國立交通大學

資訊工程學系

碩士論文

利用方向小波轉換分析進行視訊編碼之位元分配

Curvelet Domain Analysis for Video Coding Bit Allocation

研究生：蔡雅婷

指導教授：蔡淳仁　博士

中華民國九十六年六月

利用方向小波轉換分析進行視訊編碼之位元分配

Curvelet Domain Analysis for Video Coding Bit Allocation

研究生：蔡雅婷　　Student：Ya-Ting Tsai

指導教授：蔡淳仁　　Advisor：Chun-Jen Tsai

國立交通大學

資訊科學與工程研究所

碩士論文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master

in

Computer Science
June 2007
Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

利用方向小波轉換分析進行視訊編碼之位元分配

學生：蔡雅婷　　　　　　　　　　　　　指導教授：蔡淳仁博士

國立交通大學資訊科學與工程研究所

摘要

本論文主旨在於探討人類視覺特性於視訊編碼中之位元配置方式，並根據論文中提出的方法利用方向小波轉換分析視訊影片中的原始圖像及殘值圖像的結構性特徵，且依此分析結果設計一個位元分配的策略，使得在視覺上較為重要的區域可得到較多的位元，以期達到較佳的視覺品質。本論文中使用方向小波轉換分析的主因為它可在不同方向上做子頻帶分解，所以相較於離散餘弦轉換或其他基於可分離轉換的小波轉換，方向小波轉換能顯現出更多的結構性資訊。論文中提出的位元分配的策略嘗試在非結構性區域節省位元及在結構性區域增加視覺品質。在 MPEG-4 簡易版類別編碼器上的實驗結果顯示，提案方法在所有測試案例中皆有較好的表現，能在人眼視覺較為重視的區域得到較佳的影像品質。

Curvelet Domain Analysis for Video Coding Bit Allocation

Student：Ya-Ting Tsai　　　　　　　　　　Advisor：Dr. Chun-Jen Tsai

Computer Science and Engineering College of Computer Science
National Chiao Tung University

Abstract

This paper proposes a video bit allocation scheme based on Curvelet domain analysis. The proposed algorithm analyzes the structural characteristics of the intensity and motion-compensated residual images of a video sequence in curvelet domain to determine a bit-allocation policy so that visually important regions will be allocated with more bits. Curvelet transform is adopted in this thesis for such analysis because it performs sub-band decomposition in various directions so that more structure information is revealed in curvelet domain than in DCT or other wavelet domains based on separable transforms. The proposed bit-allocation policy tries to save bits in unstructured regions and increase quality in structured regions. Experiments using standard test sequences coded with an MPEG-4 simple profile video encoder show that the proposed bit allocation method has better performance (achieves higher PSNR's) in the regions most human observers care about in all test cases.

Acknowledgement

I would like to express my gratitude to all those who supported and encouraged me to complete this thesis. First of all, this paper owes much to the thoughtful and helpful suggestions and comments of my advisor, Professor Chun-Jen Tsai. I gratefully acknowledge helpful discussions with him on several points in the paper. Then, my special thanks are due to my seniors, juniors, and classmates, especially my senior, Chien-Peng Ho for consistent encouragement and valuable advice. During these days at National Chiao-Tung University, I enjoy the moments I have with all MMES Lab members. Finally, I would like to thank my dear parents and brother for their support and encouragement.

Content

1. Introduction 1

2. Previous Work 4

2.1. Transform Analysis 4

2.2. Properties of Human Visual System 7

3. Study and Analysis of Curvelet Transform 10

3.1. Why Curvelet Transform 10

3.2. Fundamentals of Curvelet Transform 11

3.3. Mathematical Formulation of Curvelet Transform 13

3.4. Implementation of Digital Curvelet Transform 17

3.4.1. Take Fourier transform into frequency domain 18

3.4.2. Band-pass filtering 19

3.4.3. Polar interpolation 24

3.4.4. Inverse Fourier transform 26

3.5. Interpretation of the Curvelet Transform Coefficients 26

3.5.1. Curvelet Coefficients in the Coarsest and the Finest level 27

3.5.2. Curvelet Coefficients in the Middle Levels of Resolutions 27

4. Proposed Bit Allocation Framework 30

4.1. Analysis on curvelet transform coefficients 30

4.1.1. The composition of the curvelet transform 31

4.1.2. Image type and the presentation of the related coefficient 33

4.2. The Otsu Algorithm 35

4.3. Statistical method to analyze the coefficients 38

4.4. Proposed Bit Allocation Scheme 41

4.5. Determination of the Weighting Threshold 43

5. Experiment and Analysis 46

5.1. Result of the proposed bit allocation scheme 46

5.2. Result of Proposed Bit Allocation Scheme with Weighting Threshold 52

5.2.1. Number of structured and unstructured blocks of the two method 53

5.2.2. The result of proposed scheme with linear formula 54

6. Conclusion and Future Work 62

7. Reference 65

List of Figures

Figure 2.1. Reduction of blocking artifacts using DCT 5

Figure 3.1 The anisotropy scaling relation for curves. 12

Figure 3.2 The curvelet tiling in the frequency domain 17

Figure 3.3 Decomposition of image into frequency domain 18

Figure 3.4 Decomposition of Meyer wavelet in resolution of scale 0 to scale 4 20

Figure 3.5 Scaling function and wavelet function of Meyer window function 21

Figure 3.6 Angular scaling in the East quadrant region 24

Figure 3.7 The directional decomposition in the 4th level coefficients. 26

Figure 3.8 Perform 2D IFFT on lowest level 26

Figure 3.9 Result of coarest level of curvelet coefficients 27

Figure 3.10 Coefficients of each resolution level 28

Figure 3.11 Position mapping of curvelet coefficients 29

Figure 4.1 Position mapping of curvelet coefficients in the coarsest level resolution 31

Figure 4.2 Curvelet coefficients in the finest level resolution 32

Figure 4.3 Curvelet coefficients in the 2nd, 3rd and 4th level resolution 32

Figure 4.4 Curvelet coefficients in the 2nd, 3rd and 4th level resolution 34

Figure 4.5 The coordinate formed by curvelet coefficients 35

Figure 4.6 The histogram separated by Otsu threshold selection method 38

Figure 4.7 The histogram and coefficient of variation of six blocks in Stefan image 40

Figure 4.8 Block diagram of the proposed bit allocation model 42

Figure 4.9 GOP structure of a video sequence. 45

Figure 4.10. Threshold selection of CV. 45

Figure 5.1 Comparison of visual quality in Stefan Sequence 49

Figure 5.2 Comparison of visual quality in Football Sequence 50

Figure 5.3 Comparison of visual quality in Bus Sequence 51

Figure 5.4 Comparison of visual quality in Stefan Sequence 56

Figure 5.5 Comparison of visual quality in Football Sequence 58

Figure 5.6 Comparison of visual quality in Bus Sequence 60

Figure 6.1 The directional decomposition in the 4th level coefficients. 63

List of Tables

Table 3.1 Input Parameter to curvelet codec 17
Table 3.2 Four basic parts of 1-D Meyer window 22
Table 3.3 Parameters of each resolution level 22
Table 3.4 The band-pass filtered coefficients of image Lena 23
Table 3.5 Formula of Meyer Window 25
Table 4.1 Analysis of frequency components. 44
Table 5.1 Parameter settings of the three experiments 47
Table 5.2 Resulted data of the three sequences 47
Table 5.3 Ratio of improvement in Stefan sequence 49
Table 5.4 Ratio of improvement in in Football sequence 51
Table 5.5 Ratio of improvement in in Bus sequence 52
Table 5.6 Automatic selection of threshold using method I. 53
Table 5.7 Automatic selection of threshold using method II. 54
Table 5.8 Average result of Method I 55
Table 5.9 Average result of Method II 55
Table 5.10 Ratio of number of regions with better PSNR and worse PSNR in Stefan sequence 56
Table 5.11 Saved bits in unstructured regions in Stefan sequence 57
Table 5.12 Ratio of number of regions with better PSNR and worse PSNR in Football sequence 59
Table 5.13 Saved bits in unstructured regions in Football sequence 59
Table 5.14 Ratio of number of regions with better PSNR and worse PSNR in Stefan sequence 60
Table 5.15 Saved bits in unstructured regions in Bus sequence 60

1. Introduction

Digital distribution of video is becoming popular today due to the applications such as digital television, digital camcorder, DVD player, etc. For the decades, video compression is an important technology in multimedia applications. A great number of digital video codec standards have been published; including MPEG-2, MPEG-4, and H.264, etc. Many research efforts have been put into the design of encoders that can achieves the best overall quality based on these standard codecs.

In video codec, the coding scheme can be divided into several processes: predictive coding, transform coding, quantization (e.g. rate-distortion coding), and entropy coding. Transform coding tries to decorrelate the components of video data and to centralize the energy of video data in order to facilitate rate-distortion coding and entropy coding. In recent years, many different transforms have been published to improve compression efficiency. Among these transforms, the most popular transforms are Fourier Transform, Discrete Cosine Transform (DCT) and Wavelet Transform.

Discrete Cosine Transform is more widely used for image and video coding than Fourier Transform since the performances of these two transforms are similar but the operation of DCT is simpler: DCT involves only real number operations instead of complex number operations. Wavelet Transform becomes more popular in recent years since it captures both frequency domain and spatial domain information in one compact representation. The wavelet transform performs very well on one dimensional signal since it can represent signal discontinuity in a more compact form than DCT does, but not as good as expected on two-dimensional data.. The main reason is that to reduce computational complexity, most practical wavelet transform implementation uses separable 1-D transform. The contours in a two-dimensional image can be oriented in any directions. However, the separable wavelet transform only captures signal discontinuities in horizontal and vertical directions.

Recently, a new sub-band decomposition method called curvelet transform has been proposed. Unlike the wavelet transform, curvelet transform decomposes data components into multi-directional data sets and it also maintains the multi-scale spatial information similar to that of a wavelet transform. However, it is not easy to find a critically-sampled curvelet domain representation of an image, therefore, curvelet transform are not used for general image or video compression applications. It is more often used to separate high frequency components due to noises and high frequency components due to signal discontinuities of the image data.

In a video codec, the module that controls the size of the bitstreams for different coding units is called the rate control module. Finding a good trade off between video data rate and visual quality (distortion) is one of the key issues of a high performance rate control scheme. Most encoder tries to estimate the rate-distortion function of a video sequence during encoding. However, the distortion measures are usually MSE or MAD-based so that it does not precisely reflect the visual importance of the video data. In general, the importance of a coding unit is related to the sub-band data in the frequency domain. For example, the components of high-band frequency data are less visually important than the low-band frequency data at the same spatial resolution scale.

It has been shown [19][35] that a structured-region in image has more visual importance than an unstructured-region in image. A region full of random textures (or motion-compensated residuals in residual images) is usually hard to encode and not easy for human eyes to discern the degree of distortions. This kind of image component is referred to as unstructured regions. On the other hand, a region whose textures are simple, with discontinuities in only few directions is referred to as structured region, and any distortion in such regions can be picked up easily by human eyes.

In general, it is not easy to classify between structured and unstructured regions in an image, especially when the definition of "structureness" is dependent on human observers. In this thesis, the curvelet domain analysis is proposed to achieve this goal.

The major advantage of curvelet transform is to decompose input data into frequency coefficients of several directions at each spatial resolution scale. Therefore, in this thesis, we try to analyze video data in curvelet transform domain by its directional presentation in order to classify image regions into structured and unstructured regions and to achieve better bit allocation for video compression. The goal of the proposed technique is to save bits in unstructured regions since human eyes can not discern the distortion. The saved bit budget will be allocated to structured regions to improve its visual quality.

The organization of the thesis is as follows. Chapter 2 introduces some related work of rate control schemes and the perceptual models of human visual systems. Chapter 3 introduces the curvelet transform, including the mathematical definition and its architecture. In chapter 4, the proposed method will be described in detail. The experimental results will be shown in chapter 5. Finally, the conclusions and discussions will be given in chapter 6.

2. Previous Work

As mentioned in chapter one, the main purposes of the transform coding process are to find a more compact data representation and to facilitate video data analysis for rate-distortion coding. For several decades, a great number of different transforms has been studied. Among these transforms, the most popular transforms used for transform coding are the Discrete Cosine Transform (DCT) and the Wavelet Transform. However, since these transforms do not decompose the frequency sub-bands alone image edges, existing transform domain representations are not compact at signal discontinuities. A new transform, curvelet transform, tries to provide multi-resolution and multi-directional signal decomposition, is introduced by Cand`es et al.[22] [23][24]. What is more, Human Visual System (HVS) is researched to classify the regions in image whether the distortion of the region human eyes are sensitive to or not. These studies are of use in compression technique during the bit-allocation selection scheme.

The organization of this chapter is as follows. In section 2.1, we will first introduce popular transforms used in transform coding systems. And then we will analyze the pros and cons of existing transforms and briefly describe the reasons why a new transform, curvelet transform, is adopted in this thesis. In section 2.2, existing work on modeling the relation between human vision systems and the characteristics of images are discussed. Furthermore, the reason why we use curvelet analysis to distinguish between unstructured-texture regions and structured-texture regions are discussed.

2.1. Transform Analysis

For a long time, Discrete Cosine Transform (DCT) [1][2] is one of the most popular transforms which is used in transform coding systems. It is because DCT keeps a good balance between compactness of data representation and computational complexity [3][6].

The DCT is defined as follows [3][4]:

$$F(u,v)=\frac{C(u)C(v)}{\sqrt{MN}}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\cos\frac{(2i+1)u\pi}{2M}\cos\frac{(2j+1)v\pi}{2N}f(i,j) \tag{2.1}$$

$$C(\xi)=\begin{cases}\frac{\sqrt{2}}{2} & \text{if } \xi=0,\\ 1 & \text{otherwise.}\end{cases} \tag{2.2}$$

In equation 2.1, f(i, j) represents an entry of coefficients on the location of (i, j) and M and N mean the size in the horizontal and vertical direction separately. The two-dimensional DCT transforms it into a new function F(u,v), with integer u and v running over the same range as i and j. Equation 2.2 defines the value of multiplicator C(ξ) according to whether the DCT coefficient is the direct current (DC) component of the signal or not.

Compared to DFT, DCT can minimize the blocking artifact when coefficients are truncated or quantized. For example, as shown in Figure 2.1, the implicit n-point periodicity of the DFT can be replaced by the implicit 2n-point periodicity of the DCT. Therefore, the boundaries between adjacent sub-images become invisible because implicit 2n-point periodicity of DCT does not inherently produce boundary discontinuities.

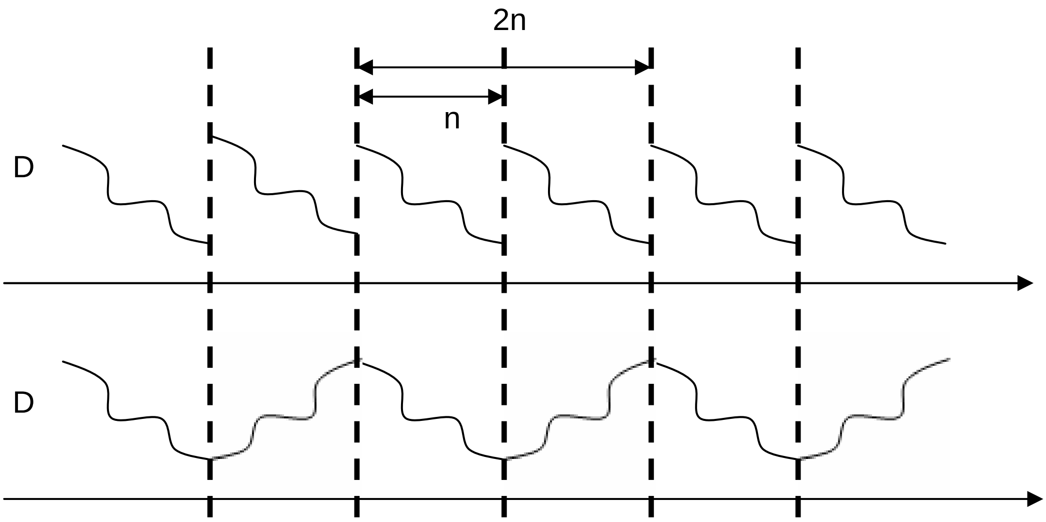


Figure 2.1. Reduction of blocking artifacts using DCT

In 1987, Mallat constructed a structure of wavelet function and the analysis and synthesis progress of signal decomposition. More importantly, Mallat first showed that wavelets are the foundation of a powerful approach to signal processing and analysis called multi-resolution theory [7]. In the next year, Daubechies proposed orthonormal and compactly supported wavelet and the theory of wavelet analysis was constructed [8].

The main advantages of wavelet transform are listed as follows. First of all, wavelet transform is computationally efficient and inherently local; therefore, it is not necessary to subdivide the original image into sub-images before applying the transform. As a result, the method eliminates the side effect of blocking artifact which is usually produced by the DCT-based compression scheme. Even more importantly, signals decomposed into wavelet domain have good resolution in both time and frequency domain. These characteristics work nicely for one-dimensional signals. Signals such as audio data using wavelet-based compression scheme perform better than those using the traditional DCT-based compression scheme. Finally, wavelet transform can decompose signals into coefficients with different levels of resolution, and the characteristic is called multi-resolution [9]. For analyzing the signal data, multi-resolution representation is very effective since the decomposed coefficients are scale-invariant interpretations. These kinds of decomposition provide a hierarchical framework to interpret the signal data.

However, for two-dimensional image or video compression, traditional wavelet transforms which use 1-D separable transforms to decompose the sub-bands has significant drawbacks. Such transforms only capture signal discontinuity efficiently in either horizontal or vertical directions. Nevertheless, the direction of signal discontinuities in two dimensional signals can vary along 360 degree of angles. Traditional separable wavelet transform fails to capture the geometry of image and edges due to the fact that the directions of contours in two-dimensional images can take arbitrary angles [10][11].

