THE HONG KONG
POLYTECHNIC UNIVERSITY

Department
of
Electronic and Information Engineering

Unified feature analysis in multiple compressed domains

by

AU Ka Man

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy

November 2004

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written nor material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

________________________________ (Signed)

________________________________ (Name of student)

Abstract

Visual information often requires huge storage space. Owing to the constraints of transmission and storage, images are usually compressed by either JPEG or JPEG2000 to reduce its storage requirement. Over the past decade, retrieval systems operated in uncompressed/spatial and compressed domains have been proposed. However, they all function in single domain only. Retrieving these kinds of images in multiple domains typically requires reconverting them into spatial domain in which features are extracted for further analysis. This approach incurs many pre-processing operations such as decompression, especially for large image archives. The objective of this study is to investigate the common features in different compressed domains so that image indexing can be done directly from their respective domains.

A fundamental difference between JPEG and JPEG2000 is their transformation schemes. JPEG and JPEG2000 employ dissimilar Block-based Discrete Cosine Transform (BDCT) and Wavelet Transform (WT) respectively. Direct comparison on BDCT blocks and WT subbands cannot expose their relationship. By employing a subband filtering model, filters in BDCT and WT can be directly compared. In accordance with our intensive mathematical analysis, BDCT coefficients can be concatenated to form structures similar to WT subband. Under the same structure, BDCT and WT filters are comparable. Considering JPEG2000 Part I and II compression schemes, commonly used wavelet kernels are involved in our comparison. Theoretical studies show that both BDCT and WT filters share common characteristics in terms of passband region, magnitude and energy spectra. Particularly, lowpass filters in the two transforms are the same for Haar wavelet. In addition, both lowpass and bandpass filters of the selected kernels provide high similarities. Outputs of the two subband models are alike. Common features can thus be extracted from the BDCT and WT subband outputs.

Though high similarities are confirmed between BDCT and WT outputs, compression may affect their similarities. This is because the compression schemes of JPEG and JPEG2000 are quite different starting from transformation to quantization. The effect of compression on their similarities is worth examining. A variety of images are compressed at various compression ratios ranging from 1.6:1 to 72:1 by BDCT in JPEG and different wavelet kernels in JPEG2000. Studies on their output spectra expose their similarities under compression. Despite of high compression, large similarities can still be found.

To validate our proposed subband filtering model, an image retrieval system is established. The system aims to search for images in different compressed domains by applying the model to partially decompressed images. To overcome the effect of shifting, scaling and rotation, translation and rotation invariant features are extracted from the images. Our simulation results present high precision and recall values at all compression ratios. Our proposed indexing algorithm concludes that relevant images can be searched from different compressed domains, regardless of the wavelet kernel or compression ratio used.

Both theoretical and experimental studies confirm that common features can be extracted directly from multiple compressed domains irrespective of the value of the compression ratio and the use of BDCT and WT kernels. Relevant JPEG and JPEG2000 images can be retrieved from one and the other without incurring full decompression.

Acknowledgments

I extend my most sincere gratitude to my Supervisor and my best friend, Dr. Bonnie N.F. Law, for her continuous guidance and encouragement. Her down-to-earth research approach has been a significant influence on me. She gave me informative suggestions and valuable advises concerning my studies. Her effort has been indispensable and contributed to every success of my research. Deserving of special thanks is my Co-Supervisor Prof. W.C. Siu, Director of the Center for Multimedia Signal Processing, for his support and invaluable advice. I am deeply impressed with his subjective judgment on research, science and global direction. It is beyond doubt that this will positively influence my future research and career.

I particularly thank Dr. Kenneth Lam, Dr. Daniel Lun, Dr. Y.L. Chan, Mr. Patrick Chan, Mr. K.O. Cheng, Miss K.L. Lau, Mr. W.L. Hui, Mr. H.K. Cheung, Mr. K.T. Fung, Mr. K.C. Hui, Mr. W.H. Wong, Mr. T.P. Yip, Mr. Rainbow Fu, Mr. K.H. Lin and Mr. C.M. Lai. Their expert knowledge and outstanding support have helped me in solving many difficulties. I am deeply grateful to my best friends, Mr. Alvin Cheung, Mr. Eddy Lo, Mr. Jim Tse, Mr. K.Y. Cheong, Miss Karen Leung and Mr. Tommy Lau. Without their constant encouragement, my work would not have been completed so smoothly.

My gratitude goes to the Department of Electronic and Information Engineer, the Center for Multimedia Signal Processing, and the Hong Kong Polytechnic University, for providing me a comfortable working environment and generous financial support.

Finally, and above all, I acknowledge the love, patience and encouragement of my family.

