sum of absolute transformed difference (SATD) in form of

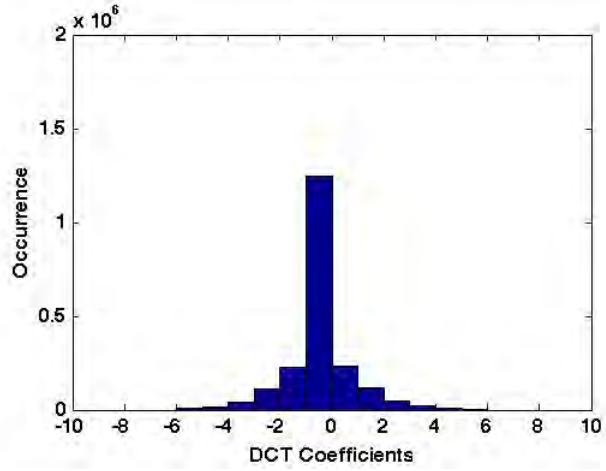
$$SATD(s,c) \quad = \quad \sum \Big| T\Big\{ s[x,y] - c[x,y] \Big\} \Big|, \tag{5.2}$$

where T is a certain orthonormal transform. In H.264/ AVC, T is chosen as the separable Hardmard transform due to its simplicity. This RDO optimization allows the encoder to

select the best prediction without the exact knowledge of prediction residuals in the form of DCT coefficients.



(a)



(b)

Figure 5.7: Histograms of (a) MOR data in form of pixel differences and (b) MOR data in form of DCT coefficeints.

However, if we apply the RDO procedure to the MOR signal, the prediction results will be poor. The reasons will be explained below. Fig. 5.7(a) shows a histogram of the SOR signal in the form of pixel values before the SOR prediction takes place. We can see that the MOR signal in pixel domain already has a much reduced dynamic range. Fig.

5.7(b) show the histogram of the same SOR signal AFTER being DCT transformed. We see that even without further prediction and compensation, the SOR data in the form of transformed DCT coefficients already has an extremely small dynamic range of (-8, 8). Recall in Fig.5.1, the SOR data for fine quantization stepsize (QP=12) also has a much uniformed distribution for all scanning positions.

Therefore, we conclude that the MC and SC signal in the MOR layer, exhibit some unique features such as very limited dynamic range and uniform distribution of the nonzero coefficients. Hence, the best prediction copy obtained through H.264/AVC's RDO optimization process in the motion search in pixel domain might not be able to translate into the true best match after the residual is transformed and compensated. Therefore, we need to develop a different prediction scheme that can target this type of signal characteristics and provide an accurate prediction.

One way to achieve this goal is to introduce additional DCT and quantization process during the motion search phase to evaluate every prediction candidate. This allows the encoder to have an exact knowledge of the prediction results. However, this will require the addition of extremely large amount of computation to be spent on the extra DCT and quantization processes into the already most complex motion search module. Therefore, to maintain an effective motion compensated prediction process for the MOR data while keeping the evaluation process as simple as possible, we introduce the MOR prediction in frequency domain in Sec. 5.3.3.

5.3.3 MOR Prediction in Frequency Domain

The block flow diagram of the system is illustrated in Fig. 5.5. The higher order residual input block will go through the standard DCT module first to achieve frequency separation and quantization. The quantized DCT coefficients will be used in the SOR or TOR prediction in the frequency domain. This predictor design allows the RDO process to be able to evaluate the prediction results directly in the form of DCT coefficients on the fly; thus, making sure the final prediction copy can improve the coding performance. We will further illustrate the RDO design in Sec. 5.3.4.

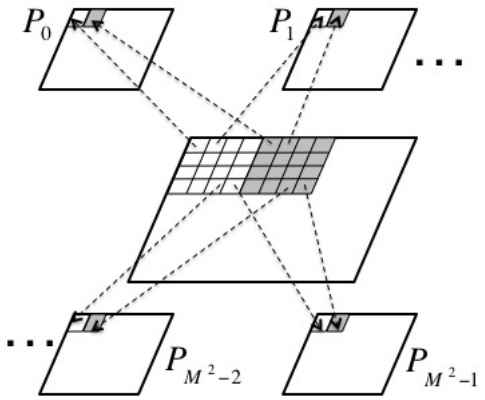


Figure 5.8: Re-grouping of the same frequency coefficients to obtain planes of DCT coefficents, denoted by P_i, where $i = 0, 1, \cdots, M^2 - 1$.

To perform the prediction in frequency domain, the target block first goes through an MxM DCT and quantization outside of the coding loop. Then, each coefficient of

$\{k_0(j), k_1(j), ...k_{M^2-1}(j)\}$ in this $M \times M$ DCT block of block $B(j)$ is extracted into an individual corresponding coefficient plane of $\{P_0, P_1, ...P_{M^2-1}\}$. This coefficient extraction process can be mathematically expressed as

$$P_i = \bigcup_{j=0}^{N} k_i(j), \tag{5.3}$$

This coefficient extraction process as shown in Fig.5.8 is a graphical representation of aggregating the same frequency components from different blocks into a closer plane.