To remedy this shortcoming, another method of sub-band decomposition called curvelet transform has been published. Unlike the wavelet transform, curvelet transform not only decomposes data components into multi-direction data sets but also maintains the characteristic of multi-scale spatial resolution of wavelet transforms [12][13]. In this thesis, curvelet transform is used as a tool to help analyzing the contours of the two-dimensional images. The detail description of curvelet transform is presented in next chapter.

2.2. Properties of Human Visual System

In resent years, many researches on Human Visual System (HVS) are published, hoping to find a computational model for the behavior of human eyes [14][15]. These researches are important to image processing and coding [16][36].

When human observers look at still images, the perceptual importance of each region in images is not the same. Many characteristics of the image regions such as the shape of objects, the contrast of luminance, the location of objects, the size of the full objects, and the articulation of the objects will affect the perceptual importance significantly. Furthermore, whether an object is in the foreground or the background may also affect its perceptual importance [17].

Many studies of the relation of human eye movements and the features have been published [15]-[19]. When humans look at a still image, they move their eyes several times a second. Therefore, the features of the region that human eyes stop to gaze every time can be considered as the features that can attract human eyes. Enhancement of the visual quality of these regions is a higher priority task than improving other regions'. Some researchers classify the features of the regions that attract human eyes' attention into three main groups. First, human observers always take priori notice of the regions that contain the faces [18]. Secondly, the regions that have higher spatial contrast intensity would attract human observers than other regions. Third, human eyes also tend to look at the regions that the correlations of the intensities of the nearby image pixels are weaker [19].

The regions which have large contrast intensities include two different types of textures: the structured-textured regions and the unstructured-textured regions. More precisely speaking, these two kinds of regions can be discriminated by the representation of the edges (signal discontinuities) in the region. First of all, the structured region means that the number of the edges of objects inside the specific region is relatively little and the lengths of the edges are long (structured stimuli). On the other hand, the unstructured region means that the number of the edges of objects inside the specific regions is quite large, the position and the direction of the edges are quite random, and the lengths of the edges are small (random stimuli).

As a result, the distribution of edge pixels in unstructured region is scrambled and entropy of this kind of region is usually higher than that of the structured regions. However, human observers usually have trouble discerning the distortion in the unstructured regions. In other words, we can dispatch fewer bits to the unstructured regions because human eyes are less sensitive to its distortions. On the other hand, since human eyes are more sensitive to the distortions in the structured regions, allocating more bits to structured regions can enhance the visual quality more significantly than allocating more bits to unstructured regions.

For a video sequence, previous discussion on the structure of textures can be extended to the temporal domain as well. If the motion of an object from one frame to the next is smooth and can be tracked easily by eye movements, the texture of the object will have a stable projection on the retinas. Therefore, it would be easy for human to discern coding distortion of the sequence on this particular object. On the other hand, if the motion is random, it would not be easy for human eyes to get a stable image on the retina and the coding noises would not be apparent to a human observer. The type of motion can be analyzed from the motion-compensated error residual images. Again, random, small edges in the error residual images around the area of the object means that the object is moving randomly, while an area with structured residual images means that the object is moving smoothly.

In this thesis, we propose a new model that can distinguish the unstructured regions from the structured ones. Therefore, the new model can be used in the bit-allocation process to enhance the visual quality of coded bitstreams.

3. Study and Analysis of Curvelet Transform

The goal of this thesis is to design a video bit allocation model based on visual behavior. As the previous chapter describes, we perform video data analysis in curvelet domain. Therefore, before we present the perceptual model-based bit allocation algorithm, we first introduce the 2-D curvelet transform in this chapter. First of all, we must study the theory and characteristics of curvelet transform. Secondly, it is important to understand the digital implementation of curvelet transform and the meaning of transformed coefficients in order to design the bit-allocation algorithm for perceptual-based video coding.

This chapter is organized as follows: We begin in section 3.1 by showing the reason why we must use curvelet transform to analyze the video data. Section 3.2 describes the main features of curvelet transform. In section 3.3, the mathematical formulation of curvelet transform is presented. Furthermore, section 3.4 presents the implementation of digital curvelet transform. Finally, the representation of transformed coefficients is introduced in section 3.5.

3.1. Why Curvelet Transform

For the last two decades, many transformations based on multi-scale decomposition have been published [7]-[14]. Today, especially in the field of signal processing, multi-scale and multi-resolution based transformations such as wavelets are becoming the popular decomposition methods. Multi-scale transforms have many advantages [21]. First of all, with multi-resolution transform, compressed data can be transmitted in scalable fashion. That is, low resolution data can be transmitted before high resolution data. Secondly, using multi-scale transform is convenient for data mining in large datasets. Thirdly, signal noise removal, for example, in image restoration is more effective in the multi-resolution transform domain. As a result, there are an increasing number of studies of multi-scale and multi-resolution transformations recently.

In last few years, a multi-scale based transform, curvelet transform, was developed to improve the limitations of traditional multi-scale transforms.[22] [23][24] Generally speaking, the curvelet transform is applied using a pyramid structure with multi-resolution. In each scale of the pyramid, the curvelet coefficients records frequency components along different directions [25] [26].

Comparing to traditional 2-D wavelet transform for images, curvelet transform is an over-complete system that contains more sub-band information and therefore it handles some problems better than traditional wavelets [27]. In curvelet transform domain, the representation of edges in an image region (at a particular scale) can be analyzed from multi-directional decomposition of the spatial edges into frequency components. With multi-directional frequency decomposition and multi-resolution characteristics of the curvelet transform, we can obtain more information regarding the structure of the image textures (or the motion-compensated residuals). To be more specific, in the proposed bit allocation scheme, we first analyze the image data according to their frequency components along each edge direction. This analysis discriminates a region with structured texture (or motion-compensated residuals) from a region with unstructured texture (or motion-compensated residuals). Finally, the result of the analysis is used in the bit allocation decision in the video rate control mechanism. All the processes will be described in detail in next chapter.

3.2. Fundamentals of Curvelet Transform

The basic idea of curvelet transform arises from anisotropy scaling relation for curves which is also called the curve scaling law [23].

$$\text{width} \propto \text{length}^2 \qquad (3.1)$$

Figure 3.1 illustrates the basic idea of curvelet transform [23][24]. First of all, suppose there exists a curve u = u(v) in the (u, v) orthogonal coordinate system. In general, we can use the Taylor series expansion to expand the equation u=u(v) as in Equation 3.2.

$$u(v) \approx u(0) + u'(0)v + \frac{u''(0)}{2}v^2, \qquad \text{when } v \approx 0 \qquad (3.2)$$

Figure 3.1 shows that the curve of u=u(v) can be locally approximated by a basis function with rectangle with width w and length ℓ. The relation of the width and length is w=u(ℓ/2).

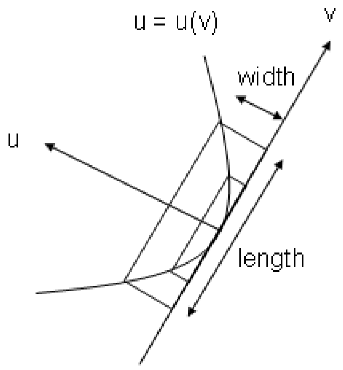


Figure 3.1 The anisotropy scaling relation for curves.

Moreover, since the v-axis is tangent to the curve at the origin (0, 0), the value of the u(0) and u’(0) is zero. As a result, we can obtain Equations 3.3 and 3.4.

$$u(v) \propto \frac{v^2}{2} \tag{3.3}$$

$$w \propto \frac{\ell^2}{8} \tag{3.4}$$

In conclusion, if we construct a correct multi-resolution scale for two-dimensional curves, we will get better approximations when the scale becomes finer.

The advantage of curvelet transform comes from a flexible multi-resolution and directional image expansion using curve segments. To be more specific, curvelet transform is a multi-resolution decomposition method. If the total number of resolution is N, the 1st level is the coarsest level and level N being the finest level N. Coefficients in the coarsest level and the finest level are not decomposed by directional filters, so they do not contain the information of directional frequency component analysis. On the other hand, coefficients from the 2nd level to the N-1th level are decomposed by two-dimensional band pass filter first and then by directional filter latter. Therefore, for these levels, the coefficients will be separated by many angular wedges, and each wedge contains frequency components of the image signals decomposed along a specific orientation. The procedure of coefficients separation by angular wedges is called parabolic scaling.

3.3. Mathematical Formulation of Curvelet Transform

First of all, we define a two dimensional space, R^2, with four variables which are a spatial-domain variable x, a frequency-domain variable ω, and variables r and θ in polar coordinates in the frequency-domain [21]. Conceptually, the principal filters are based on two basic windows, which are radial window, W(r), and angular window, V(t), respectively.

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \quad r \in \left(\frac{3}{4}, \frac{3}{2}\right); \tag{3.5}$$

$$\sum_{\ell=-\infty}^{\infty} V^2\left(t - \ell\right) = 1, \quad t \in \left(-\frac{1}{2}, \frac{1}{2}\right) \tag{3.6}$$

Equation 3.5 is the radial window W(r), which decomposes the image data in Fourier domain as the band pass filter. The window is smooth, nonnegative and real-valued, and its argument r is positive and real valued. Equation 3.6 is the angular window V(t), which decomposes the image data in Fourier domain into several wedges that contain different directional coefficients. The filter is also smooth, nonnegative and real-valued, and its argument t is real valued. The argument j represents the scale of the coefficients, and ℓ means the direction of the coefficients.

$$U_j\left(r, \theta\right) = 2^{-\frac{3j}{4}} W\left(2^{-j} r\right) V\left(\frac{2^{\left\lfloor \frac{j}{2} \right\rfloor} \theta}{2\pi}\right), \quad \text{for } j \geq j_0 \tag{3.7}$$

The radial and angular window can form a frequency window U_j as shown in Equation 3.7. Like Equation 3.5, the argument j represents the scale of the coefficient. To be more specific, j_0 is just the 1st level of the decomposition scale and $\lfloor j/2 \rfloor$ is the integer part of j/2. In curvelet transform, the directional decomposition starts from the 2nd level of resolution and ends at the last level of resolution. In other words, the 1st level (coarse scale) decomposition can only produce the roughly low pass filtered coefficients. For other scales, U_j can decompose image data into polar wedges that contain different directions.

More precisely, we can use the waveform to represent the frequency window.

$$\mathrm{U}_{\mathrm{j}}(\mathrm{r},\theta)=\mathrm{U}_{\mathrm{j}}(\omega_1,\omega_2)=\mathrm{U}_{\mathrm{j}}(\omega) \tag{3.8}$$

$$\hat{\varphi}_{\mathrm{j}}(\omega)=\mathrm{U}_{\mathrm{j}}(\omega) \tag{3.9}$$

In Equation 3.8, we change the form of frequency window form polar coordinate to orthogonal coordinate. And then we can use the waveform in Equation 3.9 to define the frequency window.

Let's introduce two parameters that indicate the position of coefficient in the polar wedge.

$$\theta_\ell = 2\pi \cdot 2^{-\lfloor \mathrm{j}/2 \rfloor} \cdot \ell, \qquad \begin{aligned}&\text{where} \quad \ell = 0,1,\ldots \\ &\text{and} \quad 0 \le \theta_\ell \le 2\pi\end{aligned} \tag{3.10}$$

$$\mathrm{x}_{\mathrm{k}}^{(\mathrm{j},\mathrm{k})} = \mathrm{R}_{\theta_\ell}^{-1}\left(\mathrm{k}_1 \cdot 2^{-\mathrm{j}}, \mathrm{k}_2 \cdot 2^{-\mathrm{j}/2}\right) \qquad \text{where } \mathrm{k} = (\mathrm{k}_1,\mathrm{k}_2) \in \mathrm{Z}^2 \tag{3.11}$$

$$\mathrm{R}_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \mathrm{R}_\theta^{-1} = \mathrm{R}_\theta^{\mathrm{T}} = \mathrm{R}_{-\theta} \tag{3.12}$$

The Equation 3.10 expresses the rotation angles θ_ℓ that we use to indicate the direction of coefficients. The next Equation 3.11 shows the position of coefficient $\mathrm{x}_{\mathrm{k}}^{(\mathrm{j},\mathrm{k})}$ in the spatial domain that is controlled by the translation parameter k. And the notation R_θ^{-1} is the inverse (and transpose) of rotation by θ radians as what Equation 3.12 shows.

Therefore, at decomposition scale $2^{-\mathrm{j}}$, at the orientation of rotation angles θ_ℓ, and at the position $\mathrm{x}_{\mathrm{k}}^{(\mathrm{j},\mathrm{k})}$, basic curvelets can be defined as follows:

$$\varphi_{j,\ell,k}(x) = \varphi_j\left(R_{\theta_\ell}\left(x - x_k^{(j,\ell)}\right)\right) \tag{3.13}$$

As a result, we can use the inner product operation on an element f and a curvelet to produce a curvelet coefficient, and the formulation is represented as Equation 3.14.

$$\begin{aligned} c(j,\ell,k) &= \left\langle f, \varphi_{j,\ell,k} \right\rangle \qquad \text{where } f \in L^2\left(R^2\right) \\ &= \int_{R^2} f(x)\overline{\varphi_{j,\ell,k}(x)}\,dx \end{aligned} \tag{3.14}$$

We can also translate the curvelet coefficient of Equation 3.10 into the frequency domain operation as Equation 3.15 shows:

$$\begin{aligned} c(j,\ell,k) &= \frac{1}{(2\pi)^2}\int \hat{f}(\omega)\overline{\hat{\varphi}_{j,\ell,k}(\omega)}\,d\omega \\ &= \frac{1}{(2\pi)^2}\int \hat{f}(\omega)U_j\left(R_{\theta_\ell}\omega\right)e^{i\left\langle x_k^{(j,\ell)},\omega\right\rangle}\,d\omega \end{aligned} \tag{3.15}$$

Figure 3.2 shows the diagram of decomposition by curvelets in frequency domain. This figure represents five decomposition level of resolution. The 1st level is the coarsest level, which is only decomposed by low pass filter. Therefore, the 1st level coefficients are non directional. Other levels are composed of angular wedges. The dotted region is one of the angular wedges in the 5th level coefficients. At scale 2^{-j}, the length of each wedge is $2^{-j/2}$, and the width of each wedge is 2^{-j}.

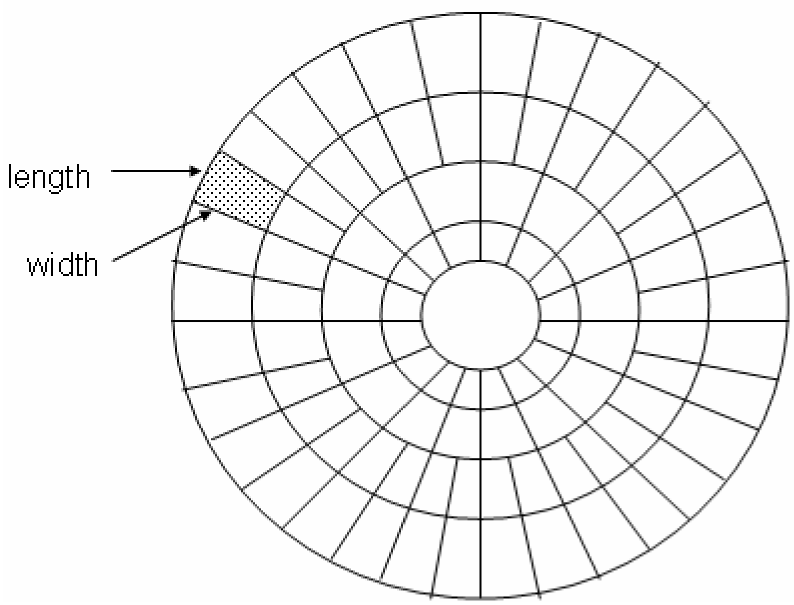


Figure 3.2 The curvelet tiling in the frequency domain

3.4. Implementation of Digital Curvelet Transform

In this section, we will describe the procedure for computing 2-D curvelet transform [21][28]. In short, the 2-D curvelet transform can be computed using unequally-spaced fast Fourier transform (USFFT). The Fast Fourier Transform library used in the program can be obtained in [29].

Some input parameters to the algorithm are described in Table 3.1, where *nx* and *ny* are the input image width and height, and *ns* is the number of image resolution scale for wavelet decomposition, which is a result of $\log_2(nx)$-3. In addition, Meyer wavelet [30] is used for the wavelet transform in the algorithm and n_φ is the degree of the Mayer window function.

nx	512
ny	512
ns	6
n_φ	3

Table 3.1 Input Parameter to curvelet codec

First of all, we convert the component of image, RGB, to YC_BC_R format to get the Luma component Y. And then we use the Luma component Y as the input of the curvelet transformation. The procedure of computing curvelet transform is composed of four steps and each of the steps will be described in the following sections. Section 3.4.1 describes the Fourier transform procedure which transforms the image inputs into frequency components. Section 3.4.2 introduces the band pass filtering process that decomposes the frequency data into several resolution scales. The polar scaling method is described in section 3.4.3. Finally, section 3.4.4 shows how the coefficients are converted back to spatial domain via inverse Fourier Transform.

3.4.1. Take Fourier transform into frequency domain

First of all, since we need to scale the image data by different resolution in the following steps, we must transform the image data into frequency domain. Assume that the input image data is in YC_BC_R format. To obtain Fourier samples of the image, a two-dimensional Fast Fourier Transform is applied on the Luma components(Y channel of data), and the transform coefficients is normalized by dividing by $(nx{\cdot}ny)^{0.5}$. As Figure 3.3 indicates, the low-band coefficients are centralized in the center of the image since this representation facilitates repetitive decomposition of the coefficients.