Table of Contents

List of Figures

List of Tables

Abbreviations

1D	One dimensional
2D	Two dimensional
35D	35 dimensional
BDCT	Block-based Discrete Cosine Transform
B93	Biorthogonal 93
B97	Biorthogonal 97
DB4	Daubechies 4
DCT	Discrete Cosine Transform
DVD	Digital Video Disc
FCH	Fuzzy Color Histogram
MSB	Most Significant Bit
JPEG	Joint Photographic Experts Group
LSB	Less Significant Bit
QBIC	Query By Image Content
RCT	Reversible Component Transform
TIFF	Tagged Image File Format
VCD	Video Compact Disc
WT	Wavelet Transform
WWW	World Wide Web

List of Publications

International Journal Papers

1. K.M. Au, N.F. Law and W.C. Siu, “Unified Feature Analysis in JPEG and JPEG2000 Compressed Domains“, submitted to Pattern Recognition

International Conference Papers

1. K.M. Au, N.F. Law and W.C. Siu, “Unified feature analysis in different compressed domains,” *2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia*, Vol. 1, pp. 71-75, 15-18 Dec. 2003

2. K.M. Au, N.F. Law and W.C. Siu, “Spatial-spectral feature analysis in JPEG and JPEG-2000,” *International Symposium on Intelligent Multimedia, Video and Speech Processing*, Vol. 1, pp. 378-381, 20-22 Oct. 2004

3. K.M. Au, N.F. Law and W.C. Siu, “Direct image retrieval in JPEG and JPEG2000”, *IEEE International Conference on Image Processing 2005*, 11-14 Sept, 2005

Chapter 1

Introduction

1.1 Overview

Visual information has increased enormously in our daily life. Pictures are taken by digital devices such as Digital Camera and stored in computers in digital format. Movies can be captured by Digital Video Camera and stored in Digital Video Disc (DVD) or Video Compact Disc (VCD). As a result of integration of the visual information into multimedia systems, the World Wide Web (WWW) has become very popular [23]. Owing to the enormous amount of information to be stored, compression is one of the ways to reduce the storage requirement of large image archives. Many compression schemes have been proposed to reduce storage size at the expense of visual quality. Compression also provides a solution for transmitting a large amount of information under the constraints of limited transmission bandwidth. If one wants to find a relevant image from a multimedia storage system, an effective image searching system is extremely important in this case. The face recognition system for security purposes is one of these useful applications. However, most of the systems have to store millions of images. This results in a large storage requirement. To reduce this requirement, images in these archives are always compressed. When images are in a compressed form, an effective image indexing system in the compressed domains is desirable.

In this chapter, I would like to explain what digital visual information is and why we have digital image processing. As images may be compressed effectively by exploring their characteristics during the compression process, image characteristics will be discussed. Then, an overview of some existing compression schemes will be given. Also, differences between different compression schemes will be summarized. Considering that image retrieval is one of the important applications of image processing, image retrieval systems will be further explained. Finally, the objective and the organization of this thesis will be presented.

1.2 Digital Image Processing

It is generally accepted that pictures appear on surface permanently. People show their wonderful painting skills by drawing pictures on papers or walls. Photographers capture beautiful scenes by traditional film cameras. As a result of advances in new technologies nowadays, pictures can be drawn using computers. Photographs can be taken by digital cameras as well as digital video cameras. Even the hand-drawn pictures can be stored in computers by using a scanner to make a copy of it and save the copy in the computer. These pictures, namely digital images, are stored in digital format by the use of new technologies. With the help of digital image processing, a digital image can be easily modified or edited without doing the whole picture. Image enhancement, restoration and object segmentation are the common applications of digital image processing. Using simple software tools, visual quality can be easily improved by changing the color contrast or image brightness. Image restoration allows us to reduce the distortion of images, such as the distortions due to noise, artifact and blurring. Digital image processing provides the convenience to produce high quality pictures.

1.3 Image compression standards

Due to the advantages of the digital medium, there is a large amount of digital images in multimedia systems. Imagine that when we have many archives of digital images, a large storage space is required. However, every piece of storage space means cost, especially to a large image archive. To reduce the cost of storage, one of the solutions is to compress the images so that less storage size is needed. Many image compression schemes have been investigated and proposed in the past decades. Two examples are Joint Photographic Experts Group (JPEG) and Tagged Image File Format (TIFF) [20,53,36,37]. The aim of image compression is to use less storage space to store the original images. This is possible because part of the visual information in pictures is redundant. Thus, compression schemes preserve the non-redundant information or the characteristics of images but neglect the unimportant information. JPEG image compression scheme is a commonly used image compression standard, especially in WWW application. JPEG2000 is another newly developed image compression standard [14], which provides better image quality and stronger compression power than JPEG. The choice of using which compression algorithm depends on users' applications, or even their preference. Therefore, users may use various compression standards to compress their visual data. Inside one database or storage device, images are highly possible to be stored in different compressed domains.

During the compression process, images in the spatial domain are transformed into the spatial-frequency domains. The advantage of this transformation is that variation of image contents are clearly shown in the frequency spectrum in the frequency domain. To store the contents of an image, we can store them in the frequency domain instead of the spatial domain. Despite the use of either lossless or lossy compression schemes, less storage space is required to store the image in the compressed domains than in the original spatial domains. This is because image features are transformed into frequency domain where compression undergoes. Quantization and entropy coding on the frequency coefficients aim to use less bit-streams to encode the correlation between the coefficients. Thus, the whole transformed image is represented by a minimized length of bit-stream at a specific compression ratio. The shorten bit-stream then reflects the reduced storage size after compression.