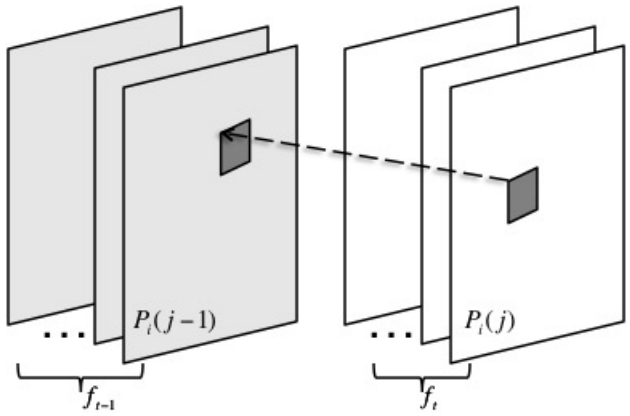


Figure 5.9: MCP in frequency domain for each frequency plane.

The coefficients on each coefficient plane of P_i are further grouped into a s×s partitions of np_l. This re-arranged s×s partition will then go through a compensated prediction phase which is similar to the MCP in spatial domain except that the compensation is done on each individual coefficient planes as follows. See Fig.5.9. Given a SOR/TOR frame of F_t, the frequency extraction process will generate a series of coefficient planes of $\{P_0, P_1, ...P_{M\times M-1}\}$ corresponding to each frequency component. For a DCT transformed and quantized SOR/TOR block $Q[n_x^T]$ of size $M \times M$ at location (i, j) on P_i, the prediction

error of a SOR block in frequency domain is made up of $l = \frac{M \times M}{s \times s}$ number of np_l obtained as

$$\hat{n_x}^T = \sum_l np_l(i + \delta_{i,l}, j + \delta_{j,l}), \tag{5.4}$$

where each translational vectors $(\delta_{i,l}, \delta_{j,l})$ to point to one predicted partitions within the reference planes.

5.3.4 Rate-Distortion Optimization

The RDO process for MOR prediction is based on the same principle of minimizing the Lagrangian function in Eq. (5.1). However, note that in our MOR scheme, MCP with RDO process is performed after the DCT and quantization on the transformed and quantized DCT coefficients. Hence, the distortion is fixed and will not be changed for each search candidate. Therefore, the Lagrangian cost function can be reduced to minimize rate only as

$$J(s, c) = R_{res}(s, c) + R_{mv}, \tag{5.5}$$

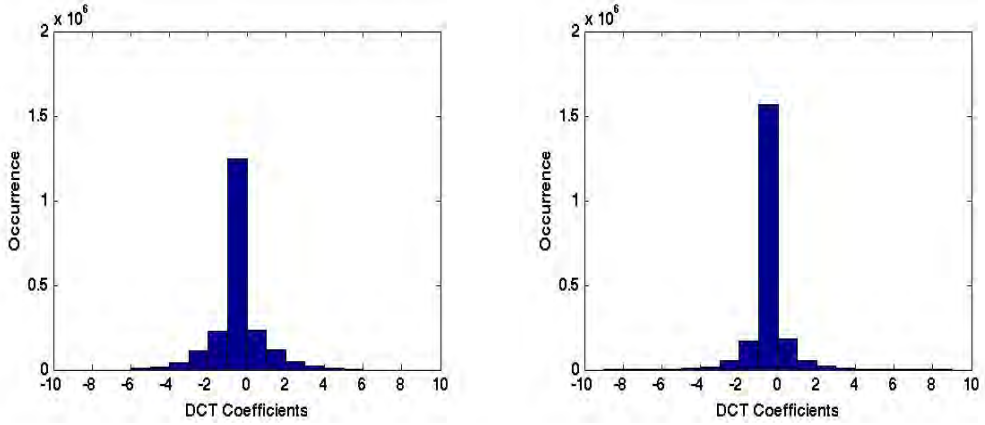
where R_{res} is the bits required to encode prediction residual and R_{mv} is the bits to encode motion vectors. As we are working in the frequency domain already, we can estimate the rate of the prediction residual based on the ρ-domain's approach [20] as

$$R_{res} = \theta \cdot (1 - \rho), \tag{5.6}$$

To estimate the bit used to encode motion vectors, we see that the as the search is operated in a much reduced search range. The MV data has a similar distribution compared to the residual coefficients. Hence, we can consider each MV to use the same bits as a nonzero coefficients. Thus, we can further simplify Eq. (5.5) as

$$J(s,c) \quad = \quad N_{nzTC} + N_{nzMV}, \tag{5.7}$$

where N_{nzTC} is the number of nonzero transformed coefficients and the N_{nzMV} is the number of nonzero motion vectors. The effectiveness of the proposed RDO method can be observed by comparing the histograms of DCT coefficients before prediction and after prediction in Fig. 5.10(a) and (b).