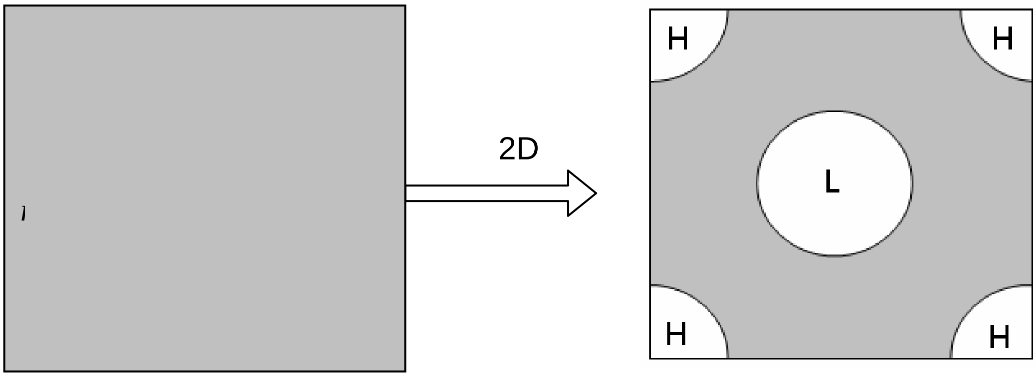
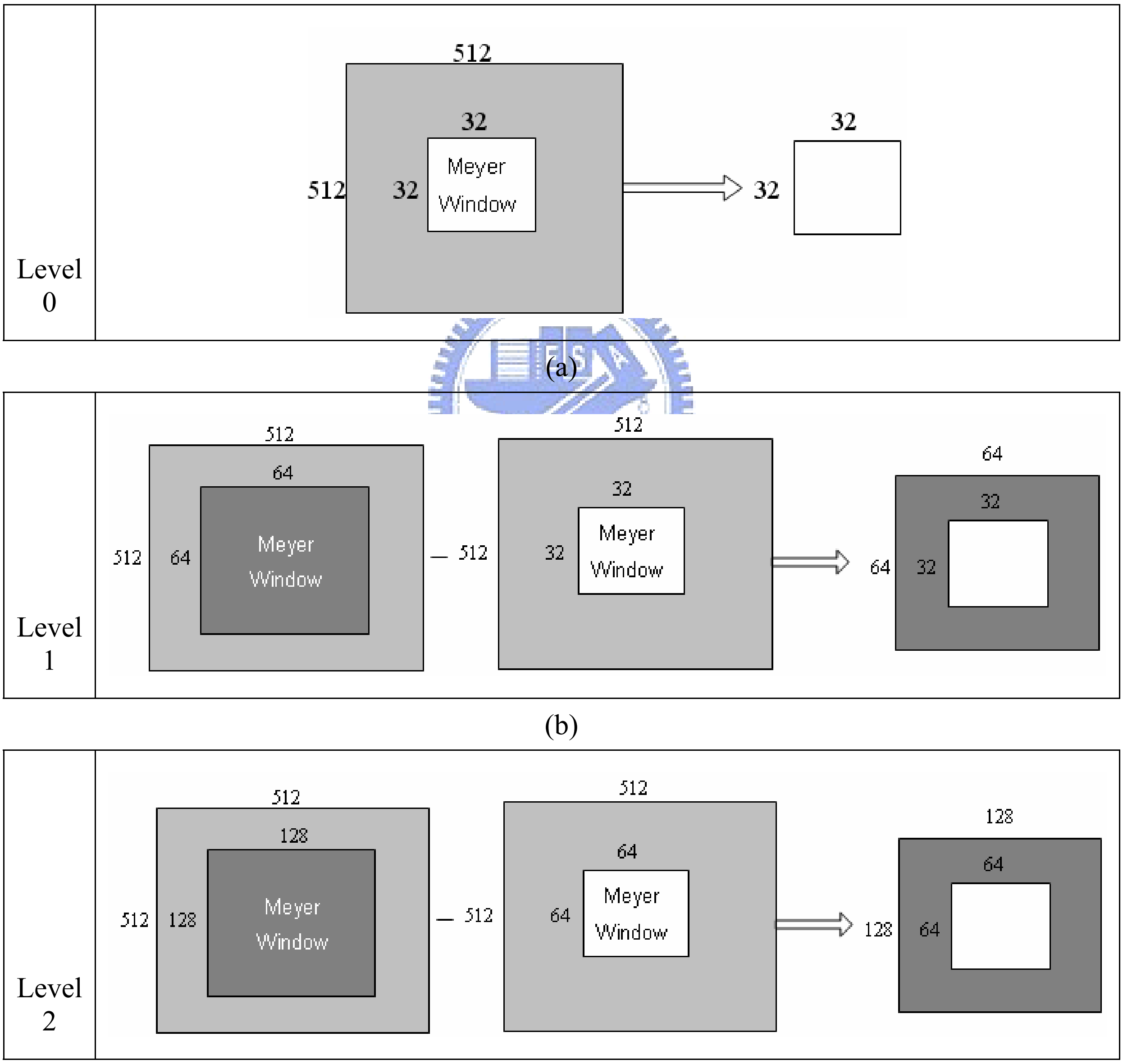


Figure 3.3 Decomposition of image into frequency domain

3.4.2. Band-pass filtering

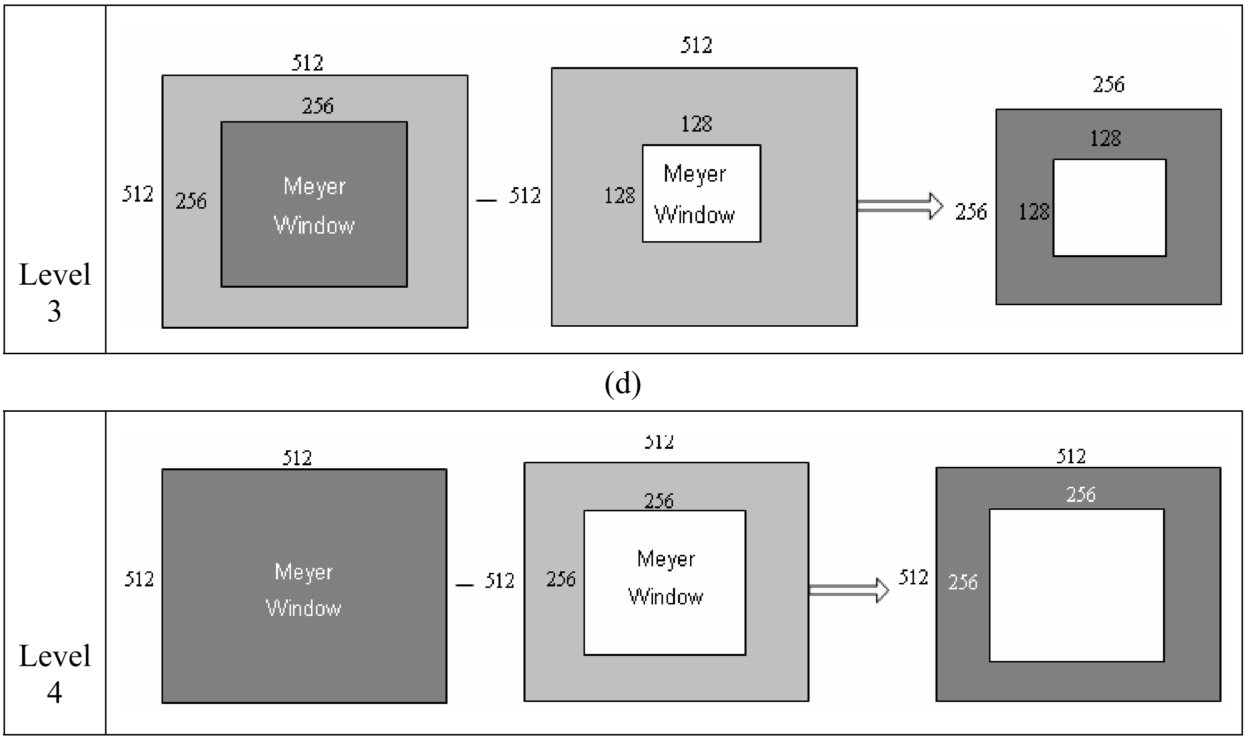
In the second step, we have to obtain frequency coefficients in different resolutions. First of all, we must create different levels of wavelet transform window function to decompose the coefficients obtained from previous step. Figure 3.4 illustrates how the band-pass filters are applied.



(a)

(b)

(c)



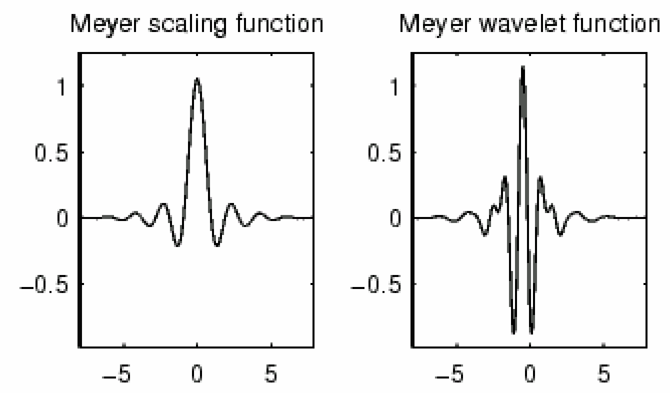
(d)

(e)

Figure 3.4 Decomposition of Meyer wavelet in resolution of scale 0 to scale 4

Figure 3.4 shows the decomposition of Meyer wavelet in different resolution. The coefficients in coarsest level are filtered by the low pass filter with size 32x32, and the procedure is indicated by Figure 3.4(a). Figure 3.4(b) to Figure 3.4(e) represents the coefficients in resolution of scale 1 to scale 4, respectively. For these scales, the filters are composed by the subtraction of two low-pass filters in order to form a band pass filters.

During the band-pass filtering procedure, the filters we apply are based on the low-pass Meyer window function. The scaling function and wavelet function of the Meyer window function is shown in Figure 3.5 [28] [30].



(a)

Wavelet Function :

$$\hat{\psi}(\omega) = (2\pi)^{-\frac{1}{2}} e^{\frac{i\omega}{2}} \sin\left(\frac{\pi}{2} v\left(\frac{3}{2\pi}|\omega|-1\right)\right) \quad if \quad \frac{2\pi}{3} \le |\omega| \le \frac{4\pi}{3}$$

$$\hat{\psi}(\omega) = (2\pi)^{-\frac{1}{2}} e^{\frac{i\omega}{2}} \cos\left(\frac{\pi}{2} v\left(\frac{3}{4\pi}|\omega|-1\right)\right) \quad if \quad \frac{4\pi}{3} \le |\omega| \le \frac{8\pi}{3}$$

and

$$\hat{\psi}(\omega) = 0 \qquad if \ |\omega| \notin \left[\frac{2\pi}{3}, \frac{8\pi}{3}\right]$$

where

$$v(a) = \mathrm{a}^4 (35 - 84\mathrm{a} + 70a^2 - 20\mathrm{a}^3), \ a \in [0,1]$$

(b)

Figure 3.5 Scaling function and wavelet function of Meyer window function

The detail procedure of decomposition is described as follows. First of all, we generate one-dimensional Meyer window of degree 3 by combining four basic parts, see Table 3.2. The one-dimensional Meyer window is leading with a zero coefficient and then followed by the 4 parts with the order 4 3 2 1 2 3 4. Therefore, the one-dimensional Meyer window will be composed of [zero 4 3 2 1 2 3 4]. The two-dimensional Meyer window is constructed by point-wise multiplication of two one-dimensional Meyer windows.

Part	The actual value of the filter
1	1…1
2	cos(πv/2(3a/2)) where *a* are dyadic points $(2^{i+1}/3\ldots2^{i}/3\text{-}1)/2^{i}\text{-}1$
3	cos(πv/2(3a/2))where *a* are dyadic points $(2^{i}\ldots2^{i+1})/2^{i}\text{-}1$
4	1…1

Table 3.2 Four basic parts of 1-D Meyer window

After generating the two-dimensional Meyer window, we filter the low frequency components by the Meyer window. To be more specific, the range of resulted low frequency components in resolution level i, sx_i and sy_i, can be calculated by Table 3.3. The filter coefficient we use in level i, where i are 1 to 4, is the square of coefficient i minus square of coefficient i-1, and then get its square root. In level 5, we use the coefficient of square roots of 1 minus square of coefficient 4. After the decomposition, we can get six different scales with size $sx_i \cdot sy_i$.

More over, the size of the coefficients is controlled by the level of their scale. And the relation of size and scale can be show as Table 3.3. To be more specific, sx_i and sy_i are width and height of level i.

Level	i	$sx_i = sy_i = min(2^{i+2},nx)$	*Window*(i)
0	i_0=3	32	i_0
1	i_1=4	64	$(i_1^2\text{-}i_0^2)^{0.5}$
2	i_2=5	128	$(i_2^2\text{-}i_1^2)^{0.5}$
3	i_3=6	256	$(i_3^2\text{-}i_2^2)^{0.5}$
4	i_4=7	512	$(i_4^2\text{-}i_3^2)^{0.5}$
5	i_5=8	512	$(1^2\text{-}i_4^2)^{0.5}$

Table 3.3 Parameters of each resolution level

One example of the band-pass filtered coefficients in each resolution is shown in Table 3.4.

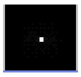
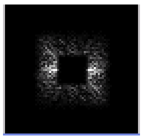
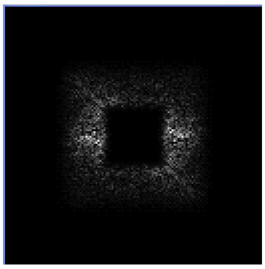
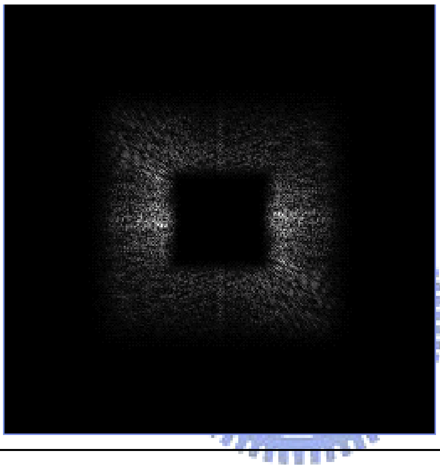
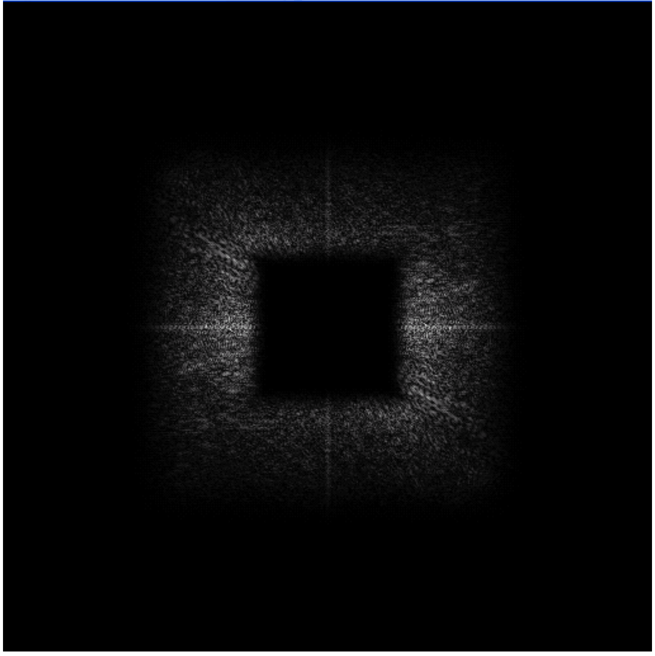
Result	Decomposition of Meyer wavelet
Level 0	
Level 1	
Level 2	
Level 3	
Level 4	

Table 3.4 The band-pass filtered coefficients of image Lena

3.4.3. Polar interpolation

The frequency coefficients in different resolutions computed by the band pass filter must be re-sampled to form directional frequency components. This can be done by interpolating the coefficients obtained from the previous procedure along vertical and horizontal directions. After resampling, the coefficients may be rearranged into four groups, namely west, east, north and south, based on the directions of directional decomposition.

For example, for the coefficients in the east quadrant, the whole procedure of the directional decomposition is shown in Figure 3.6.

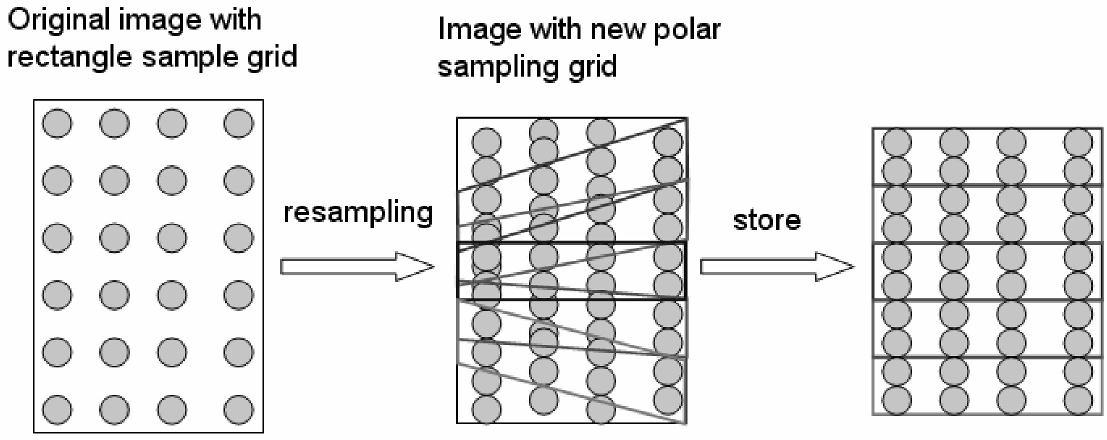

Original image with rectangle sample grid
resampling
Image with new polar sampling grid
store

Figure 3.6 Angular scaling in the East quadrant region

First of all, we can obtain the rectangle shaped coefficients in the East quadrant. Secondly, we apply column-wise one-dimensional inverse Fast Fourier Transform on the coefficients. And then we have to resample the coefficients into a shape of wedges (shown in the middle picture in Figure 3.6). From Figure 3.6, one can see that, for each column, x sampling gird stays the same, but y sampling grid must be rearranged along each column. Therefore, we have to calculate the new indices and interpolate the new coefficients. We obtain the new indices by performing a one-dimensional Meyer Window, see Table 3.5.