1.4 Content based image retrieval

With the rapid growth of the Internet and multimedia systems, the use of visual information has increased enormously. Lots of visual information can be found in the Internet or any multimedia systems. In many applications, users would like to find the relevant visual information from the systems. Thus, finding out relevant visual information from these systems efficiently becomes a valuable research topic in recent years. Text-based image retrieval methods are used in the early indexing system [59,21]. Images are first annotated with text and keywords. Image searching is simply accompanied by finding out similar image description and keywords. However, numerous keywords can be used as the description of the same object. As the keywords to be used differ from person to person, the indexing results are not reliable. Therefore, using the content of the images instead of keywords in indexing can be more reliable.

Over the past decade, many retrieval systems based on image contents have been proposed, such as the QBIC system [43] from IBM, the Virage system [1] by Virage and the Photobook system [2] from the MIT Media Lab. These indexing systems find the features (meaningful information) inside the image contents directly from the spatial domain. Thus, visual contents of images such as color, shape, texture and spatial layout are classified as image features. By comparing these features between different images, images with similar visual contents can be searched. These retrieval systems provide more reliable indexing results than text-based image retrieval methods, but cannot be applied to images which are in the compressed domain. This is due to the fact that features of images in the compressed domains are in the frequency domain rather than in the spatial domain. Therefore, retrieval systems operate in the spatial domain cannot extract features from the compressed images. To search compressed images in multimedia systems, image indexing in the compressed domains is desirable.

1.5 Image retrieval in compressed domains

As we have discussed in section 1.3, a large amount of visual information is in compressed domain due to the benefit of storage reduction. Image retrieval systems in spatial domain that are introduced in section 1.4 are no more applicable. To search relevant visual data in compressed domains, one can fully decompress all the compressed data back to their original spatial domain and then find out those images with similar visual contents. However, this process requires huge computational resources and thus will be extremely time consuming for a large image archive. As many applications are image searching from the WWW or a large image database, it is very inefficient to entirely decompress all the images in the database. Therefore, compressed image retrieval systems must also operate in the compressed domains.

Numerous image indexing systems in the compressed domains have been proposed, for example, the WaveGuide [42,26,65]. These retrieval systems aim to find image features in the compressed domains such as frequency contents of the image and preserved energy after compression. One major constraint of these retrieval systems is that they are domain specific. This means that they are designed for, and can only retrieve relevant images from, a particular compressed domain. As images can be compressed in various types, retrieval systems operate in single domain may lose similar images in other compressed domains. To perform image indexing in different compressed domains, a compression independent method is needed.

1.6 Feature extraction in different domains

As mentioned in the previous sections, the major functionalities of image indexing systems are to compare the features of images and search images carrying the most similar characteristics. In the spatial domain, indexing systems mainly extract characteristics directly from the image pixel intensity. Therefore, characteristics like shape, color, statistical analysis can be obtained easily based on the intensity level. Different from the spatial domain, images in the compressed domains are transformed from the spatial domain. Therefore, images are converted from spatial intensity level into the frequency spectra. By analyzing the frequency spectra, characteristics preserved by the transformation can be found. Therefore, different transformation schemes may preserve different information in the frequency spectra.

As mentioned in subsection 1.3, there are a number of image compression schemes. In practice, images are possibly compressed by different compression schemes in the multimedia systems. Since different compression schemes employ different transformation kernels during the compression process. Therefore, information in the transformed domains that are considered as important characteristics may differ from one compression scheme to another. As different compression schemes identify different sets of characteristics, the retrieval systems suitable for every compression scheme are quite different. This leads to a situation that image retrieval systems are specially developed for a particular compression scheme. For example, JPEG and JPEG2000 use Discrete Cosine Transform (DCT) and Wavelet Transform (WT) respectively as their transformation kernels. The two different transformation kernels result in different information preserved. Therefore, special retrieval systems are required to deal with dissimilar features preserved by the compression schemes [9,8,56,17,13,22,73,40,62,52].

Inside a real multimedia system, images with different image formats may exist in the same image archive. To search images with similar contents in different domains, the traditional image indexing systems are ineffective. However, though different transformation kernels are used in different compression schemes, they aim to transform images from the spatial domain to frequency domain. Thus, some compression schemes may preserve similar features of the images in the compressed domains. In order to search images in such database, we need a retrieval system that can extract common features in different compressed domains. Until now, very few studies have been done in investigating such kind of systems. To fill in this gap, our research aims to investigate how to find the common features between different compression schemes to realize image indexing in different compressed domains. And we aim to establish an image retrieval system that allows image searching in different compressed domains.

1.7 Objective of this thesis

Images can be heterogeneous in uncompressed or compressed domains. With the rapid growth of multimedia systems, there are wide applications of visual information in multimedia systems, for example, in digital library systems, and video on demand applications, etc. To search for relevant visual data from such large multimedia systems, image indexing and retrieval techniques have become important in modern applications. Many retrieval systems have been proposed recently. However, most of them operate in the single image nature. In other words, they can only search images of single type, but are unable to find out others. Even though there are retrieval systems that operate in different image domains, they can only search images in one uncompressed and one compressed domains by converting them to the same domain. However, in modern image archives, images are compressed into various image natures. Thus, the traditional image indexing and retrieval systems are not efficient in searching similar images in heterogeneous domains. To fit the modern multimedia systems, a novel image indexing system will be proposed to cope with the indexing problem of heterogeneous image natures.