(a) (b)

Figure 5.10: The DCT coefficients histograms of MOR data after MOR prediction in frequency domain.

5.3.5 Pre-search Coefficient Optimization for TOR

To further improve the coding efficiency and to facilitate a fast search, we propose to add a pre-search coefficient optimization phase for third-order residuals. This process is based on the observation that after the coefficients are extracted to individual frequency planes, the last few high frequency planes $P_i, (i = 13, 14, 15)$ are more sparsely populated with nonzero coefficients. This is mainly due to the fact that short-range correlations in the TOR has a very random distribution. This sparse distribution of nonzero coefficients could potentially have a very detrimental effect on the entropy coder, as the CABAC is designed to perform most efficiently when there is a long consecutive run of zeros. Hence, to facilitate the entropy coder, we introduce this pre-search coefficients optimization process for TOR. This optimization happens before the coefficient extraction, and the detailed flow diagram is shown in Fig. 5.11.

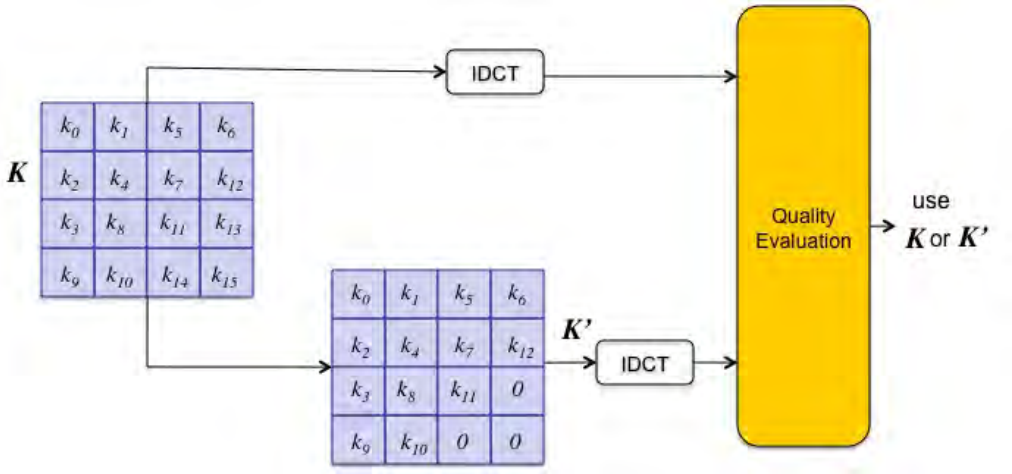


Figure 5.11: The block diagram of the pre-search DCT coefficients optimization process for TOR.

In this pre-search optimization phase, we first examine the DCT coefficient block K. If the DCT block has only less than C_z number of nonzero coefficients in the last three high frequency scanning positions ($SP = 13, 14, 15$). We zero out these positions and perform an IDCT. This partially zero out block is compared to the original incoming block I. Note that this I is the pixel residual from FOR. If the optimization error is lower than the predefined empirical threshold ϕ, we will use this optimized block K' and proceed further to the frequency extraction phase. Otherwise, we will take the original block K instead. This presearch optimization phase helps to boost the consecutive run of zeros in the higher frequency plans and thus allows the later entropy coder to have a better compression results. The complete system diagram is shown in Fig. 5.12.