Meyer window is combined by the 4 parts	
Part 1	sin(πv/2(3*a*))
Part 2	sin(πv/2(3*a*))
Part 3	cos(πv/2(3*a*))
Part 4	cos(πv/2(3*a*))

Table 3.5 Formula of Meyer Window

Furthermore, the interpolation method is calculated by one sine window such as Equation 3.16. Finally, the resampled coefficients is stored in a rectangular shape region in transform domain for the purpose of easy accesses.

$$\mathrm{w} = \sin(\frac{\pi}{4}(1+\sin(\pi(t+\frac{1}{2})))) \cdot \sin(\frac{\pi}{4}(1+\sin(\pi(\frac{1}{2}-t)))) \qquad (3.16)$$

Note that for the east and west groups of coefficients the interpolation is done column-wise while for north and south groups of coefficients, the interpolation is done row-wise. Figure 3.7 shows the coefficients before and after the procedure of the directional decomposition. Coefficient in the right part is the result of the red ladder shaped region (East quadrant) in the 4th level resolution after applying the polar interpolation. To be more specific, the region which is framed in yellow designates the first angle in the resolution.

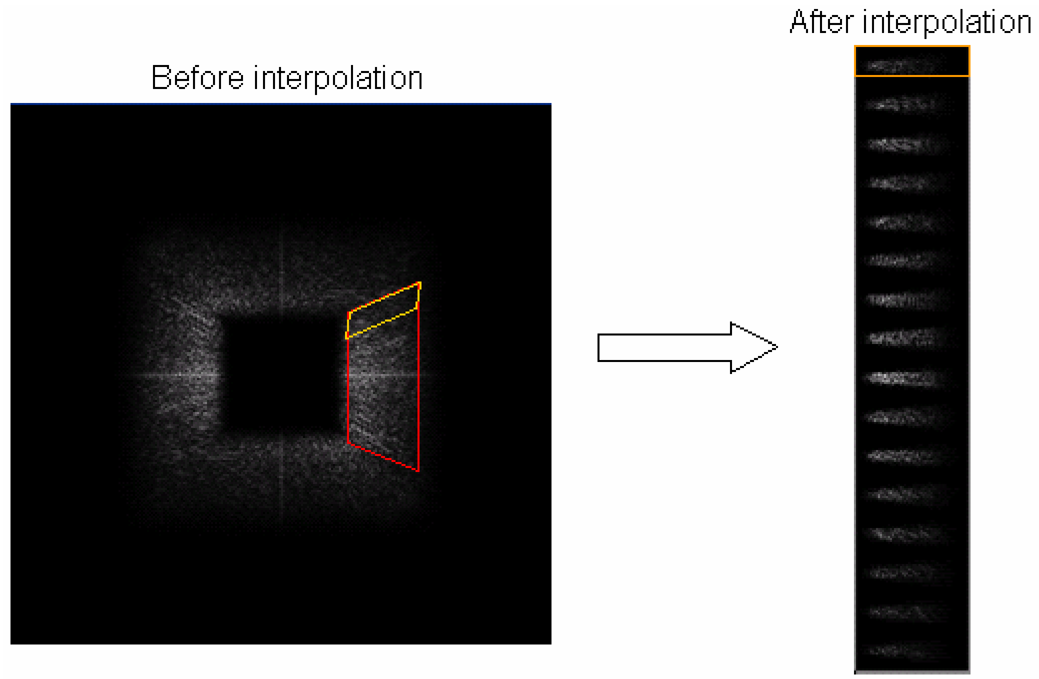


Figure 3.7 The directional decomposition in the 4th level coefficients.

3.4.4. Inverse Fourier transform

In the final step shown in Figure 3.8, two-dimensional Inverse Fast Fourier Transform is applied on the coefficients to transform the data back to spatial domain. The pixel data is normalized by multiplication with $(nx{\cdot}ny)^{0.5}$ in order to cancel the original normalization we apply on the coefficients.

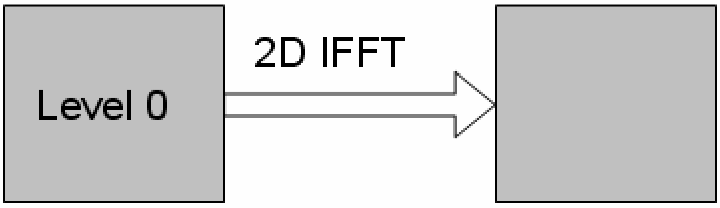


Figure 3.8 Perform 2D IFFT on lowest level

3.5. Interpretation of the Curvelet Transform Coefficients

In this section, we will give an interpretation to the curvelet coefficients in each level of resolution. The coefficients at the coarsest and the finest levels are not decomposed along different directions while the coefficients at other levels are decomposed along different angles.

3.5.1. Curvelet Coefficients in the Coarsest and the Finest level

As previous paragraph describes, the curvelet in the coarsest and the finest level of resolution do not contain the information of directional decomposition. To be more specific, the coarsest curvelet coefficients are the low-pass coefficients, and on the contrary, the finest curvelet coefficients are the high-pass coefficients.

Here we show an example of coarset level curvelet coefficients by stefan image with size 352×288. Size of the coarset level curvelet coefficients is 32×32. After normalizing the coarest level of curvelet coefficients, the image of coefficients is shown in Figure 3.9.

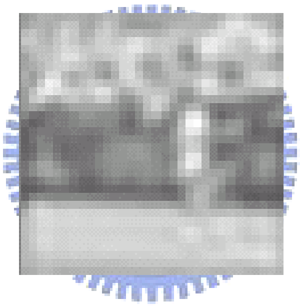
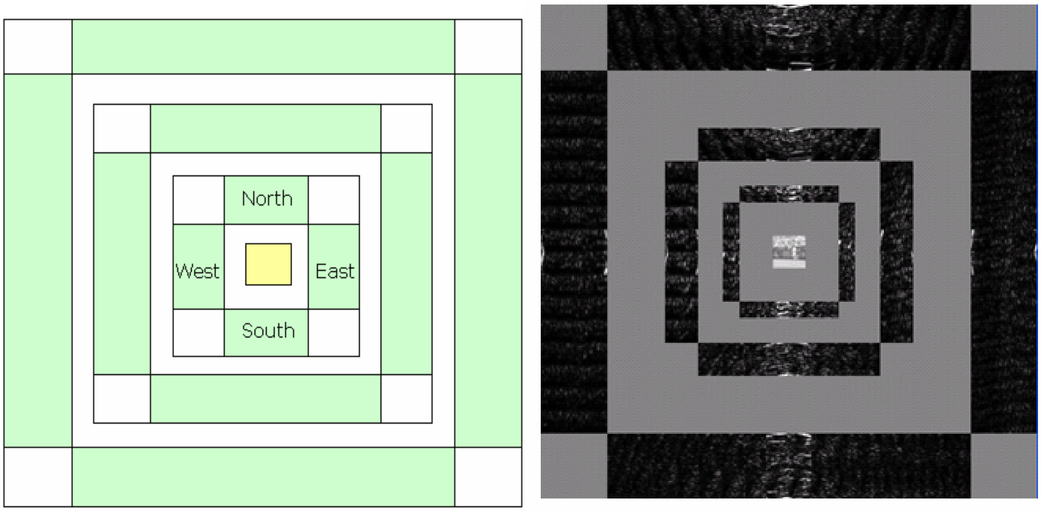


Figure 3.9 Result of coarsest level of curvelet coefficients

3.5.2. Curvelet Coefficients in the Middle Levels of Resolutions

Curvelets in the middle levels of resolution contain the information of directional decomposition. For each resolution scale, the coefficients in different directions (within a wedges area along different angles) are resampled according to their orientation. The number of angles analyzed at one scale is chosen by the resolution of the scale. For instance, there are 32 different angular wedges in the 2nd and 3rd levels, and 64 different angular wedges in the 4th and 5th levels.

As other multi-resolution based transforms, curvelet coefficients can be presented in a spatial image. Figure 3.10 (a) helps to define the position of curvelet coefficients. In the first step, we allocate the coefficients of the coarsest level in the middle of image. In the second step, the coefficients of one scale are separate in to 4 parts composed of North, East, South and West. More over, we separately put the coefficients around the coarsest level coefficients from low level to high level. Figure 3.11 shows the actual curvelet coefficient in 4 levels of resolution of the Stefan image.



(a) (b)

Figure 3.10 Coefficients of each resolution level

Moreover, an example of the mapping between the positions of the curvelet coefficients and the original image pixel position is illustrated in Figure 3.11. The mapping function of the coarsest-level coefficient is simple since it is just a subsampled version of the original image. However, the mapping of the other levels of coefficients is not as trivial since their samples are rearranged in the frequency domain. However, one can still derive the mapping function precisely by inversing the band pass filtering and the coefficient storage procedure described in this chapter. More details of the mapping function will be discussed in next chapter.

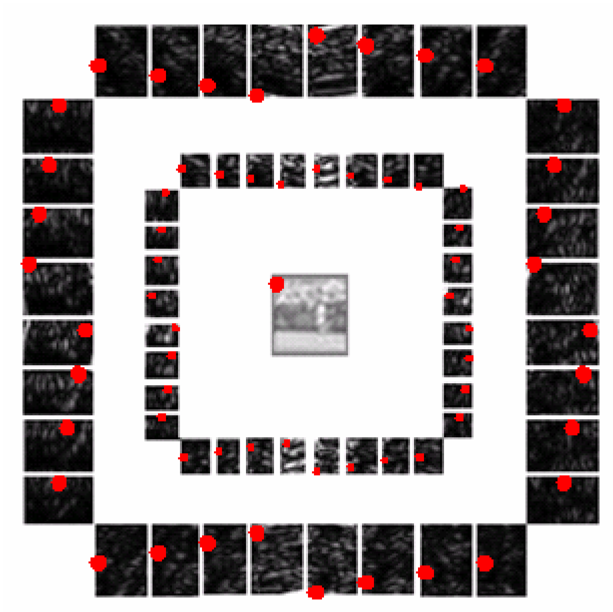


Figure 3.11 Position mapping of curvelet coefficients

4. Proposed Bit Allocation Framework

In order to design a video bit allocation policy that matches human perceptual behavior, we must design a measure to distinguish between structured and unstructured regions (or motion-compensated residuals). The texture structure can be classified by analyzing the distribution of the angular frequency components in the curvelet domain. In the proposed framework, the analysis is done by classification of the histogram of frequency components across different angles. The assumption is that for a structured region, the histogram should have clear peaks (large magnitude frequency components) at few angles. On the other hand, for an unstructured region, the histogram will be nearly uniformly distributed.

The organization of this chapter is listed as follows. In section 4.1, some analyses on curvelet coefficients at different resolutions are presented. In section 4.2, we describe the Otsu thesholding algorithm, which is used in the proposed framework for histogram classification. Section 4.3 formulates the proposed statistical measure that calculates the degree of texture structure randomness in a specified region. Finally, section 4.4 presents the proposed bit allocation scheme for MPEG-4 simple profile.

4.1. Analysis on curvelet transform coefficients

Although curvelet transform provides decomposed frequency components at different resolutions, the coefficients at the coarsest and the finest resolutions do not arrange frequency components according to the direction of the transform windows. In this thesis, these coefficients are referred to as the first group of coefficients. The second group contains coefficients at other intermediate resolutions. These coefficients are arranged according to the direction of the transform windows. In section 4.1.1, we will describe the detail information of the curvelet coefficients in the two groups. In section 4.1.2, we will show the spatial mapping between the original image positions and the corresponding curvelet coefficients.

4.1.1. The composition of the curvelet transform

We have glanced at some features in section 3.5. In this section, we will put more emphasis on the meaning of coefficients in a curvelet domain image and their relation to the original image.

First, we explain the components in the first group that do not contain the information of directional decomposition. For the coarsest-level coefficients, they contain the low frequency components and the size is diminished into fixed size of 32×32 pixels. Therefore, the position mapping between the original image and the transformed coefficients is a direct down-scale.

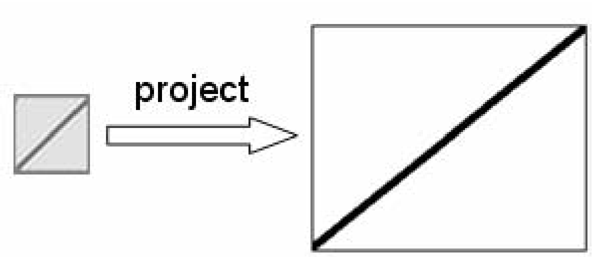


Figure 4.1 Position mapping of curvelet coefficients in the coarsest level resolution

Let's take the simple edge image that the only edge starts from right-up corner to left-bottom corner as an instance. Figure 4.1 shows the way of position mapping between curvelet coefficients in the coarsest level resolution and the original image. The image of normalized coarsest coefficients is listed left whose size is 32×32 pixels, and the original image is listed right whose size is 352×288 pixels. We can easily find out the original spatial properties are remained in the curvelet coarsest level coefficients. Similarly, Figure 4.2 is the image of normalized finest level coefficients. Curvelet coefficients in the finest level resolution are the collection of high frequency components of the original image, and their total size is the same as the original image. Therefore, the positional mapping is direct one-to-one mapping. In the proposed bit-allocation method, neither coefficient in the coarsest nor finest level of resolutions is used.

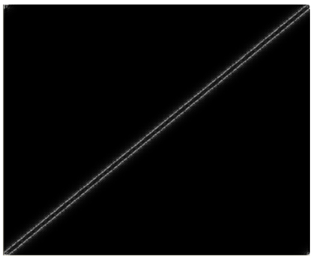


Figure 4.2 Curvelet coefficients in the finest level resolution

For the second group of coefficients that contain the information of directional decomposition, the number of levels depends on the original image size. For example, if the original image has 352×288 pixels, we can obtain three middle levels of resolution which are the 2nd level, 3rd level and 4th level respectively, see Figure 4.3.

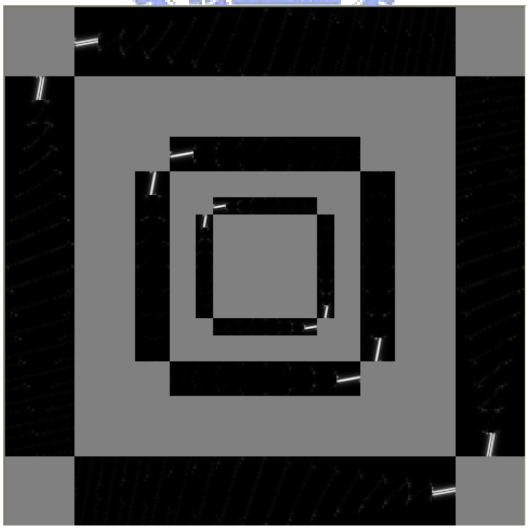


Figure 4.3 Curvelet coefficients in the 2nd, 3rd and 4th level resolution

For the 2nd and 3rd level, coefficients are separated into 32 angles. However, there are 64 angular wedges in the 4th level. That is to say, the total number of separated angles in one resolution is twice every two levels.

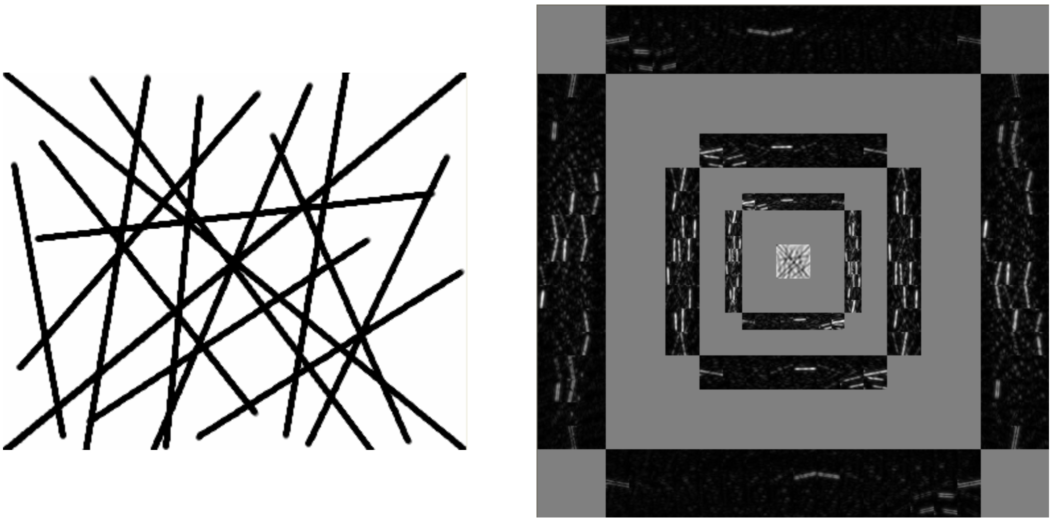
As Figure 4.3 shows, because the original image contains only an edge starts from the right-up corner to the left-down corner, we can easily find out the coefficients are centralized in the region of left-up and right-down. In short, curvelet coefficients centralize the energy into the orthogonal angles of direction of the curves.