In this thesis, reviews will be made on various image natures and retrieval systems. As images may exist in different compressed domains, it would be time consuming to perform image indexing by decompressing all the information. To cope with this problem, an image indexing algorithm will be proposed and discussed in detail. In order to verify whether common features exist in different compressed domains, we will analyze their compression schemes. Then, simulation of indexing in different compressed domains will be carried out to validate our analysis.

1.8 Organization of the thesis

This thesis is organized as follows. Overviews on image natures, image compression standards and retrieval systems will be given in chapter 2. As JPEG is commonly used in image compression and JPEG2000 will be the future trend of compression standard due to its advantages, this thesis will focus on these two compression schemes. A fundamental difference between them is their transformation schemes. In chapter 3, we will analyze the spectral characteristics of Discrete Cosine Transform and Wavelet Transform which transform visual information from spatial domain to spatial-frequency domain in JPEG and JPEG2000 compression standards respectively. From their spectral characteristics, we can determine their similarities. Further analysis will be made in chapter 4 to find out the similarities between JPEG and JPEG2000 images. This reveals the degree of commonality between their compressed outputs. Based on their output similarities, common features in JPEG and JPEG2000 compressed domains will be extracted in chapter 5. Both mathematical studies and experiments will be carried out to find out the representative characteristics in these two domains. Simulation results of our proposed retrieval algorithm in different compressed domains will be produced to validate our theoretical studies. Discussion on our experimental results will be made in chapter 6. Moreover, we will evaluate the effectiveness of the feature extraction in JPEG and JPEG2000. Finally, conclusion of the studies will be drawn in chapter 7. To carry the studies further, we also describe some future development of feature extraction in different compressed domains.

Chapter 2

Reviews on image compression and retrieval systems

2.1 Introduction

As a result of advances in new technologies nowadays, visual information such as pictures on papers or any other surfaces can be stored in computers. The integration of visual information to multimedia systems become so popular that a large amount of information has to be stored digitally. However, due to the limitations on storage space and transmission bandwidth, compression is used to reduce the storage requirement as well as the bandwidth effectively. Many compression schemes have then been proposed. As the multimedia systems store a large amount of useful information, one may want to search valuable information from them. An effective search/retrieval system is important in this application. In fact, information may be compressed by different compression schemes in the multimedia systems. Retrieval system that allows retrieving information from different compressed domains is thus desirable.

In this chapter, a brief introduction on image components is firstly introduced. We will then briefly discuss some of the basic features such as mean and histogram of pixels in image description. In fact, compression algorithms make use of these features by preserving them during the compression process. Examples include the commonly used JPEG and the newly developed JPEG2000. Using effective compression schemes, images can be stored in multimedia system effectively. In some applications, users may want to search for relevant information from these systems. Many retrieval systems that work in the compressed domains have been proposed. Researches on feature extraction and image indexing in JPEG and JPEG2000 compressed domains will be explored. Finally, we will give a summary to conclude this chapter.

2.2 Image Components

Images are drawn dot-by-dot by their color components or grayscale intensity level. Each dot in any image is named pixel in the spatial domain. Any image analysis in the spatial domain is named pixel level analysis. The dimension of an image is defined by the number of pixels in its horizontal and vertical directions. For example, if there is an $N \times M$ pixels image, the width and the height of the image are N pixels and M pixels respectively. The total number of pixels in this example image is the product of N and M.

The visual information of each pixel in the image can be described by either the grayscale intensity level or color space components. From the image representation point of view, the main difference between gray and color images is their dimensional space. Color image is represented in three dimensional color space, which provides more visual information than single dimensional gray values. The grayscale image is the oldest and the most basic nature of images, and appears in, for example, monochrome film pictures. A high gray intensity level actually reveals a bright pixel. In contrast, a low intensity value shows a dark pixel [68].

In real-world images, most of them are color images as perceived by humans. Color images provide more information than monochrome images since they contain not only the intensity information but also color information. Therefore, color images are more attractive than monochrome images and are widely in existence in multimedia systems. A number of color spaces have been defined, for example, RGB, CMY, YIQ, YUV, CIE L*a*b*, CIE L*u*v*, HSV, HIS and HLS [10]. As these color spaces are designed for specific applications, there is no agreement on which color space is the best. RGB color space is commonly used in image representation. It contains three primary color components, red, green and blue. RGB are mainly used in color camera sensors. Though there are only three primary colors, any color can be produced by mixing the three components together. YIQ and YUV are two other color spaces originated from video broadcasting. Y is the luminance and I, Q, U and V are the chrominance components. Besides RGB, YIQ and YUV, there are many other color schemes [10]. These color spaces are designed according to the human perceptual system and are being used in some image encoding schemes.

2.3 Image Features

Images usually contain plain areas, texture patterns, objects, edges and curves [10]. It is well-known that images are non-stationary in nature. The non-stationary characteristic of images means that statistics is not constant all over an image. As an example, Figure 2.1 shows that the histogram of the Lena image varies from region to region. The non-constant histogram analysis thus demonstrates the non-stationary image nature. In order to characterize the non-stationary nature, both global and local features of an image are required to be analyzed separately, usually in spatial and frequency domains [48].