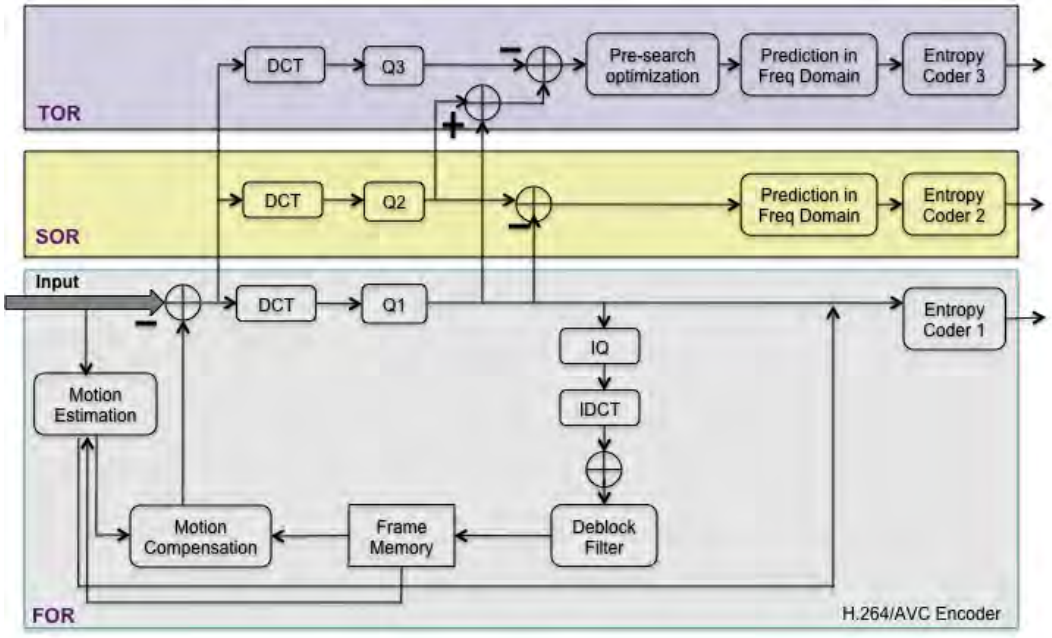


Figure 5.12: The block diagram of the proposed MOR coding scheme with pre-search DCT coefficients optimization for TOR.

Table 5.1: Coding efficiency comparison of the proposed MOR scheme v.s. H.264/AVC for high-bit-rate coding.

Sequence	Resolution	ΔBit Rate (%)
Pedestrian	1920x1080	-18.41
Rush Hour	1920x1080	-18.46
Riverbed	1920x1080	-11.40
Vintage Car	1920x1080	-16.71
Average		**-16.42**

5.4 Experimental Results

In this section, experimental results for the proposed MOR-based coding scheme is presented and compared with H.264/AVC [2] to demonstrate the superior performance of the proposed coding framework. Only the Luminance channel is compared. Four test YUV sequences of High Definition (HD) format at 1920x1080 resolution were used. They were: Pedestrian, Rush Hour, Riverbed and Vintage Car. The results were averaged over 5 P frames for each test sequence. The benchmark H.264/AVC codec used the high profile, with full (high complexity mode) RDO enabled, 1/4pel MCP, and CAVLC as its entropy coder.

For the proposed MOR, $QP_1 = 30; QP_2 = 22$, and $QP_3 = 16$. If MOR's desired quantization step size is larger than 16, for example QP=18, TOR will not be performed and QP_2 will be changed to the target quantization stepsize, *i.e,* QP=18. DCT size is set to 4×4 and the partition sizes for SOR and TOR data with MCP in DCT domain is set to $s_{SOR} = 4 \times 4$ and 2×2 and $s_{TOR} = 2 \times 2$. A zero-order binary arithmetic coding engine is used for entropy coder. The coding efficiency comparison is listed in Table. 5.1 based on [18].

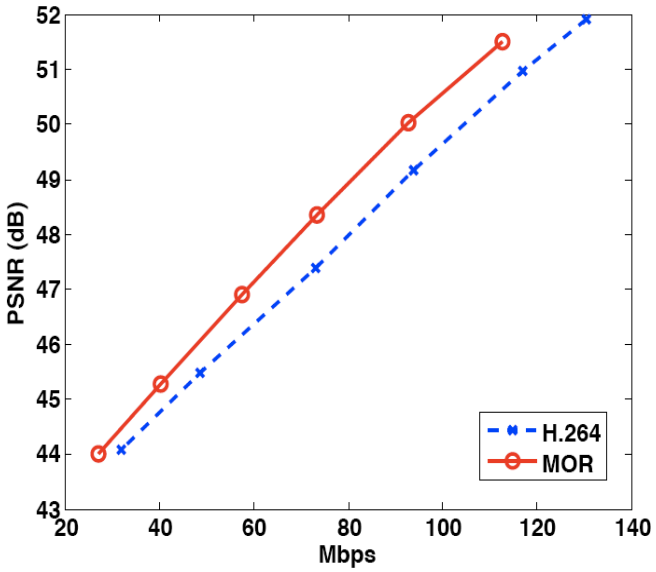


Figure 5.13: Rate-Distortion curves for Pedestrian sequence.

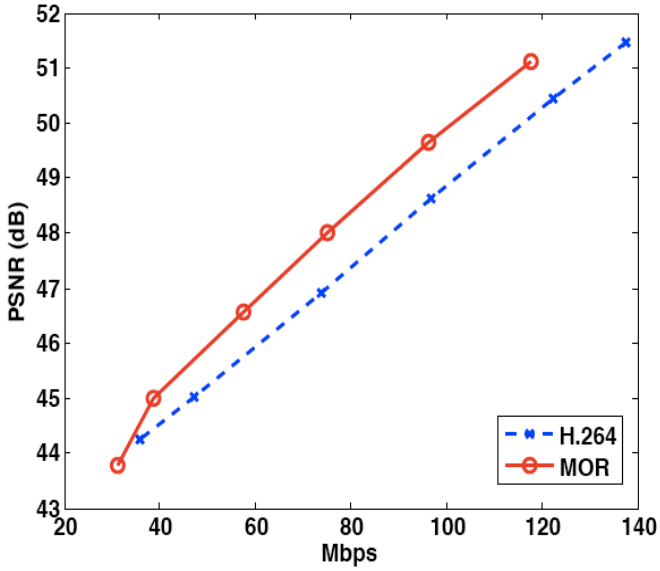


Figure 5.14: Rate-Distortion curves for Rush Hour sequence.