More over, the position mapping function of the middle levels is different from that of the first group. In section 3.5.1, as Figure 3.11 shows, we have glanced at the relation of positions between the curvelet coefficients and the original image. Because the directional decomposition process re-samples the coefficients in the frequency domain, and the direction of re-sampling varies according to the specified angles, the actual positions of the final coefficients are shifted in some direction. We can classify the directions of shifting into two groups. First, for East and West quadrants, coefficients in the vertical direction are simply proportional, but coefficients are shifted by the angle of the orientation in the horizontal direction. Secondly, for North and South quadrants, coefficients in the horizontal direction are simply proportional, but coefficients are shifted by the angle of the orientation in the vertical direction.

4.1.2. Image type and the presentation of the related coefficient

As Figure 4.3 shows, we can easily understand the distribution of the curvelet coefficients. Since the original image only contains a simple edge, the positions where the coefficient appears are simple, too.

Let’s see a more complex example. If the original image contains multiple multi-directional edges, such as the image in Figure 4.4(a), the distribution of its coefficients is much more complicated.



(a) (b)

Figure 4.4 Curvelet coefficients in the 2nd, 3rd and 4th level resolution

In Figure 4.4(b), we can see that the coefficients are distributed in multiple angular wedges. Furthermore, it is natural that the angular wedges which the coefficient appears are different in each resolution. Therefore, curvelet transform can determine whether the curves in an image are complicated or not according to the directional decomposition in the middle levels of resolutions.

In our proposed scheme, we take each angular wedge in each resolution scale as a data unit. The actual process is to calculate the magnitude of one angular wedge in one resolution, and the magnitude becomes the representative value of the energy in the orientation in the resolution. Secondly, since we can get three resolution scales with directional decomposition, the coordinate formed by the magnitudes can be shown as in Figure 4.5.

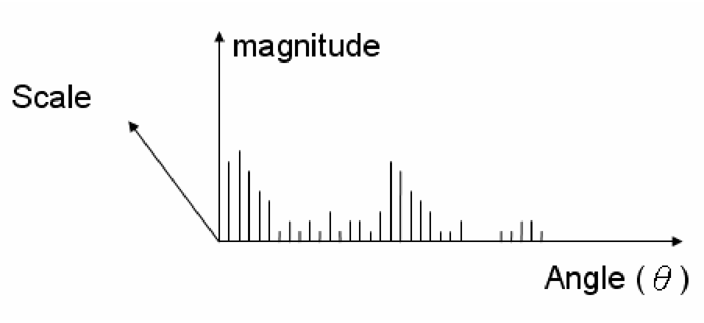


Figure 4.5 The coordinate formed by curvelet coefficients

It is a three dimensional coordinate which is formed by the scale, magnitude, and angle. The angle indicates the angle of the orientation which starts at 0° and ends at 360° in the direction of clockwise. However, the total angles in each resolution scale are different. To be more specific, the 2nd, 3rd level resolution contains 32 angles respectively, but the 4th level resolution contains 64 angles. In the plane of angle θ and magnitude, we can see the figure as histogram. Naturally, value of the magnitude is according to its quantity of coefficient in the orientation. If the value of magnitude is larger, it means more data in this direction. Therefore, we can analyze the composition of histogram to see whether the direction in the image is structured image or not.

As a result, we can take advantage of the property in curvelet transform to analyze the input video data in our proposed bit-allocation scheme. In next section, we will introduce the Otsu algorithm to help analyzing the curvelet coefficients for the sake of determine whether a small area in an image contains complicated curves or not.

4.2. The Otsu Algorithm

For picture processing, the technique of selecting histogram threshold is very useful for many applications, such as object extraction or edge detection [31]. Therefore, there are a variety of techniques proposed for threshold selection. In this section, we will introduce a typical threshold selection algorithm from gray-level histograms. The method is proposed by

Nobuyuki Otsu in 1979 [32].

Generally speaking, given a histogram of an image, there is a deep and sharp valley between every two neighbor peaks. The position of the bottom of each valley is the threshold we want to obtain. In the algorithm proposed by Otsu, the histogram threshold can be derived form the viewpoint of discrimination analysis.

First of all, we assume the number of gray level in an image is L. The total number of pixels N is the summation of n_i for level i=1, 2,…, L. The probability distribution is as Equation 4.1 shows:

$$p_i = \frac{n_i}{N}, \qquad p_i \geq 0, \qquad \sum_{i=1}^{L} p_i = 1 \tag{4.1}$$

Secondly, we can separate all pixels into two classes by a threshold at level k. C_0 is the group which contain the pixels with level 1 to level k, and C_1 contain the pixels with level k+1 to level L. The probabilities of class occurrence of C_0 and C_1 are listed in Equation 4.2.

$$\begin{aligned} \omega_0 &= \Pr(C_0) = \sum_{i=1}^{k} p_i = \omega(k) \qquad \text{where} \quad \omega(k) = \sum_{i=1}^{k} p_i \\ \omega_1 &= \Pr(C_1) = \sum_{i=k+1}^{L} p_i = 1 - \omega(k) \end{aligned} \tag{4.2}$$

Moreover, the probabilities of class mean levels of C_0 and C_1 are listed in Equation 4.3.

$$\begin{aligned} \mu_0 &= \sum_{i=1}^{k} i \Pr(i \mid C_0) = \sum_{i=1}^{k} i p_i / \omega_0 = \mu(k)/\omega(k) \qquad \text{where} \quad \mu(k) = \sum_{i=1}^{L} i\, p_i \\ \mu_1 &= \sum_{i=k+1}^{L} i \Pr(i \mid C_1) = \sum_{i=k+1}^{L} i p_i / \omega_1 = \frac{\mu_T - \mu(k)}{1 - \omega(k)} \end{aligned} \tag{4.3}$$

For any choice of k, we can get the relation in Equation 4.4.

$$\omega_0\mu_0 + \omega_1\mu_1 = \mu(\mathrm{L}) = \sum_{\mathrm{i}=1}^{\mathrm{L}} \mathrm{i}\,\mathrm{p}_\mathrm{i} = \mu_\mathrm{T} \ , \qquad (4.4)$$
$$\omega_0 + \omega_1 = 1$$

Based on the formula and variable above, we can introduce the discriminate criterion measure, between-class variance, in Equation 4.5 to evaluate the threshold k [33].

$$\sigma_\mathrm{B}^2 = \omega_0(\mu_1 - \mu_\mathrm{T}) + \omega_1(\mu_1 - \mu_\mathrm{T}) = \omega_0\omega_1(\mu_1 - \mu_0)^2 \qquad (4.5)$$

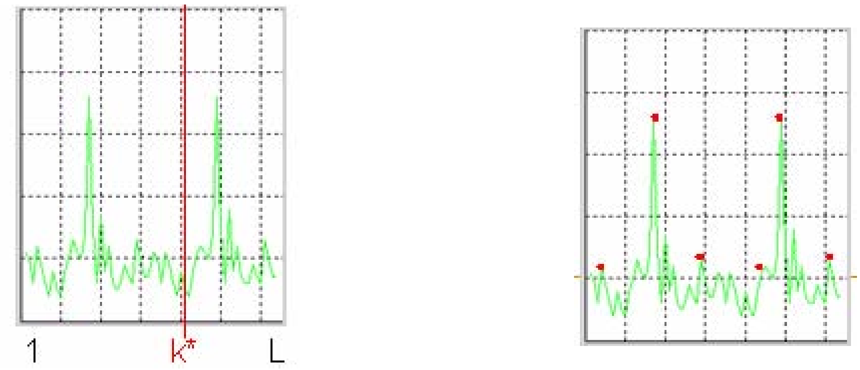
Since we can assume that if the two classes are distinguished by good threshold, the number of between-class variance must be large. Therefore, we can easily obtain the conclusion that the best threshold that separates the two groups must derive the maximum between-class variance over other thresholds. This relation can be formulated in Equation 4.6 where k^* is the best solution of histogram threshold.

$$\sigma_\mathrm{B}^2(\mathrm{k}*) = \max_{1\le \mathrm{k} < \mathrm{L}} \sigma_\mathrm{B}^2(\mathrm{k}) \qquad (4.6)$$

Furthermore, the method can easily be extended into a multi-threshold case. Equation 4.7 shows the example of two-threshold selection method which can produce four peaks in the gray-scale histogram.

$$\sigma_\mathrm{B}^2(\mathrm{k}_1^*, \mathrm{k}_2^*) = \max_{1\le \mathrm{k}_1 < \mathrm{k}_2 < \mathrm{L}} \sigma_\mathrm{B}^2(\mathrm{k}_1^*, \mathrm{k}_2^*) \qquad (4.7)$$
$$\text{where } \mathrm{C}_0 \text{ for } [1,\cdots,\mathrm{k}_1], \mathrm{C}_1 \text{ for } [\mathrm{k}_1+1,\cdots,\mathrm{k}_2], \mathrm{C}_2 \text{ for } [\mathrm{k}_2+1,\cdots,\mathrm{L}].$$

In the previous section, we know that the input histogram is the curvelet coefficient in one resolution scale. The level indicates the angle of the angular wedges. In other words, the input data is the histogram of 32 or 64 level histogram. Figure 4.6(a) represents the result of single threshold produced by Otsu algorithm where k* is the best threshold to separate the histogram into two groups.



(a) **(b)**

Figure 4.6 The histogram separated by Otsu threshold selection method

Furthermore, in our proposed scheme, we extend the threshold selection method to five-thresholds. As a result, we can obtain six classes in the whole histogram. And then as Figure 4.6(b) shows, we will compute the value of the peak in each class. As a result, the rise and fall of each histogram is different according to its composition of data in each orientation. We can assume that if the variety of histogram is bigger, the direction of edges in the image is more complicated. Therefore, we can use the six mountain peaks in the histogram to indicate the degree of complication of the image. The actual computational method is described in the continuous section.

4.3. Statistical method to analyze the coefficients

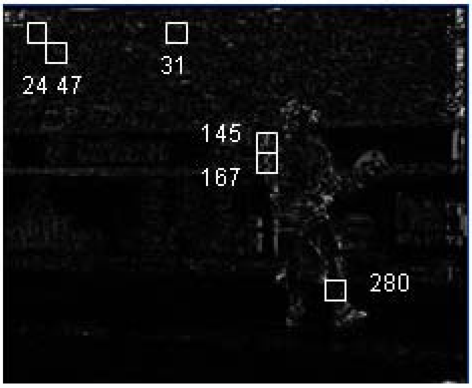
In the section, we use the coefficient of variation to measure the degree of variation of our histogram. The coefficient of variation (CV), which is also called “relative variability”, is a measure of dispersion of a probability distribution [34]. To be more specific, CV represents the ratio of standard deviation to the mean value.

We do not use standard deviation to analyze our data because has interpretable meaning under the condition that the mean value of every sample is the same. In other words, standard deviation represents the degree of variability relative to the mean value. However, in our histogram of curvelet coefficients, the average magnitude of each angular wedge is definitely different. Therefore the coefficient of variation is used instead.

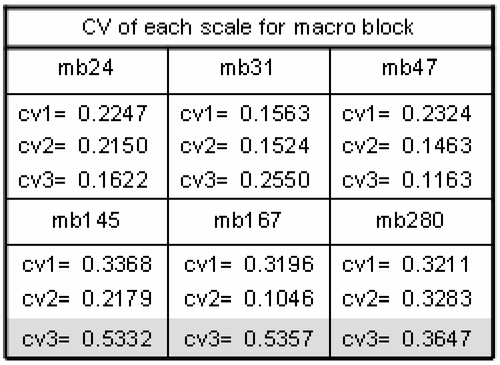
$$CV = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\bar{x}} \quad \text{where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.8}$$

Equation 4.8 shows the formula of the coefficient of variation (CV). It is easily seen that CV is the value the standard deviation divided by the mean. The measurement of the coefficient of variation is better in datasets with markedly different means or with different units of measurement. Our input dataset just match the first type.

Here we list six classical examples of histogram and its coefficient of variation in Figure 4.7. We take one residual frame of Stefan sequence with resolution CIF as our example. First of all, we divided the image into 396 blocks with size 16 by 16. Therefore, we can obtain curvelet histogram of each block in each resolution level.



(a)



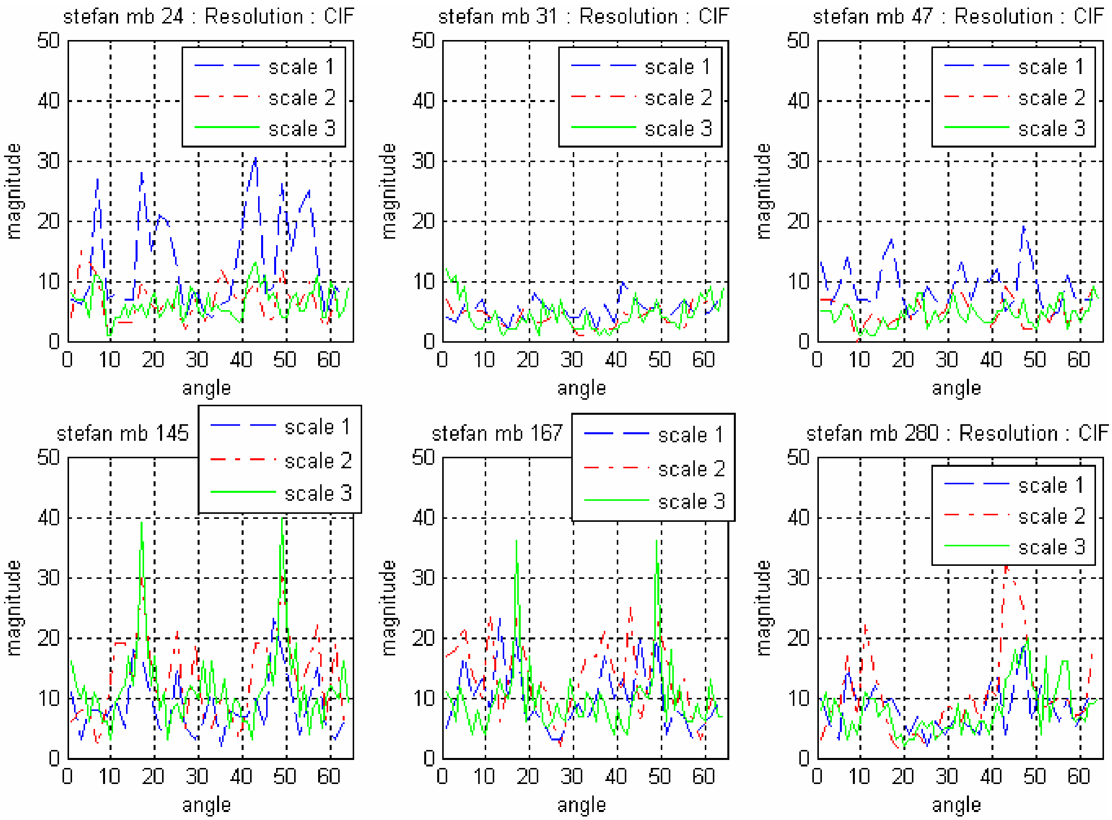
CV of each scale for macro block		
mb24	mb31	mb47

cv1= 0.2247
cv2= 0.2150
cv3= 0.1622
	cv1= 0.1563
	cv2= 0.1524
	cv3= 0.2550
		cv1= 0.2324
		cv2= 0.1463
		cv3= 0.1163

mb145	mb167	mb280

cv1= 0.3368
cv2= 0.2179
cv3= 0.5332
	cv1= 0.3196
	cv2= 0.1046
	cv3= 0.5357
		cv1= 0.3211
		cv2= 0.3283
		cv3= 0.3647

(b)



(c)

Figure 4.7 The histogram and coefficient of variation of six blocks in Stefan image

Figure 4.7(a) shows actual positions of blocks we select. In Figure 4.7(b), it lists the coefficient of variation of each block in each level. To be more specific, cv1 means the coefficient of variation of the first level with directional decomposition, and cv2 means the coefficient of variation of the second level with directional decomposition, etc. And the curvelet histogram of the relative block is presented in Figure 4.7(c). As the histogram shows, one can see that the value of variation and the number of peaks exist some relation. Based on the number of peaks and the magnitudes of peaks in the histograms, the image region can be classified into several types of images. The first type is that the region doesn't contain any clear edges at all, such as block 31, and the second type is that the region contains many small edges, such as block 24 in the first resolution level with directional decomposition. Both these two kinds of images are considered as unstructured image since their texture has complex edges in them. On the other hand, if the magnitudes of the peaks in the histogram are strong, it means that the distribution of the edges in the block is simple and clear. For example, block 145, 167, and 280 in the third resolution level shows such case.

4.4. Proposed Bit Allocation Scheme

The bit allocation algorithm for video coding must determines the quantization parameter based on the visual importance of a coding block. The input to the bit-allocation algorithm is a macroblock of video data. For intra-coded blocks, the input data is the image pixels while for inter-coded blocks, the input data is the motion-compensated error residuals. After curvelet transform, one can obtain the coefficients that are separated by their direction of contour and resolution. And then we can directly take each angular wedge in each resolution scale as a data unit as described in section 4.1. To be more specific, we integrate the coefficients by calculating the magnitude of one angular wedge in one resolution, and the magnitude becomes the representative value of the energy in the orientation in this resolution. As a result, the display of curvelet coefficients can be expressed as in Figure 4.5 a three dimensional coordinate which is formed by the scale, magnitude and angle.