Statistical measurement at the pixel level reflects features in the spatial domain [70,18]. As shown in figure 2.1, the regional histogram differs from one region to another. The difference between histograms can be regarded as a feature. The statistical measurement should also differ from one image to another since image contents are different. Therefore, feature can be extracted either from the whole image or from any region in the image. This fact results in two kinds of visual content descriptors, i.e, global and local descriptors. A global descriptor extracts visual features from the whole image, whereas a local descriptor extracts features from regions or objects to describe the image. Commonly used statistical measurements such as mean, variance, standard deviation and histogram can capture the statistical characteristics at different regions or the whole image [61,25,47,18,29,57,64]. Statistical measurement can either be obtained from the luminance or chrominance components in an image.

Each piece of visual information contains a different frequency content. Plain areas and edges usually dominate at the lowpass and bandpass band respectively [48]. This is due to the fact that edges belong to sharp changes and discontinuities, which show a large magnitude in the frequency spectral analysis. In contrast, plain areas are smooth and their spectra are in the low frequency regions. The different spectral characteristics, can also be taken as image features. Hence, a statistical measurement can be done on frequency transformed pixel values (transform coefficients) in order to measure how much information exists in the frequency domain. Like statistical measurement in the spatial domain, the statistical measures such as mean, variance, standard deviation, energy and histogram are also applicable in the frequency domain [19,30,27,31,69,64].

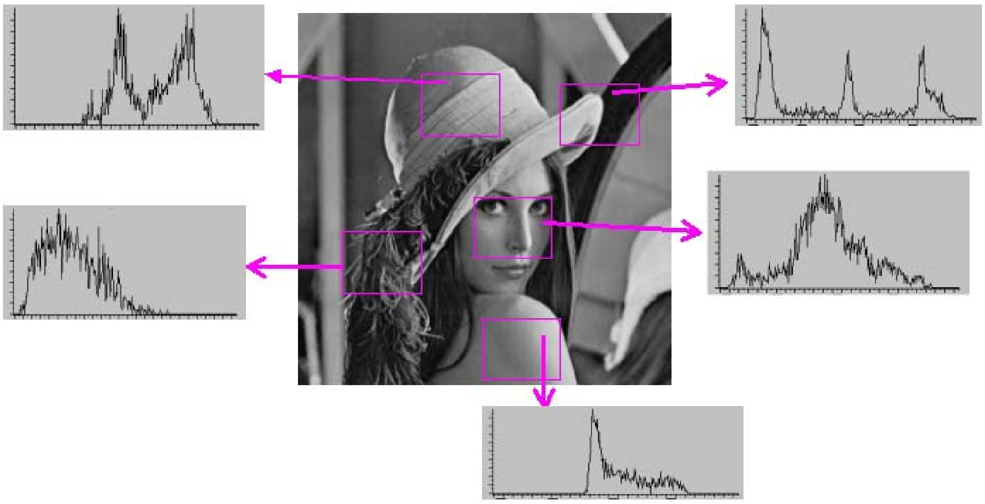


Figure 2.1: Non-stationary characteristic of image.

2.4 JPEG image compression standard

An effective compression scheme should provide a small "image" file size while still preserve most of the image characteristics. Among some recently proposed compression schemes [36,37,68], JPEG is the most common one used in WWW and multimedia systems. There are two compression modes, lossless and lossy, in the JPEG compression standard specified by the JPEG committee [36].

JPEG compresses images using the Discrete Cosine Transform (DCT) in a block-by-block basis [67]. The block structure is used so that the stationary assumption in DCT can hold. DCT assumes that the statistical measurement is the same within a whole DCT block. However, it was shown in section 2.3 that images are non-stationary. Considering the whole image as a block and applying DCT on the whole image cannot represent the non-stationary image properly. To fit into the stationary assumption in DCT, block-by-block operation can be employed so that pixels inside every small block can be assumed to be stationary. Thus the block-by-block DCT transforms images from spatial to spatial-frequency domain.

To compress an image, the following procedures are involved. They are transformation, quantization and entropy coding as illustrated in figure 2.2. Images are firstly sub-divided into 8×8 blocks and shifted from unsigned integers to signed integers. Each block then undergoes a DCT independently. Figures 2.2 and 2.3 show the procedure to encode/decode a single component, i.e., either a luminance or chrominance component. To encode a grayscale image, the input image block is shifted to signed grayscale intensity level. Luminance and chrominance color components in color image are regarded as multiple grayscale images. Thus, the luminance and chrominance components are compressed either in the color planes one by one or by interleaving the 8×8 blocks from each color plane in turn. Though no color transformation is actually defined in JPEG, it is expected that color images should be in YCbCr representation. Therefore, to deal with RGB component images, YCbCr transformation is employed to remove any correlation between RGB components and converted into luminance and chrominance channels. Following forward DCT, quantization, 'zig-zag' scanning and entropy coding are then used to reduce the number of bits to represent the DCT coefficients. During the process of quantization, insignificant coefficients, i.e., coefficients close to zero, are quantized to zero. 'Zig-Zag' scanning organizes the two dimensional (2D) DCT coefficient block into a one dimensional (1D) coefficient stream such that important low frequency coefficients and insignificant high frequency coefficients are placed at the front and at the back of the stream respectively (Figure 2.4). To support the JPEG progressive mode, spectral selection is the simplest method. Successive scans are used to partition the DCT coefficients stream after the 'zig-zag' scan. The first, second and remaining scans code the DC, one or more AC and the additional AC coefficients of each block respectively until all of the AC coefficients have been processed. The resultant 1D coefficient steam then undergoes entropy coding which encodes data according to its statistical occurrence. In other words, frequent symbols are represented using a short code word whereas infrequent symbols are represented by a long code word. The resultant output is the compressed image data that are stored in multimedia systems.