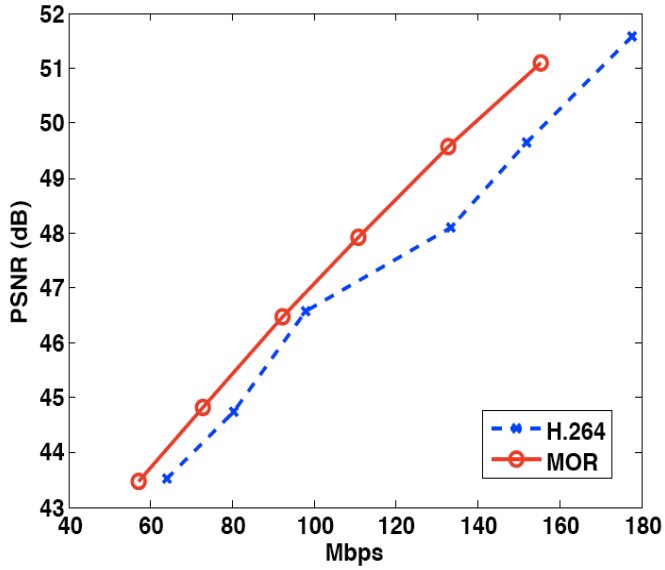


Figure 5.15: Rate-Distortion curves for Riverbed sequence.

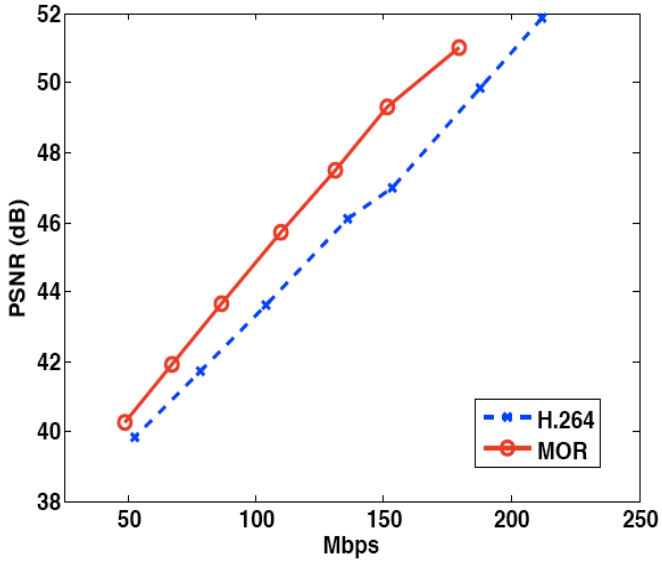
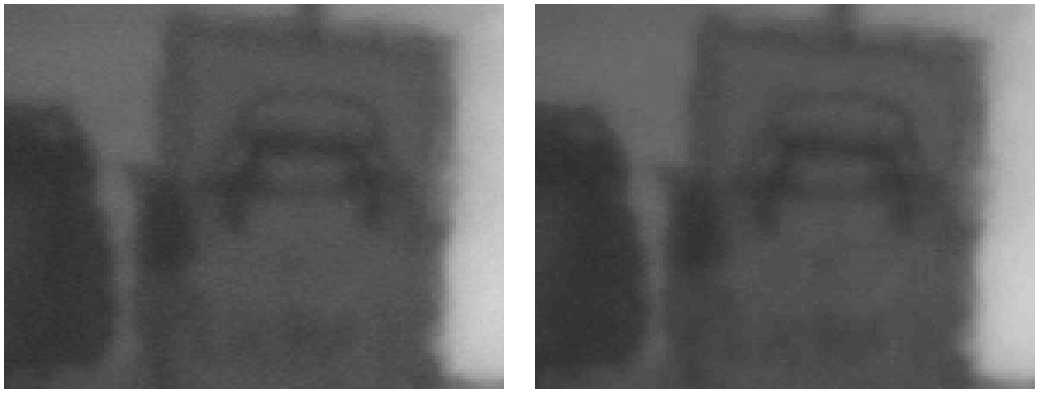


Figure 5.16: Rate-Distortion curves for Vintage Car sequence.

Secondly, we compare the rate distortion performance for the high-bit-rate coding scenario in Fig.5.13 to Fig.5.16. We see that the proposed MOR reduces the coding bit rate of H.264/AVC by up to 23% depending on the quality requirements.



(a) (b)

Figure 5.17: Decoded Rush Hour frames with (a) MOR and (b) H.264 at 60Mbps.

(a) (b)

Figure 5.18: Decoded Vintage Car frames with (a) MOR and (b) H.264 at 80Mbps.