Next, as described in section 4.2, for each plane of angle and magnitude, we analyze the mountain peaks of the histogram. Each mountain peak represents the gathering of direction of edges. To classify the complexity of the region, the coefficient of variation (CV) to analyze the mountain peaks of the angular histogram. On one hand, if the value of each mountain peak in one histogram varies slightly, it will be represented by a small CV and it means that the direction of edge in the block is not obvious. Of course we can indicate the image as an unstructured region. On the other hand, if the value of each mountain peak in one histogram varies dramatically, it will be represented by a large CV. It means that the direction of edge in the block is obvious, and we can indicate the image as a structured region.

In the proposed method, we separate all images into three groups. They are group of unstructured regions, group of structured region, and group of well structured regions. Since human eyes are less sensitive to images of unstructured regions than images of structured regions, we can adjust the way of bit allocation according to our analysis of the image. As a result, images of unstructured regions can be seen as unimportant regions, so we can diminish bits of the regions in a compression technique. On the other hand, we can increase bits of the well structured regions in order to enhance the performance of compression.

Figure 4.8 shows the block diagram of the proposed bit allocation algorithm. Blocks in the first line is the original encoding procedure of an MPEG-4 simple profile encoder, and blocks in the second line is the modified encoding flow.

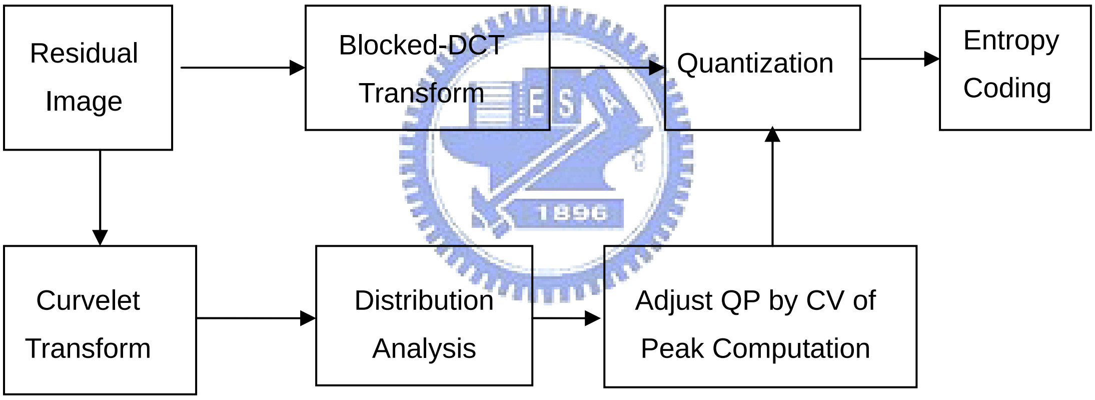


Figure 4.8 Block diagram of the proposed bit allocation model

In the process of determining the complexity of images and adjusting QP based on the CV of histogram peaks, Equation 4.8 is proposed.

$$d_{QP} = \mathrm{round}(\frac{CV_{max} - T_{min}}{T_{maxu} - T_{min}} - 0.5) \quad \text{where CVmax is the maximum of CV} \tag{4.8}$$

If d_{QP} is equal to 1, then check the minimum of CV:

If CV_{min} is smaller then T_{maxl}, then change the value of d_{QP} from 1 to 0

In Equation 4.8, T_{min} controls the boundary of decreasing QP. If the maximum coefficient of variation in three resolution levels (CV_{max}) is less than T_{min}, the macro block is considered unstructured region, and QP of the macro block is reduced.

On the other hand, the way to judge whether the macro block is a strictly structured region or not is similar but contains one extra condition. T_{maxu} and T_{maxl} control the boundary of increasing QP. If the maximum coefficients of variance (CV_{max}) is larger than T_{maxu} and minimum coefficients of variance ($CV_{min.}$) is not less than T_{maxl}, the macro block is considered strictly structured region, and QP of the macro block is increased.

By using the formula to computing updates of quantization parameters for each macroblock, we can obtain three kinds of updated quantization parameters. The first group of quantization parameter is the same as the original quantization parameter. This means that the composition of image is a normally structured region, and we do not have to increase or decrease its bits. The second group of quantization parameter corresponds to the original quantization parameter plus one. This type of regions means that the composition of image does not contain obvious edges, so it is typically unstructured regions. We can decrease its bits and the compression result does not cause obvious degradation to human eyes. In addition, we can allocate the saving bits to other regions that the human observers are more sensitive to. And this is the behavior of the third type. The third group of quantization parameter equals the original quantization parameter minus one. This type of regions means that the composition of image contain clear edge structures, which is referred to as well structured regions. Therefore, we can increase its bits to enhance performance by human eyes, since the improvement of visual quality in this kind of region can dramatically catches human eyes.

4.5. Determination of the Weighting Threshold

In the section, the selection of thresholds mentioned in previous section is described. In general, the degree of presentation of signal discontinuity contains a relationship to the sampling frequency. Therefore, in order to consider the weighting of each resolution scale, we must analyze the range of minimum and maximum sampling frequencies in each directional resolution scale as listed in Table 4.1 .

		Min		Max	
Scale	Direction	F (cycles/samples)	T (cycle)	F (cycles/samples)	T (cycle)
1	Horizontal	16/352	22	32/352	11
	Vertical	16/288	18	32/288	9
2	Horizontal	32/352	11	64/352	5.5
	Vertical	32/288	9	64/288	4.5
3	Horizontal	64/352	5.5	128/352	2.75
	Vertical	64/288	4.5	128/288	2.25

Table 4.1 Analysis of frequency components.

From Table 4.1, one can see that the proportions of mean frequencies in these three scales are 1:2:4. Consequently, the formula for the overall CV (combining information form all resolution scales) is computed as in Equaltion 4.1. Note that the sum of the coefficients is one and CV_1, CV_2, and CV_3 are the CV's for different resolution scales.

$$CV = 0.14 \cdot CV_1 + 0.29 \cdot CV_2 + 0.57 \cdot CV_3 \tag{4.1}$$

Next, we must determine the threshold CV for structured and unstructured regions. The threshold is estimated by a pre-analysis step for each group of picture (GOP). One example of GOP structure which contains nine frames is shown in Figure 4.9.

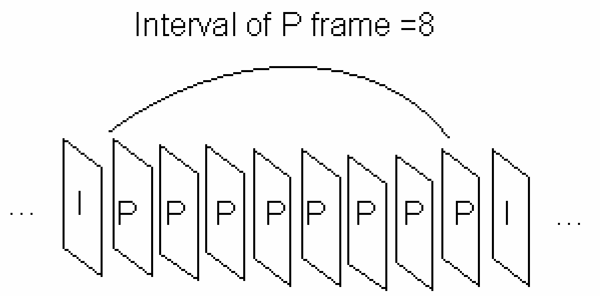


Figure 4.9 GOP structure of a video sequence.

Two possible methods are tested to determine the CV threshold at a particular scale. Both methods are based on estimating the boundaries between well structured regions (SR) and unstructured region (USR).

Method I: the threshold of SR is computed as the value of CV that makes the regions with top 1/3 CV values being counted as well structured regions. Then, the threshold for the unstructured regions is selected so that it decreases the bitrate for the unstructured regions (n_1%) so that the overall bitrate stays the same.

Method II: the threshold for SR is computed as that in method I, and the threshold for USR is determined so that blocks of the last n_2% number are considered as USR.

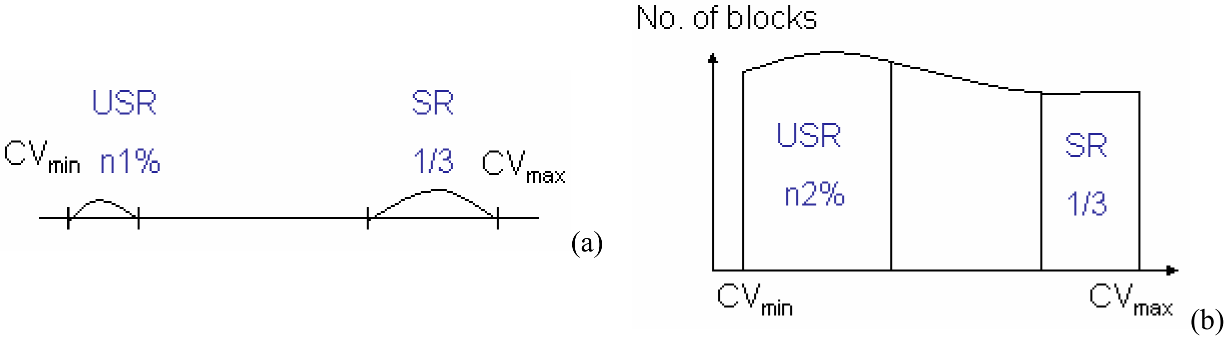


Figure 4.10. Threshold selection of CV.

5. Experiment and Analysis

In this section, the performance of the proposed curvelet-based bit allocation scheme is investigated using the MPEG test sequences STEFAN, FOOTBALL, and BUS in CIF resolution. An MPEG-4 Simple Profile encoder is used for the experiments.

5.1. Result of the proposed bit allocation scheme

The goal of our proposed scheme is to achieve better visual quality given same the target bitrate constraint. That means that we have to enhance the performance of the regions that human observers are more sensitive to by allocating more bits to them. And the process should not increase total bits a lot. In this chapter, we conducted some experiments to show the efficiency of the proposed algorithm. PSNR and SSIM are used as measures to evaluate the performance of the video. For divided regions of image, we only use PSNR to evaluate their visual quality since SSIM is good at extracting structural information and it does not work well in small size regions.

Three MPEG test sequences STEFAN, FOOTBALL, and BUS in CIF (352×288) resolution are used to test the performance of the proposed curvelet-based bit allocation scheme. The first 120 frames are used to conduct the experiment for each sequence. Table 5.1 lists the setting parameters of the experiments. T_{maxu}, T_{maxl} and T_{min} are the manually selected thresholds of coefficient of variation (CV) of curvelet coefficients used in the image region classification algorithm. Results with automatic selection of thresholds will be presented in section 5.2. QP is the default quantization parameter, and Luma Bitrate is the original bitrate of luma components produced by the MPEG-4 simple profile without the proposed scheme.

case	sequence	T_{maxu}	T_{maxl}	T_{min}	QP	Luma Bitrate
1	Stefan	0.62	0.27	0.23	12	616
2	Football	0.58	0.25	0.24	9	611
3	Bus	0.57	0.28	0.18	9	1071

Table 5.1 Parameter settings of the three experiments

First of all, the resulted data of the three sequences are listed in Table 5.2. The field “Original” indicates that the sequence is compressed by the original MPEG-4 simple profile, and the field “Modified” indicates that the sequence is compressed by the proposed bit allocation scheme. QP is the value of quantization parameter for the whole sequence. Furthermore, QP from the proposed algorithm varies within a range of ±1 from the original encoder. Total Bitrate is the total bitrate of the whole sequence. Luma Bitrate indicates bitrate of luma components of the whole sequence. Next, Header Bitrate and Chrome Bitrate mean bitrates of header and chrominance components of the whole sequence, respectively.

Sequence	Stefan		Football		Bus	
Type	Original	Modified	Original	Modified	Original	Modified
PSNR	30.16714	30.17565	33.84276	33.86667	31.75857	31.75973
SSIM	0.92485	0.92554	0.88017	0.88093	0.90463	0.90473
QP	12	11~13	9	8~10	9	8~10
Total Bitrate	790	796	1294	1299	1250	1254
Luma Bitrate	618	616	928	927	1071	1071
Header Bitrate	144	152	273	279	161	166
Chroma Bitrate	27	27	92	92	16	16

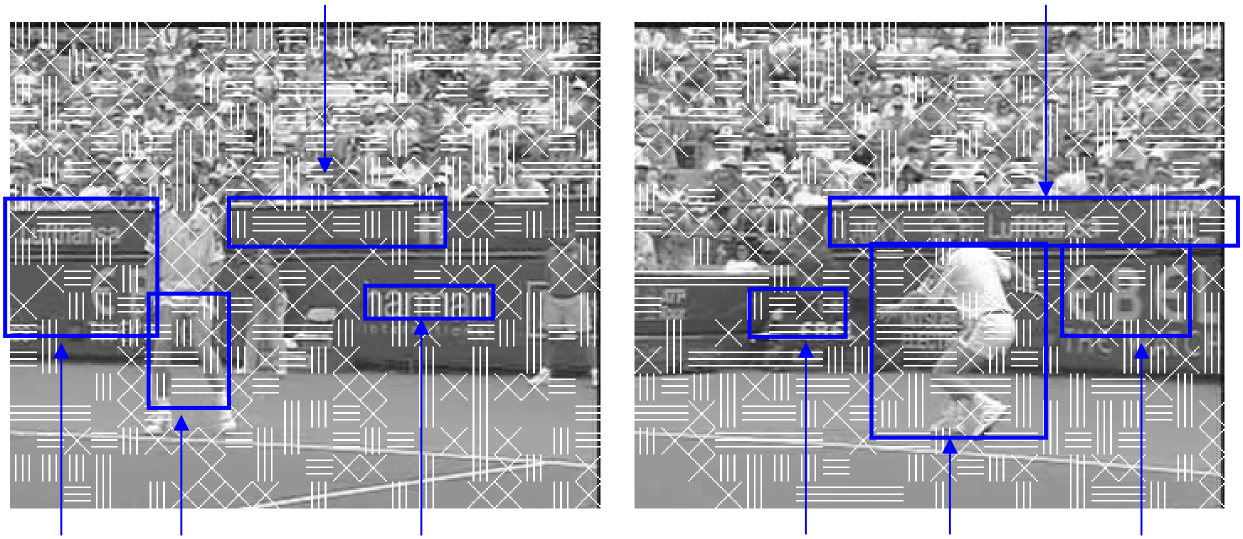
Table 5.2 Resulted data of the three sequences

As the numbers in gray cells show, the proposed scheme increases visual quality without increasing the bitrates of luminance components. Since our proposed scheme allocates different number of bits to macro blocks according to their composition of directional edges, we may increase bits of header data definitely. Therefore, the experiments only focus on the variation of bits of luminance components.

Next, for each sequence we list two kinds of typical frames to analyze the result of visual quality and bits our proposed scheme causes. Figure 5.1, Figure 5.2 and Figure 5.3 show the result of visual quality in three sequences respectively. In these figures, we divide whole frame into 396 macro blocks with size 16x16. And we label the macro blocks of three kinds of properties we are interested in. First of all, macro block in label | indicates the block which has better visual quality in PSNR measurement and more bits of luminance components than the original MPEG-4 simple profile. Secondly, macro block in label — indicates the block which has better visual quality in PSNR measurement but less bits of luminance components. Thirdly, macro block in label ╳ indicates the block which has worse visual quality in PSNR measurement and less bits of luminance components. We can easily see the distribution of visual quality by dividing the frame into two groups. Group of region with label ╳ indicates worse visual quality and group of region with label — and | indicate better visual quality than the original image.

For the Stefan sequence, human observers may pay special attention to tennis player and the area with obvious edges such as words on the wall. On the other hand, the regions that audiences on the grandstand and the flat regions are mostly human observers are not sensitive to relatively. In Figure 5.1(a), the major movement in the 51st frame is the tennis player moving towards the right hand side. In the proposed scheme, the regions with clear and obvious directional edges will be considered structured region. As a result, performances of this kind of regions such as tennis player's legs, words on the wall and lines on the ground are mostly enhanced. Nevertheless, the regions of audiences on the grandstand and the flat regions will be seemed to unstructured regions since their directions of edges are complicated.

So bits of these regions are usually decreased and it may cause worse visual quality.



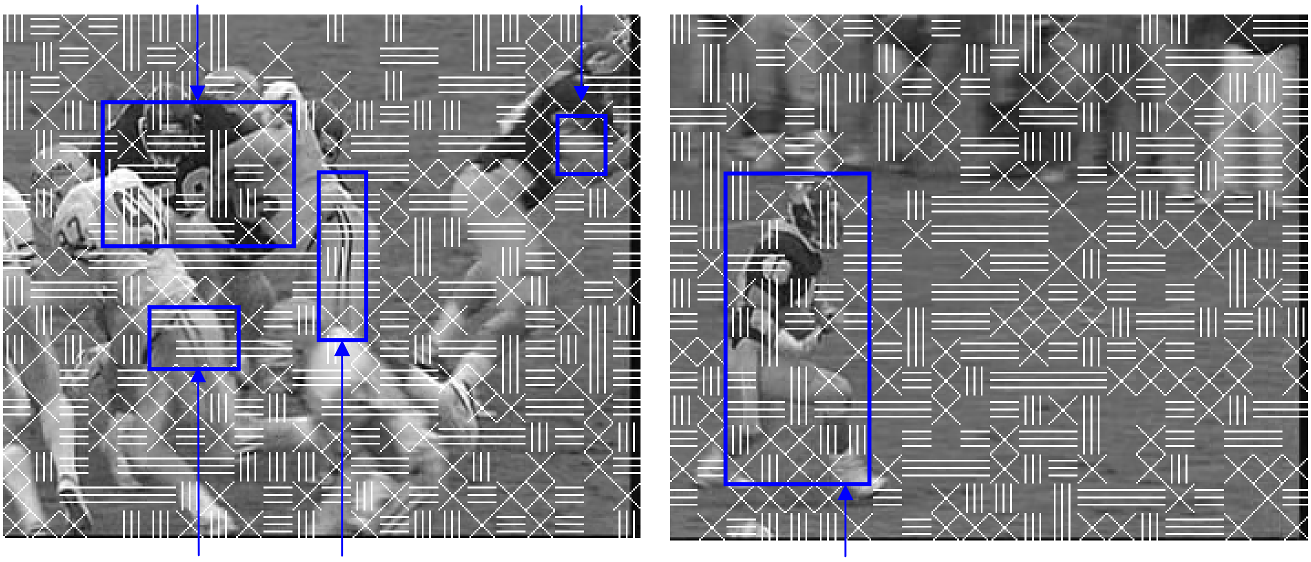
(a) (b)

Figure 5.1 Comparison of visual quality in Stefan Sequence
(a)The 51st frame in Stefan. (b) The 96th frame in Stefan.