To reconstruct from the encoded image (Fig. 2.3), entropy decoding, re-organization and de-quantization are the first three procedures to reconvert compressed image data back to DCT transformed image. Then, inverse DCT is carried out to the outputs of de-quantizer. Since forward DCT and inverse DCT are a pair of forward and backward operations, outputs of inverse DCT operation are actually the inputs of forward DCT assuming perfect reconstruction. By undergoing the inverse DCT operation, the reconstructed image data are obtained.

DCT employs cosine function as its kernel as shown in Equation 2.1. By considering the input 8×8 pixel block as a 64-point discrete signal, the frequency content of the pixel block can be extracted into different frequency regions. The forward DCT decomposes the input spectrum into low and high frequency components. As the DC coefficient represents the average of the 8×8 block, it usually takes up a significant fraction of the total block energy. The other 63 AC coefficients relate to the frequency spectrum of the image block at different frequency regions. Therefore, the AC coefficients take up the remaining small fraction of the total block energy. Usually, a typical image contains lots of plain areas but few edges. Thus, the frequency content decreases with higher frequency. In other words, higher frequency AC coefficients such as AC_{77} are often insignificant, i.e. close to zero, whereas the lower frequency AC coefficients such as

AC_{01} are significant. Compression is then achieved by encoding only the significant coefficients but ignoring the insignificant coefficients. Inverse DCT at the decoder reverses the above procedure to reconvert the DC and AC coefficients back to 8×8 pixel block.

Since DCT is performed on a block-by-block basis, it ignores the continuity between pixels in the original spatial domain. The DCT blocks are treated as independent units and are considered to have no relationship between them. At high compression ratio, the fine details of the blocks are discarded since the AC coefficients are often regarded as insignificant information. This breaks the continuity of pixels between blocks in the reconstructed image. Therefore, blocking artifacts is the main artifacts of JPEG, especially at high compression ratios.

The DCT equation is

$$X_{m,n}(u,v) = \frac{1}{4}C(u)C(v)\sum_{i=0}^{7}\sum_{j=0}^{7} x_{m,n}(i,j)\cos\frac{(2i+1)u\pi}{16}\cos\frac{(2j+1)v\pi}{16}$$

$$x_{m,n}(i,j) = \frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7} C(u)C(v)X_{m,n}(u,v)\cos\frac{(2i+1)u\pi}{16}\cos\frac{(2j+1)v\pi}{16}$$

where $X_{m,n}(u,v)$ is the DCT coefficient in block *(m,n)* at position *(u,v)*, (2.1)
$x_{m,n}(i,j)$ is the input pixel at location *(i,j)* and
$C(u), C(v) = 1/\sqrt{2}$ for $u, v = 0$; $C(u), C(v) = 1$ otherwise