Thirdly, we examine visually of the decoded frames using both MOR and H.264/AVC. In Fig.5.17, we show a side-by-side comparison of Rush Hour frames decoded using the two schemes at the same bit rate. In Fig.5.18, we show a side-by-side comparison of Vintage Car frames decoded using the two schemes at the same bit rate of 80Mbps. We

can observe that the MOR scheme consistently provides a sharper decoded frame with much less coding artifacts.

5.5 Conclusion and Future Work

In this paper, we first examine the impact of high-bit-rate coding to the existing state-of-art codec such as H.264/AVC. We then performed the correlation analysis on the video signals to show that there exist the medium- and short-range correlations in video which result in performance degradation when the coding bit rate becomes higher. Then, we proposed a novel MOR coding scheme that handles the long-, medium- and short-range differently in high-bit-rate coding. The proposed MOR coding scheme employed the H.264/AVC in the coding of the FOR, a frequency domain MCP in the coding of SOR and TOR. It was shown by experimental results that the proposed MOR coding scheme has an average rate reduction of 16% compared to H.264/AVC under different quality requirements. We will continue to improve the coding performance of the MOR scheme in the near future.

Chapter 6

Conclusion and Future Work

6.1 Summary of the Research

In this research, we have studied two major issues in high fidelity video coding: 1) residual processing and 2) subpel motion estimation. The research was motivated by the significant decrease in coding inefficiency of prediction residuals in today's coding schemes under the requirements of high bit rate coding and the high computational complexity associated with the subpel motion estimation.

In Chapter 3, we proposed a direct subpel MV prediction scheme with optimal subpel MV resolution estimation. Existing optimal subpel MV resolution estimation is developed using the texture characteristics of an input block. However, due to motion compensation, quantization and noise, they are not accurate in some cases. The proposed optimal MV prediction scheme can handle blocks of different characteristics by maximizing its rate-distortion (RD) gain through a flexible MV resolution while reducing the computational complexity. Two different optimal MV prediction schemes were developed based on the different shape of the error surface. The rate-distortion performance of the proposed

optimal MV prediction is about the same as that of full search with an average of 90% complexity reduction.

In Chapter 4, we conducted an analysis on the prediction residual and showed that it contains fine structured features when the coding bit rate becomes higher. These fine features were considered as film grain noise in earlier. To treat these granular noise, we introduce an extra granular noise prediction and coding scheme based on the film grain noise extraction process to extract these fine features in the residual.

A frequency-domain based prediction and compensation scheme was further proposed for granular noise data. By correlating the same frequency bands between different blocks, we could maximize the possibility between target GN block and candidate blocks that might contain similar low frequency components but different high frequency components to be considered as candidate reference blocks and vice versa. The prediction between the same frequency bands avoids the complication of sparse matrix multiplication for reconstruction as required in earlier ME in frequency domain. It was shown by experimental results that, as compared to H.264/AVC, the proposed GNPC scheme can achieve an average of more than 10% bit rate reduction in high-bit-rate coding.

In Chapter 5, we further investigate the impact of high bit rate coding from the fundamental signal characteristics. We first study the DCT coefficient distribution and show that, as the video quality requirement increases, the distribution of DCT coefficients is close to an uniform one. This explains the poor performance of traditional image/video codecs in the high bit rate region. We then performed a signal correlation analysis and showed different types of correltions in video frames. Due to the use of a fine quantization step, the quantization process can no longer be used to remove the short and

medium range correaltion effectively. Since the block-based motion-compensated predictive (MCP) coding technique is only effective in removing the long range correlation, the coding performance of the traditional video codecs degrades rapidly as the quality requirement becomes higher. Based on the study, we propose a multi-order residual (MOR) coding scheme. A coefficient optimization technique was proposed to enhance the compression performance furthermore. It was shown by experimental results that the proposed MOR scheme outperforms the state-of-the-art H.264/AVC codec by an average of 16% in bit rate saving.

6.2 Future Research Directions

To make the current research more complete, we would like to extend the current work along several directions as detailed below.

- Advanced sub-pel interpolation scheme

  H.264/AVC employs the quarter-pel ME, and there are two different sub-pel interpolation schemes. A 6-tap filter approach is used for half-pel interpolation, and a bilinear filter is used for quarter-pel interpolation. Although there exist other more sophisticated interpolation schemes that offers better performance than the 6-tap and bilinear filters, the existing codec does not adopt them due to the consideration of computational complexity. Since the proposed optimal subpel MV prediction scheme reduces the complexity of subpel interpolation by an average of 90%, it opens a new opportunity for more advanced interpolation schemes to further improve the overall RD performance.