In Figure 5.1(b), the 96th frame, the major movement in the 96th frame is the tennis player waving his rocket. Therefore, human eyes may notice the area of tennis player's whole body and the area with obvious edges such as words on the wall and lines on the ground. Performances of these kinds of regions are mostly enhanced. And similar as in Figure 5.1(a), bits of the regions of audiences on the grandstand and the flat regions are usually decreased and it may cause worse visual quality. Table 5.3 shows ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of audience is smaller than others.

	51st frame		96th frame	
PSNR	Better	Worse	Better	Worse
Audience	78	54	65	67
Words	20	4	30	9
Legs	9	2	-	-
Whole body	-	-	12	5

Table 5.3 Ratio of improvement in Stefan sequence



(a)

(b)

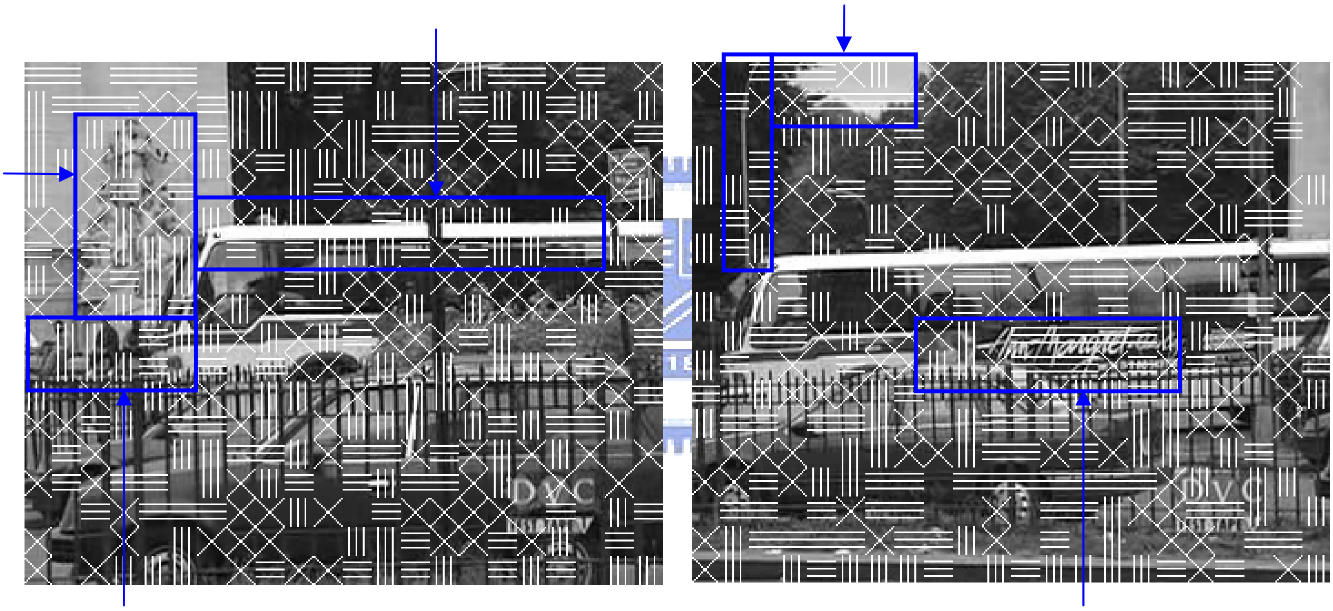
Figure 5.2 Comparison of visual quality in Football Sequence (a)The 27th frame in Football. (b) The 116th frame in Football.

Next, in whole Football sequence, human observers may pay more attention on the area of football and football player than the area of grass. Moreover, obvious edges on football player such as numbers on their sports coats or stripes on their pants may attract human eyes dramatically. In Figure 5.2(a), the major movement in the 27th frame is the football players competing for the football. In the proposed scheme, the performances of the regions we said above that humans may be more sensitive to, numbers on their sports coats or stripes on their pants, are mostly enhanced. Nevertheless, bits which are allocated to the regions of too complicated grass and the flat regions are usually decreased because these regions may be seemed to unstructured regions. And the processing may cause worse visual quality of these regions. Here we select the other kind of frame in Football sequence to analyze its result. In Figure 5.2(b), the major movement in the 116th frame is the football players running towards right with the football in his hand. In this frame, human may pay attention to the only football player and the football. In our proposed scheme, the performances of the regions we said above mostly enhanced. However, for less important regions, such as the regions of too complicated grass and flat regions, their bits are usually decreased and their visual quality may be reduced.

	27th frame		116th frame	
PSNR	Better	Worse	Better	Worse
Grass	70	57	110	102
Numbers	12	2	-	-
Stripe	13	1	-	-
Whole body	-	-	21	10

Table 5.4 Ratio of improvement in Football sequence

Table 5.4 shows ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of grass is smaller than others.



(a) (b)

Figure 5.3 Comparison of visual quality in Bus Sequence (a)The 25th frame in Bus. (b) The 92nd frame in Bus.

Last, in Bus sequence, human observers may not pay more attention on the area of complicated background such as the trees on the top of image, and complicated foreground such as railings and still car. On the other hand, the moving bus and various backgrounds are more attracted to human eyes than the region we said above generally. Here we select two different kinds of scenes of bus sequence to analyze our result. In Figure 5.3(a), the bus is just passing through the pillar with sculpture. Therefore, human observers may take their on the regions of the sculpture, human under the sculpture, the head of bus and the top of bus. In our proposed scheme, visual qualities of these regions are mostly enhanced. In the 92nd frame, as Figure 5.3(b) shows, the regions that human observer may notice a lot are listed as follows: the advertisement with photograph and words on the bus, the street light near the head of bus and the region that sky and trees are associated with. Visual qualities of these regions are mostly enhanced. Nevertheless, for the regions of complicated edges, such as trees, railings and still car, human observers often skip their detail. In our scheme, these regions may be considered unstructured region, and their visual quality may be decrease to save bits.

Table 5.5 shows the ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of grass and railings are smaller than others.

	25th frame		92nd frame	
PSNR	Better	Worse	Better	Worse
Trees	38	37	48	70
Railings	72	82	87	67
Sculpture	12	7	-	-
Passerby	6	1	-	-
Photo and word	-	-	13	4
Edge of Sky and Trees	-	-	6	2

Table 5.5 Ratio of improvement in in Bus sequence

5.2. Result of Proposed Bit Allocation Scheme with Weighting Threshold

In this section, we will show the results of our proposed bit allocation scheme with automatic CV threshold selection described in section 4.5. Section 5.2.1 will compare results in Method I and Method II and section 5.2.2 will show the detail result of the better Method of two.

5.2.1. Number of structured and unstructured blocks of the two method

Table 5.6 and Table 5.7 show quantity of blocks (%) indicated as SR and USR by Method I and Method II respectively. There are three input sequences with CIF resolution which are Stefan, Football and Bus sequences. There are two types of threshold source. Type “EACH” means that the threshold are calculated by each residual frame, and the type “GOP” means that the threshold are calculated by all residual frames in each GOP. T_{SR} is the CV value of lower bound of SR, and T_{USR} is the CV value of upper bound of USR. Blocks$_{SR}$ is the number of blocks indicated as SR, and Blocks$_{USR}$ is the number of blocks indicated as USR. Err$_{SR}$ is the error range of Blocks$_{SR}$, and Err$_{USR}$ is the error range of Blocks$_{USR}$. For example, the number of SR blocks is 5.42%+-7.71% among all blocks.

Average							
Sequence type	Threshold source	Method I					
		T_{SR} (CV)	T_{USR} (CV)	Blocks$_{SR}$ (%)	Blocks$_{USR}$ (%)	Err$_{SR}$ (%)	Err$_{USR}$ (%)
Stefan	EACH	0.47	0.11	5.42	4.68	7.71	12.74
	GOP	0.54	0.09	2.61	2.04	2.41	2.51
Football	EACH	0.48	0.16	5.31	10.55	26.26	16.47
	GOP	0.55	0.12	4.35	3.65	12.82	3.96
Bus	EACH	0.48	0.16	5.31	10.55	5.55	13.96
	GOP	0.62	0.08	1.56	1.66	3.62	1.09

Table 5.6 Automatic selection of threshold using method I.

Average							
Sequence type	Threshold source	Method II					
		T_{SR} (CV)	T_{USR} (CV)	$Blocks_{SR}$ (%)	$Blocks_{USR}$ (%)	Err_{SR} (%)	Err_{USR} (%)
Stefan	EACH	0.28	0.25	36.04	52.92	13.06	15.01
	GOP	0.27	0.23	34.15	52.07	22.61	21.10
Football	EACH	0.34	0.29	28.69	51.53	26.92	40.14
	GOP	0.34	0.36	39.07	63.75	57.99	60.31
Bus	EACH	0.27	0.18	27.71	37.42	10.29	9.30
	GOP	0.27	0.18	31.89	33.58	8.89	10.47

Table 5.7 Automatic selection of threshold using method II.

For Method I, number of blocks indicated as SR or USR is sparse and is below 11 %. However, for Method II, number of blocks indicated as SR or USR is larger than two times of Method I. For Method I, the error range of GOP is smaller than EACH because composition of CV in each frame is quite different. For Method II, the error range of EACH is smaller than GOP because selection by quantity can be more accurate. Nevertheless, in each frame, distinguishing of SR and USR is better in the method of threshold calculated by all residual frames in each GOP whatever in Method I or Method II.

5.2.2. The result of proposed scheme with linear formula

Table 5.8 and Table 5.9 show the results of testing the algorithm on three sequences using Method I and Method II, respectively. The test conditions and the meaning of the fields in the table are the same as those in section 5.1 (in particular Table 5.2).

Sequence	Stefan		Football		Bus	
Type	Original	Modified	Original	Modified	Original	Modified
PSNR	29.52022	29.52101	33.48467	33.49391	31.14866	31.14505
SSIM	0.91944	0.91938	0.87353	0.87397	0.89351	0.89323
QP	14	13~15	10	9~11	10	9~11
Total Bitrate	828	837	1177	1187	1279	1287
Luma Bitrate	655	655	841	841	1103	1103
Header Bitrate	133	141	245	256	150	157
Chroma Bitrate	39	39	89	89	25	25

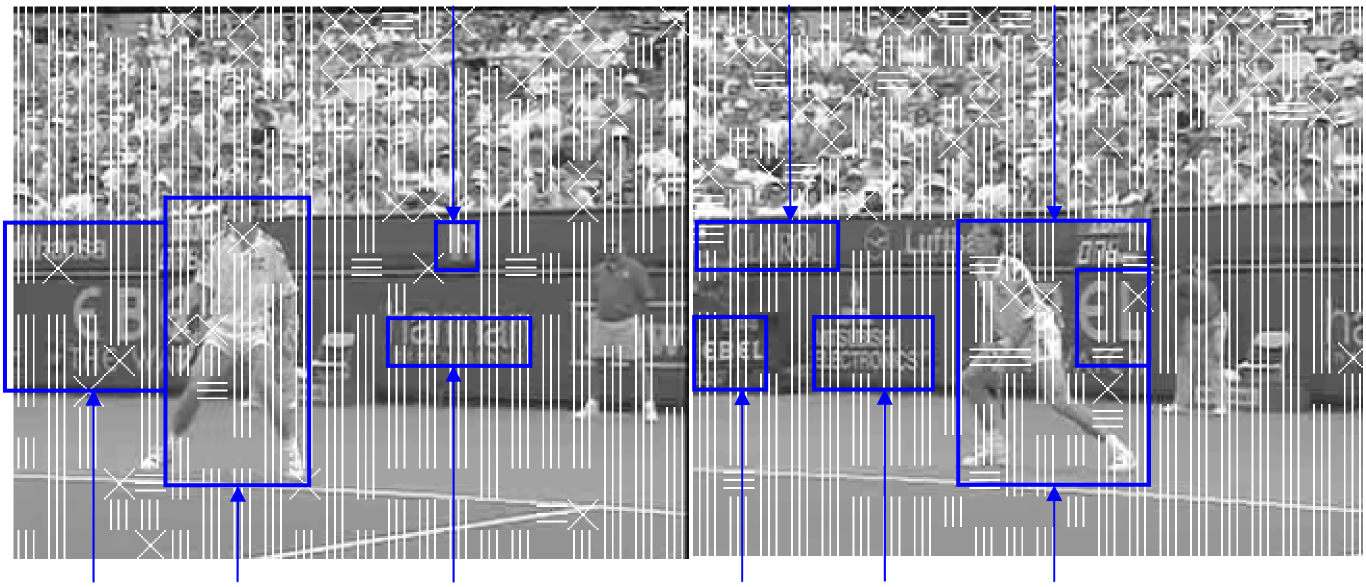
Table 5.8 Average result of Method I

Sequence	Stefan		Football		Bus	
Type	Original	Modified	Original	Modified	Original	Modified
PSNR	29.52022	29.50551	33.48467	33.46981	31.14866	31.1424
SSIM	0.91944	0.9191	0.87353	0.87381	0.89351	0.89236
QP	14	13~15	10	9~11	10	9~11
Total Bitrate	828	852	1177	1194	1279	1291
Luma Bitrate	655	655	841	840	1103	1103
Header Bitrate	133	156	245	263	150	161
Chroma Bitrate	39	39	89	89	25	25

Table 5.9 Average result of Method II

Method I perform slightly better on distinguishing between SR and USR than Method II. Moreover, on average, Method I is slightly better than Method II on video quality too.

Table 5.8 shows the result of a single frame by Method I. For each sequence we list two kinds of typical frames to analyze the result of visual quality and bits our proposed scheme causes. Figure 5.4, Figure 5.5 and Figure 5.6 show the result of visual quality in three sequences respectively. Symbols in these figures are as figures in section 5.1 defined.



(a) (b)

Figure 5.4 Comparison of visual quality in Stefan Sequence
(a)The 43rd frame in Stefan. (b) The 80th frame in Stefan.

	43rd frame		80th frame	
PSNR	Better	Worse	Better	Worse
Audience	99	33	87	45
Words	28	6	16	1
Whole body	28	8	18	6

Table 5.10 Ratio of number of regions with better PSNR and worse PSNR in Stefan sequence

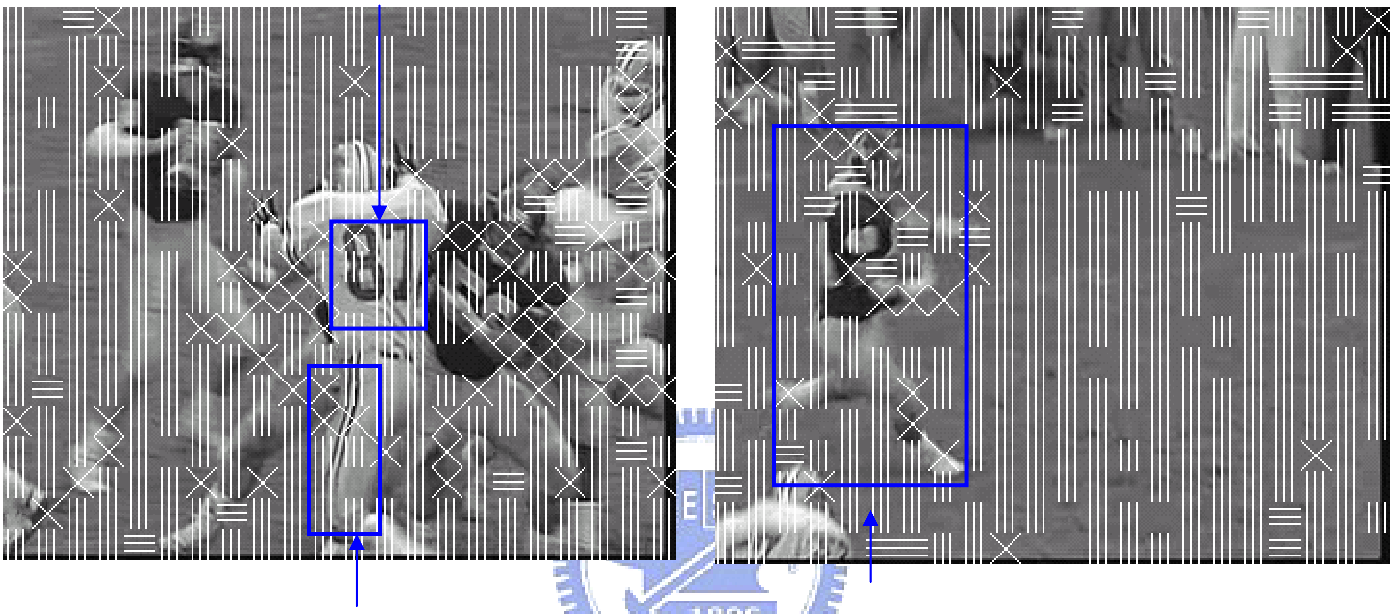
For the Stefan sequence, human observers may pay special attention to tennis player and the area with obvious edges such as words on the wall. On the other hand, the regions that audiences on the grandstand and the flat regions are mostly human observers are not sensitive to relatively. In Figure 5.4(a), the major movement in the 43rd frame is the tennis player moving towards right hand side. In our proposed scheme, the regions with clear and obvious directional edges will be considered structured region. As a result, performances of this kind of regions such as tennis player's legs, words on the wall and lines on the ground are mostly enhanced. Nevertheless, the regions of audiences on the grandstand and the flat regions will be seemed to unstructured regions since their directions of edges are complicated. So bits of these regions are usually decreased and it may cause worse visual quality.