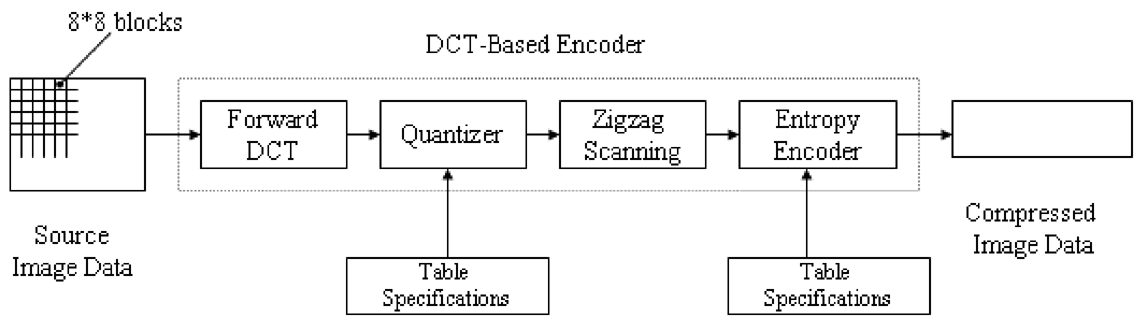


Figure 2.2: DCT-based encoder

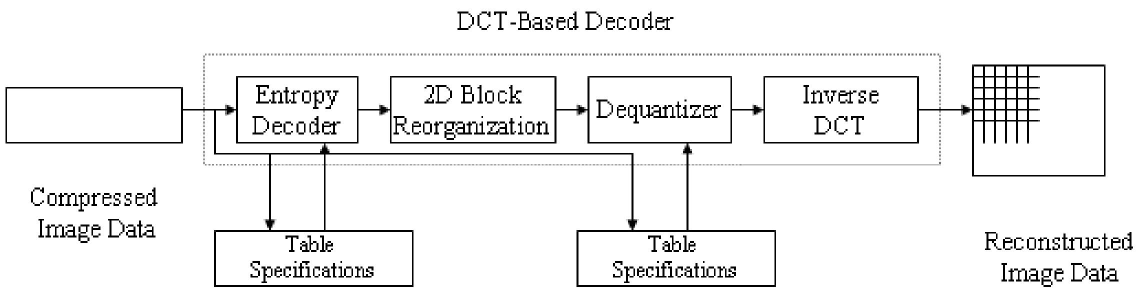


Figure 2.3: DCT-based decoder

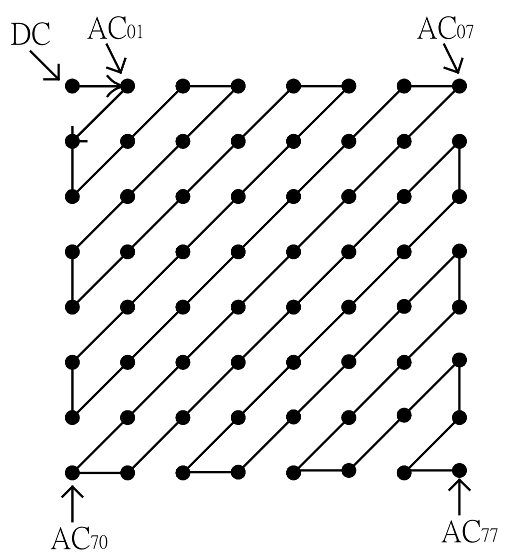


Figure 2.4: Zig-zag sequence of DC and AC coefficients in an 8×8 DCT coefficient block

2.5 JPEG2000 compression standard

JPEG2000 compression standard is a newly developed image compression scheme and released by the JPEG committee as an improvement to the JPEG compression standard [37]. Similar to JPEG, JPEG2000 provides lossless and lossy compression modes. JPEG2000 provides many advantages over JPEG [51]. For example, regions of interest coding by progression, random access to the bitstream, progressive transmission by quality, pan and zoom, compression quality and compression ratio. It compresses images using the wavelet transform (WT) [28,6], which is entirely different from the DCT used in JPEG [67]. Wavelets are well localized in both time and frequency. Thus, they can provide a multiple resolution view of an image. Due to the multiple resolution property, JPEG2000 allows progressive compression such that image quality can be refined in layers during the encoding and decoding process.

An image is firstly divided into rectangular, non-overlapping tiles with arbitrary sizes on a regular grid. Each tile should be of the same size and tiles should cover the entire image. A grayscale image can be regarded as a single component image. When the target image is a color image, either YCbCr transform or Reversible Component Transform (RCT) should be conducted to remove the correlation between RGB color data by converting into luminance and chrominance channels. After the transformation, all color components are treated as independent components.

To convert from the spatial to spatial-frequency domain, a L-level dyadic wavelet transform is applied on all components inside each tile of the image. According to JPEG2000 Part-I, floating-point wavelet kernel (9,7) and integer wavelet kernel (3,5) are used for lossy and lossless compression respectively. Note that JPEG2000 Part-II allows users to define their wavelets plus multiple wavelets. As shown in figure 2.5, after the forward wavelet transform, all wavelet coefficients undergo a uniform scalar quantization to reduce the size of the bitstream for latter coding. Uniform scalar quantization means that only one quantization step size is allowed at each subband. Following quantization, packet partition divides the quantized wavelet coefficients in each subband into a number of non-overlapping rectangular code-blocks. Entropy coding is then performed on each code-block. Code-blocks are encoded in a bitplane order such that the most significant bit (MSB) of all the coefficients are coded followed by the next MSB until the less significant bit (LSB). Progressive coding is achieved by coding the most significant information first followed by the less significant information. MQ-coder as defined in the JBIG-2 standard is adopted as the arithmetic entropy coder in JPEG2000 to compress the information in all packets [38]. Outputs of MQ-coder are the final bitstreams that are stored in JPEG2000 file format.

The algorithmic steps of the JPEG2000 decoder are just the reverse of the JPEG2000 encoder. To reconstruct the image, the bitstreams undergo MQ-decoding, de-quantization and inverse wavelet transformation. Reconstruction can also be done progressively such that a coarse image is reconstructed first and made finer successively by adding details into it.

Images are converted from the spatial domain to the spatial-frequency domain by WT decomposition as illustrated in Figure 2.6. The input $X(z)$ passes through the lowpass filter ($H(z)$) and highpass filter ($G(z)$), followed by a down-sampling by 2 operation. After passing through a number of lowpass or highpass filters, the output signal is decomposed into lowpass or highpass components. After a number of wavelet decompositions, the subbands are partitioned as shown in Figure 2.7. The symbol ‘L’ denotes the low frequency subband whereas ‘H’ represents high frequency subband. For example, HL_2 means that this subband contains high frequency and low frequency coefficients in the horizontal and vertical directions respectively, at the second level of decomposition. Figure 2.8 shows an example of the Lena image and its frequency coefficients after a three level decomposition.