- Improved QP and layer number selection in the MOR scheme

  In the proposed MOR scheme, QPs for the FOR and the SOR images were chosen empirically at fixed values. As different video streams have different rate-distortion characteristics, an adaptive QP selection scheme should improve the coding performance. Furthermore, the number of layers in the MOR scheme may be adjusted according to video characteristics. For example, a highly complex video stream which contains long-, medium-, and short-range correlations may benefit from more layers while a simple video stream may only demand the FOR and the SOR two layers. For the latter case, the use of fewer layers will reduce the layer overhead. Thus, the ability of dynamically adjusting QP and the layer number should improve the coding performance.

- Advanced prediction techniques and preprocessing of DCT coefficients

  In the proposed MOR scheme, we used a frequency-domain compensation technique. However, we may use more sophosicated prediction techniques in the SOR image. In addition, we employed a simple preprocessing technique for DCT coefficients in the MOR scheme. It is interesting to develop a more sophisticated DCT coefficient preprocessing technique to enhance the coding performance furthermore.

Bibliography

[1] "Digital Cinema System Specification V1.0," in *[Online] www.dcimovies.com/DCI_Digital_Cinema_System_Spec_v1.pdf*, July 2005.

[2] "H.264/AVC Reference software," in *[Online] Available: http://iphome.hhi.de/suehring/tml/*, July 2005.

[3] "VC-1 Technical Overview," in *[Online] Available: www.microsoft.com/windows/windowsmedia/howto/articles/vc1techoverview.aspx*, October 2007.

[4] A.M.Tourapis, O.C.Au, and M.L.Liu, "Fast motion estimation using modified circular zonal search," *IEEE Intl. Symposium on Circuits and Systems*, vol. 4, pp. 231–234, July 1999.

[5] J. Bennett and A. Khotanzad, "Modeling textured images using generalized long-correlation models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 800–809, June 1998.

[6] D. P. Bertsekas, "Nonlinear Programming," *2nd Edition, Athena Scientific, 1999*, 1999.

[7] B.Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. on Communications*, vol. 41, pp. 604–612, Apirl 1993.

[8] B.T.Oh, C.-C.J.Kuo, S.Sun, and S.Lei, "Film grain noise modeling in advanced video technology," *Visual Communications and Image Processing, Proc of SPIE Electronic Imaging*, vol. 6508, May 2005.

[9] C.D.Kuglin and D.C.Hines, "The phase correlation image alignment method," *IEEE Int. Conf. on Image Processing (ICIP2007)*, pp. 163–165, September 1975.

[10] C.Gomila and A.Kobilansky, "Sei message for film grain encoding," in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 SG16 Q.6) JVT-H022.doc*, September 2003.

[11] J. Chalidabhongse and C.-C. J. Kuo, "Fast motion vector estimation using multiresolution-spatio-temporal correlations," *IEEE Trans. on Circuits and Systems Video Technologies*, vol. 7, pp. 477–488, August 1997.

[12] S. F. Chang and D. G. Messerschmitt, "A new approach to decoding and compositing motion-compensated DCT-based images," *IEEE Intl Conf on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 421–424, April 1993.

[13] J. Cho, G. Jeon, J. Suh, , and J.Jeong, "Subpixel motion estimation scheme using selective interpolation," *IEEE Trans. on Communications*, vol. E91-B, December 2008.

[14] C.Zhu, X.Lin, and L.Chau, "Hexagon-base search pattern for fast block motion estimation," *IEEE Trans. on Circuits and Video Technologies*, vol. 12, January 2007.

[15] Y. Dai, Q. Zhang, A. Tourapis, and C.-C.J.Kuo, "Efficient block-based intra prediction for image coding with 2D geometrical manipulations," *IEEE Intl Conf on Image Processing*, October 2008.

[16] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," *Proceedings of the IEEE Intl. Conference on Computer Vision*, vol. 2, pp. 1033–1038, September 1999.

[17] T. Eude, R. Grisel, H. Cherifi, and R. Debrie, "On the distribution of the DCT coefficients," *Proc. 1994 IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, pp. 365–368, April 1994.

[18] G.Bjontegaard, "Calculation of average PSNR difference between RD-curves," in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG-M33.doc*, April 2001.

[19] H. Hang and J. Chen, "Source model for transform video coder and its application part I: fundermental theory," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 287–298, April 1997.

[20] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, December 2001.

[21] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Communication*, vol. COM-29, pp. 1799–1808, December 1981.

[22] J.Cooper, J.Boyce, J.Llach, A.Tourapis, P.Yin, and C.Gomila, "Techiques for film grain simulation using a database of film grain patterns," vol. Pattern EP 1 809 043 A1, July 2007.

[23] J.Lee, I.Choi, W.Choi, and B.Jeon, "Fast mode decision for b slice," in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 SG16 Q.6) JVT-k021.doc*, March 2004.

[24] J.W.Suh, J. Cho, and J.Jeong, "Model-based quarter-pixel motion estimation with low computational complexity," *Electronic Letters*, vol. 45, June 2009.

[25] J.W.Suh and J.Jeong, "Fast sub-pixel motion estimation techniques having lower comtational complexity," *IEEE Trans. on Consumer Electronics*, vol. 50, pp. 968–973, August 2004.