In Figure 5.4(b), the 80th frame, the major movement in the 80th frame is the tennis player waving his rocket. Therefore, human eyes may notice the area of tennis player's whole body and the area with obvious edges such as words on the wall and lines on the ground. Performances of these kinds of regions are mostly enhanced. And similar as in Figure 5.4(a), bits of the regions of audiences on the grandstand and the flat regions are usually decreased and it may cause worse visual quality. Table 5.10 shows ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of audience is smaller than others. Table 5.11 shows the saved bits of unstructured regions by our model.

Saved bits in unstructured regions		
Saved bits	43rd	80th
Audience	92	105

Table 5.11 Saved bits in unstructured regions in Stefan sequence

Next, in whole Football sequence, human observers may pay more attention on the area of football and football player than the area of grass. Moreover, obvious edges on football player such as numbers on their sports coats or stripes on their pants may attract human eyes dramatically. In Figure 5.5(a), the major movement in the 65th frame is the football players competing for the football. In the proposed scheme, the performances of the regions we said above that humans may be more sensitive to, numbers on their sports coats or stripes on their pants, are mostly enhanced. Nevertheless, bits which are allocated to the regions of too complicated grass and the flat regions are usually decreased because these regions may be seemed to unstructured regions. And the processing may cause worse visual quality of these regions. Here we select the other kind of frame in Football sequence to analyze its result. In Figure 5.5(b), the major movement in the 120th frame is the football players running towards the right with the football in his hand. In this frame, human may pay attention to the only football player and the football. In our proposed scheme, the performances of the regions we said above mostly enhanced. However, for less important regions, such as the regions of too complicated grass and flat regions, their bits are usually decreased and their visual quality may be reduced.



(a)

(b)

Figure 5.5 Comparison of visual quality in Football Sequence

(a)The 65th frame in Football. (b) The 120th frame in Football.

Table 5.12 shows ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of grass is smaller than others. Table 5.13 shows the saved bits of unstructured regions by our model.

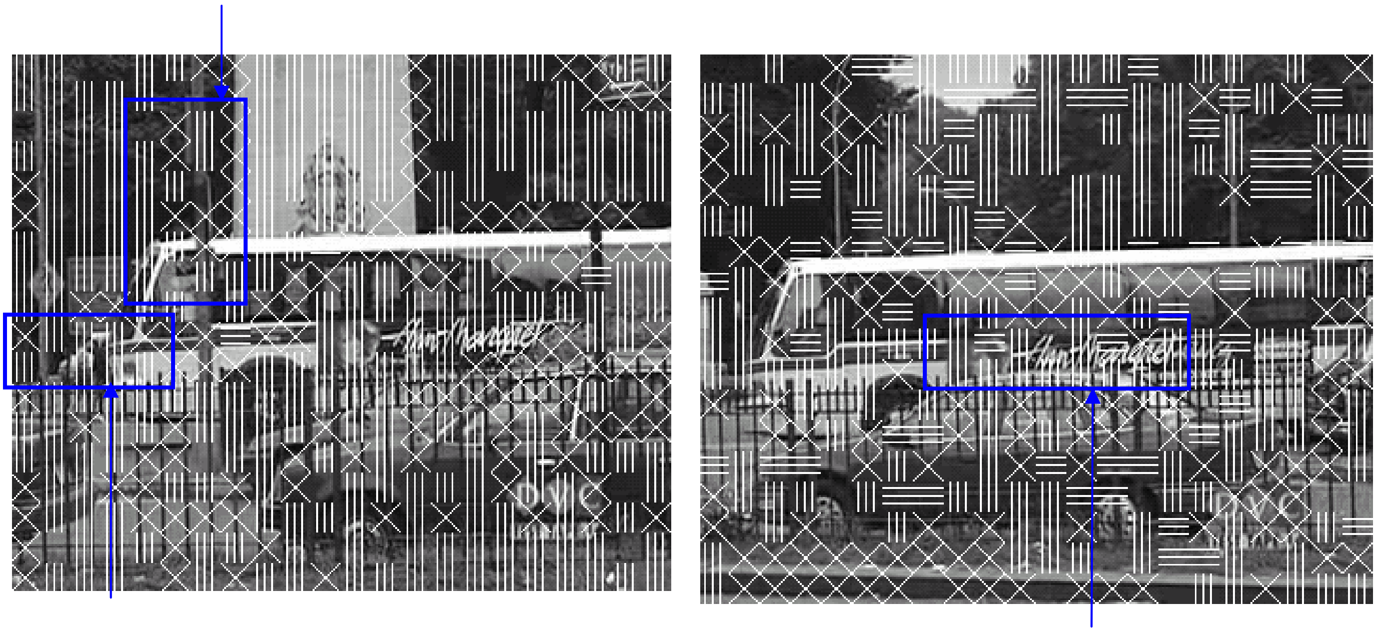
	65th frame		120th frame	
PSNR	Better	Worse	Better	Worse
Grass	79	16	214	21
Numbers	14	1	-	-
Stripe	13	1	-	-
Whole body	-	-	31	10

Table 5.12 Ratio of number of regions with better PSNR and worse PSNR in Football sequence

Saved bits in unstructured regions		
Saved bits	65th	120th
Grass	101	66

Table 5.13 Saved bits in unstructured regions in Football sequence

Last, in Bus sequence, human observers may not pay more attention on the area of complicated background such as the trees on the top of image, and complicated foreground such as railings and still car. On the other hand, the moving bus and various backgrounds are more attracted to human eyes than the region we said above generally. Here we select two different kinds of scenes of bus sequence to analyze our result. In Figure 5.6(a), the bus is just passing through the pillar with sculpture. Therefore, human observers may take their on the regions of the sculpture, human under the sculpture, the head of bus and the top of bus. In our proposed scheme, visual qualities of these regions are mostly enhanced. In the 98th frame, as Figure 5.6(b) shows, the regions that human observer may notice a lot are listed as follows: the advertisement with photograph and words on the bus, the street light near the head of bus and the region that sky and trees are associated with. Visual qualities of these regions are mostly enhanced. Nevertheless, for the regions of complicated edges, such as trees, railings and still car, human observers often skip their detail. In our scheme, these regions may be considered unstructured region, and their visual quality may be decrease to save bits.



(a) (b)

Figure 5.6 Comparison of visual quality in Bus Sequence

(a)The 47th frame in Bus. (b) The 98th frame in Bus.

	47th frame		98th frame	
PSNR	Better	Worse	Better	Worse
Trees	64	32	80	55
Railings	88	70	62	70
Sculpture	9	1	-	-
Passerby	2	0	-	-
Photo and Word	8	2	13	2

Table 5.14 Ratio of number of regions with better PSNR and worse PSNR in Stefan sequence

Saved bits in unstructured regions		
Saved bits	47th	98th
Tree	92	61
Railings	70	787

Table 5.15 Saved bits in unstructured regions in Bus sequence

Table 5.14 shows ratio of number of regions with better PSNR and worse PSNR, and it is obvious that the ratio of regions of grass and railings are smaller than others. Table 5.15 shows the saved bits of unstructured regions by our model.

6. Conclusion and Future Work

In this thesis, we proposed a video coder bit allocation scheme in Curvelet domain. A new transform, curvelet transform, which contains the property of multi-resolution and multi-directional decomposition, is introduced into the proposed bit allocation algorithm. The Otsu threshold selection algorithm is used to pick the principal edge directions in image regions. And then, coefficient of variation (CV) is used to measure the complexity of image region to determining the quantization parameters for video coding.

To be more specific, the proposed scheme classifies all macro blocks into three groups with three different quantization parameters. The first group of regions is the normally structured regions whose texture (or motion-compensated residual) is neither complicated nor simple so we do not change its allocated bits. The second group of regions is composed of unstructured regions. This type of region means that the texture is either too simple or too complicated. Therefore, we can decrease its bits and the compression result does not cause obvious distortion to human eyes. In addition, we can allocate the saved bits to the regions that human observers are more sensitive to. The third group of regions is composed of well structured regions whose texture (or motion) is clear and easily recognizable so the proposed scheme increase its bits to enhance the quality since the improvement of visually quality in this kind of region is obvious to human eyes.

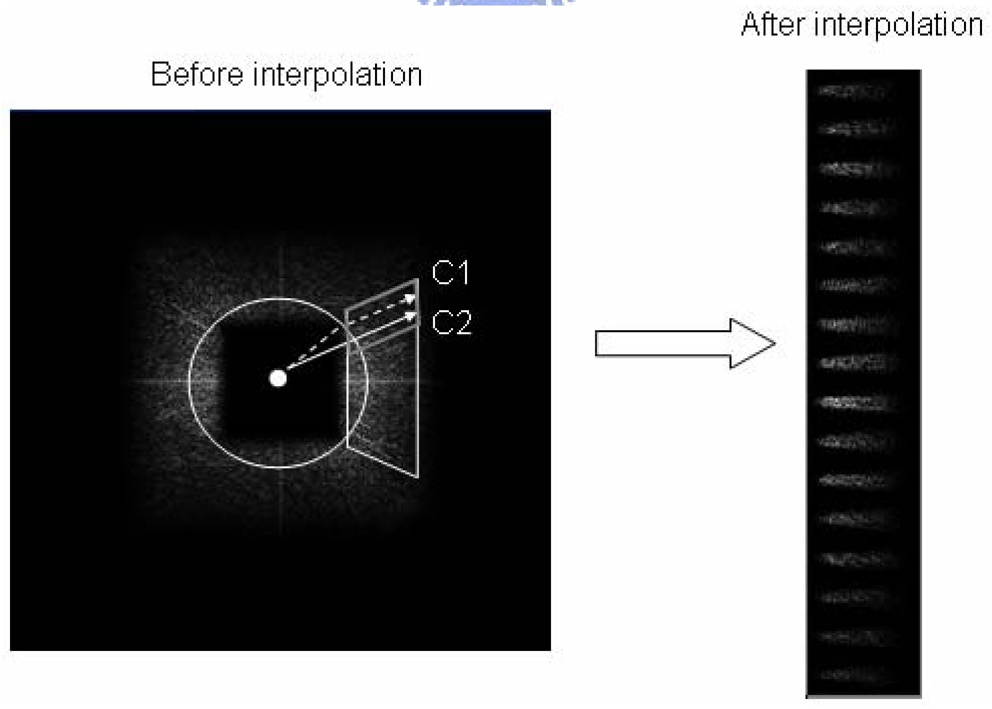
The coding performance of the proposed method is compared with the MPEG-4 simple profile encoder. Experiments show that the result of our directional complexity analysis can distinguish the groups of structured and unstructured area for all the test sequences used. The proposed method has the better performance with higher PSNR numbers in regions that human observers are more sensitive to. Even for the average result of PSNR and SSIM, our method can obtain slightly better performance given the same or lower luma bitrate.

Although the proposed bit allocation algorithm performs well, there are still some improvements that can be expected. For example, the proposed automatic threshold selection algorithm requires two pass encoding, which may not be desirable in some cases.

Secondly, some regions which are structured regions or even strictly structured regions are on the position that human observers don't care about. For example, the regions of audiences in the Stefan sequence is the typical regions that human eyes may not pay attention to. Therefore, even some of the regions are well structured in this area, it may not make sense to allocate more bits to them.

Another drawback is about the directional decomposition procedure of curvelet transform. As section 3.4.4 describes, the directional decomposition is processed by polar interpolation. However, for each angular wedge, the direction of coefficients they collect is not so accurate. Figure 6.1 shows the coefficients before and after the procedure of the directional decomposition.

Figure 6.1 The directional decomposition in the 4th level coefficients.



For instance, the direction of interpolation inside the first angular wedge is always along the direction of arrow C2. However, coefficients on the trajectory the arrow C1 do not represent the coefficients of this angle. Therefore, coefficients introduced by the polar interpolation method are not accurate enough. Consequently, the directional information we obtain is not accurate according to the influence. If the directional decomposition algorithm can be improved, the edge distribution analysis of the proposed bit allocation algorithm can be more precise and it will improve the performance of the proposed scheme too.

In summary, future improvements can be expected with these efforts.

7. Reference

[1] N.Ahmed, T.Natrajan and K.R.Rao, “Discrete cosine transform,”,*IEEE trans. Computers,* January 1974.

[2] Rao K R,Yip P,” Discrete cosine transform: algorithms, advantages, applications,” *[M].New York: Academic Press*,1990

[3] Ze-Nian Li, Mark S.Drew, ” Fundamentals of Multimedia,” 2004

[4] J.F.Blinn,” What's the Deal with the DCT?” *IEEE Computer Graphics and Spplications,*13(4):78-83,1993

[5] Rafael C. Gonzalez, Richard E. Woods, ”Digital Image Processing 2/e,” 2001

[6] Alan V. Oppenheim, Ronald W. Schafer,” Discrete-time signal processing,” 1999

[7] S. Mallat, “A Compact Multiresolution Representation: The Wavelet Model,” *Proc. IEEE Computer Society Workshop on Computer Vision, IEEE Compuiter Society Press,* Washington, D.C., pp. 2-7, 1987

[8] Daubechies I, “Orthonormal bases of compactly supported wavelets,” *Commun Pure and Appl Math,* 41(7): 909～996, 1988

[9] S. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. PAME-11,pp. 674-693, 1989

[10] Minh N. Do and Martin Vetterli, “The Contourlet Transform: An Efficient Directional Multiresolution Image Representation,” *Fellow,* IEEE, 2005

[11] P. Abrial (1,2), J-L. Starck(1), Y. Moudden(1) and M. Nguyen(2),” CURVELET TRANSFORM ON THE SPHERE,” 2005

[12] Andrei C. Jalba, Michael H.F. Wilkinson, Member, IEEE, and Jos B. M. Roerdink, “Shape Representation and Recognition Through Morphological Curvature Scale Spaces,” *Member,* IEEE, 2006

[13] Myungjin Choi, Rae Young Kim, Myeong-Ryong Nam, and Hong Oh Kim,“ Fusion of Multispectral and Panchromatic Satellite Images Using the curvelet transform,“ 2005

[14] J. L. Mannos and D. J. Sakrison, “The effects of a Visual Fidelity Criterion on the Encoding of Images,” IEEE Trans. on Information Theory, Vol.20, No. 4, Jul. 1974, pp.525-536.

[15] P. G. J. Barten,” Contrast Sensitivity of The Human Eye and Its Effects on Image Quality,“ *SPIE-International Society for Optical Engineering,* 1999.

[16] A. N. Netravali and B. G. Haskell, “Digital Pictures: Representation and Compression” *New York, NY: Plenum,* 1988.

[17] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, “Topdown control of visual attention in object detection,” in *Proc. ICIP,* vol. 1, Sep. 2003, pp. 253–256.

[18] Yarbus, A. L.,” Eye Movements and Vision,” New York: Plenum Press, 1967

[19] P. Reinagel and A. M. Zador, “Natural scene statistics at the center of gaze,” Network: Comput. Neural Syst., vol. 10, no. 1–10, 1999.

[20] V. Bhaskaran and K. Konstantinides, “Image and Video Compression Standards: Algorithms and Architectures, 2nd ed.,” Boston: Kluwer Academic Publishers, 1997

[21] E. Cand`es, L.Demanet, David, Donoho, “Fast Discrete curvelet transforms, ” Technical Report, available at http://www.curvelet.org , 2006

[22] E. Cand`es, David L. Donoho,” Curvelets: a surprisingly effective nonadaptive representation for objects with edges,” in *Curves and Surfaces IV ed. P.-J. Laurent.,*1999

[23] E. Cand`es, David L. Donoho, “New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities,” *Comm. Pure Appl. Math.* 57, pp. 219-266, 2004

[24] Minh N. Do, “Directional Multiresolution Image Representation,” Oct. 23, 2001

[25] E. Cand`es and F. Guo.,” New multiscale transforms, minimum total variation synthesis: application to edge-preserving image reconstruction,” *Sig. Process., special issue on Image and Video Coding Beyond Standards 82,* 1519–1543, 2002

[26] D. L. Donoho and M. R. Duncan.,” Digital curvelet transform: Strategy, Implementation,

Experiments," Technical Report, Stanford University, 1999

[27] J.L. Starck, E. J. Candès, " The Curvelet transform for image denoising," *IEEE Trans. Image Proc.*, 11(6): 670-684, 2002

[28] Matlab 7.0.1, The MathWorks, Inc.

[29] FFTW version 3.1.2, available at http://www.fftw.org , 2006

[30] S. Mallat, "A wavelet tour of signal processing (2nd Edition), " *Academic Press Inc.,* San Diego, CA, 1998

[31] CK. Yang, W.H. Tsai," Reduction of color space dimensionality by moment-preserving thresholding and its application for edge detection in color image," May. 17, 1995

[32] Nobuyuki Otsu," A Threshold Section Method from Gray-Level histogram," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. SMC-9, No. 1 JAN 1979

[33] K. Fukunage," Introduction to Statistical Pattern Recognition," *New York, Academic,* 1972, pp.260-267

[34] Frank, H. and Althoen, S.C," The coefficient of variation," *§C.4.b in Statistics: Concepts and Applications Cambridge, Great Britain: Cambridge University Press,* 1995, pp. 58-59

[35] C.-W.Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual Sensitivity-Guided Bit Allocation for Video Coding,"*IEEE Trans. on Multimedia*, EI, SCI, Vol. 8, Issue 1, pp. 11-18. (NSC #92-2219-E-009-006), 2006

[36] W.-C. Chang," *Human Visual System Based Bit Allocation for Video Coding,*" 2006