[26] K.H.Lee, J.H.Choi, B.K.Lee, and D.G.Kim, "Fast two-step half-pixel accuracy motion vector prediction," *Electronic Letters*, vol. 36, pp. 625–627, 2000.

[27] R. Kleihorst and F. Cabrera, "Implementation of DCT-domain motion estimation and compensation," *IEEE Workshop on Signal Processing Systems*, pp. 53–62, October 1998.

[28] T. Koga, K. Iinuma, A. Hirano, Y. Lijima, and T. Ishiguro, "Motion compensated inter frame coding for moving picture conferencing," *Proc. NTC 81*, pp. C9.6.1–9.6.5, November 1981.

[29] D. K. Kwon, M. Y. Shen, and C. C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, pp. 517–529, May 2007.

[30] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. on Image Processing*, vol. 9, pp. 1661–1666, October 2000.

[31] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. on Circuits and Systems Video Technologies*, vol. 4, pp. 438–442, August 1994.

[32] M.Schlockermann, "Film grain coding in H.264/AVC," *Joint Video Team(JVT) JVT-1034d2.doc*, September 2003.

[33] Y. Nie and K.-K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Trans. on Image Processing*, vol. 11, pp. 1442–1450, December 2002.

[34] P.Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, pp. 283–310, 1989.

[35] P.R.Hill, T.K.Chiew, D.R.Bull, and C.N.Canagarajah, "Interpolation free subpixel accuracy motion estimation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, pp. 1519–1526, December 2006.

[36] P.Yin, H.Cheong, A.Tourapis, and J.Boyce, "Fast mode decision and motion estimation for JVT/H.264," *IEEE Intl. Conference on Image Processing*, pp. 853–856, September 2003.

[37] L. R. Rabiner and B. Gold, "Theory and application of digital signal processing," *Englewood Cliffs NJ: Prentice Hall*, 1975.

[38] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. on Communication*, vol. COM-31, pp. 835–839, June 1983.

[39] J. Ribas-Corbera and D. L. Neuhoff, "Optimizing Motion-Vector Accuracy in Block-Based Video Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, April 2001.

[40] R.Rao and G.Bjontegaard, "Complexity analysis of multiple block sizes for motion estimation," in *Joint Video Team (JVT) VCEG-m47.doc*, April 2001.

[41] M. Song, A. Cai, and J. Sun, "Motion estimation in DCT domain," *IEEE Int. Conf. on Communication Technology Proceedings*, vol. 2, pp. 670–674, May 1996.

[42] A. D. Stefano, B. Collis, and P. White, "Synthesising and reducing film grain," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 163–182, June 2005.

[43] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: overview and introduction to the fidelity range extensions," *SPIE conf. Applications of Digital Image, Processing XXVII*, vol. 5558, pp. 454–474, August 2004.

[44] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, November 1998.

[45] M. Weinberger, G. Seroussi, and G.Sapiro, "The LOCO-I lossless image compression algorithm: priniciples and standarization into jpeg-ls," *IEEE Trans. on Image Processing*, May 2000.

[46] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC coding standard," *IEEE Trans. on Circuits and System for Video Technology*, vol. 7, pp. 560–576, July 2003.

[47] X.Yi, J.Zhang, N.Ling, and W.Shang, "Improved and simplified fast motion estimation for JM," in *Joint Video Team (JVT) JVT-P021.doc*, July 2005.

[48] Y.Lee, K.Han, and G.J.Sullivan, "Improved lossless intra coding for H.264/MPEG-4 AVC," *IEEE Trans. on Image Processing*, vol. 15, September 2006.

[49] R. W. Young and N. G. Kingsbury, "Frequency-domain motion estimation using a complex lapped transform," *IEEE Trans. on Image Processing*, vol. 2, January 1993.

[50] Z.Chen, P.Zhou, and Y.He, "Fast integer pel and fractional pel motion estimation for JVT," in *Joint Video Team (JVT) JVT-F017r1.doc*, December 2002.

[51] Q. Zhang, Y. Dai, and C.-C. J. Kuo, "Direct techniques for optimal subpel motion resolution estimation and position prediction," *IEEE Trans. on Circuits and Systems for Video Technology*, 2010.

[52] Q. Zhang, Y. Dai, and C.-C. Kuo, "Lossless video compression with residual image prediction and coding (RIPC)," *IEEE International Symposium on Circuits and Systems*, May 2009.

[53] Q. Zhang, S. H. Kim, Y. Dai, and C.-C. J. Kuo, "A second-order-residual (SOR) coding approach to high-bit-rate video compression," *SPIE Electronic Imaging*, January 2010.

[54] Q. Zhang, S.H.Kim, Y. Dai, and C.-C. J. Kuo, "Multi-order-residual (MOR) video coding: framework, analysis and performance," *SPIE Visual Communications and Image Processing*, July 2010.