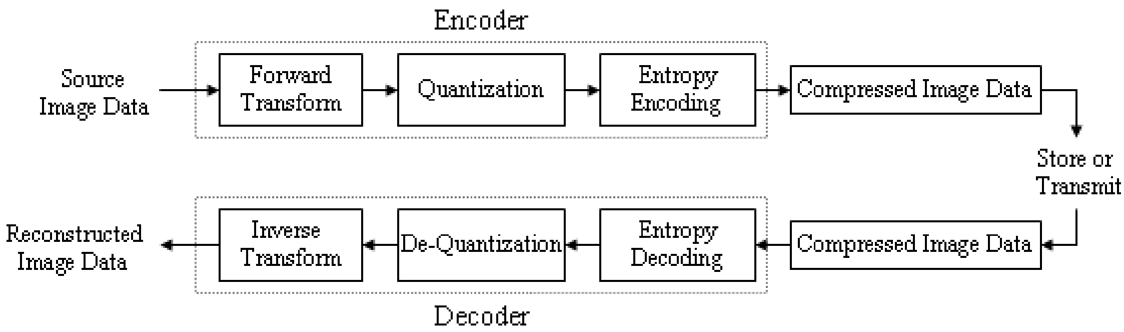


Figure 2.5: Basic structure of the JPEG2000 coding scheme.

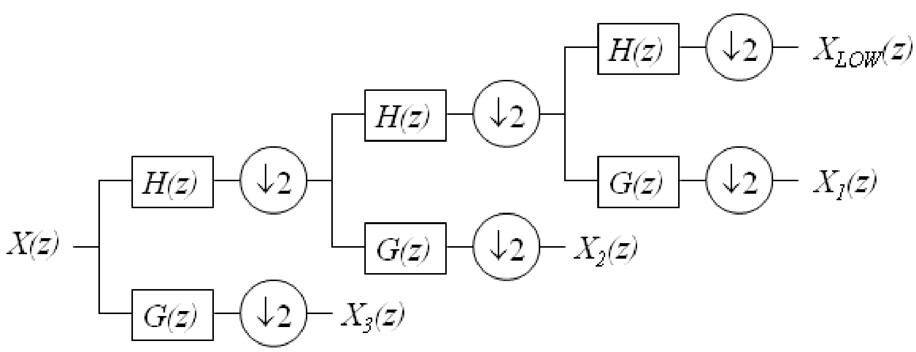


Figure 2.6: 1D WT decomposition by lowpass and highpass filters

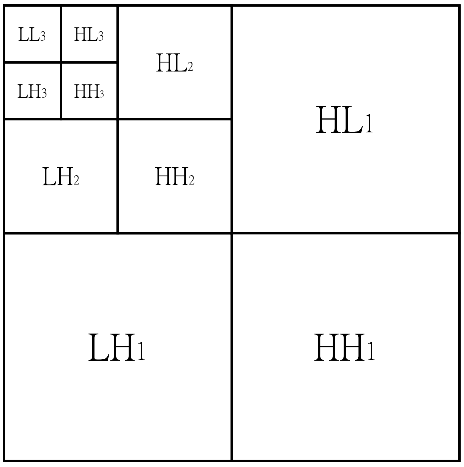


Figure 2.7: Wavelet subbands with three decomposition levels

(a) *(b)*

Figure 2.8: Lena image (a) in spatial domain and (b) its subband information after three decomposition levels

2.6 Retrieval systems in the spatial domain

Over the past few years, many retrieval systems operating in the spatial domain have been proposed. NEC ART MUSEUM is an example of an early content-based image retrieval system [63]. It finds image edge features and stores an edge map for each image in the database. It finally sorts all the images based on their scores and presents the first one to the user as a matching image. The IBM QBIC system uses multiple features for query [43]. These features include color histogram, texture, shapes and spatial relationship of objects. This retrieval system supports comparing each of the features separately. Smith and Chang developed VisualSEEK to extract regional color information and salient color regions [33,35]. The Virage image search engine [1] utilizes color, structure and texture information to extract global and local color information. All these systems operate in the uncompressed domain and aim to extract spatial information for retrieval, such as histogram, texture, structure and shapes. Carson and Ogle developed the Chabot project to identify “regions / blobs” within the image and indexing is done at the blob level [7]. Therefore, indexing each of the blobs separately is the main difference between earlier retrieval systems. Statistical measurement about pixel intensity is also studied. Mean, standard deviation, variance, histogram of grayscale values are the most commonly used measures [69, 59, 25, 47, 18, 29, 57, 64,58].

2.7 Retrieval systems in the JPEG compressed domain

JPEG is a commonly used image compression scheme in multimedia systems and the WWW. When images are compressed, feature characterizations and similarity measures operating on the compressed data is desirable. Since compression is using the smallest numbers of bit to store significant information, retrieving in the compressed domain is to compare the compressed features instead of pixel-by-pixel operation in the uncompressed domain. Using directly the compressed data can reduce computational complexity and processing time since this avoids the decompression operations. Therefore, retrieval systems in JPEG compressed domain have been investigated.

JPEG compresses images using DCT. Proposed indexing techniques involve extracting low-level features, such as color, shape and texture, by analyzing DCT coefficients [55,12,50]. To retrieve JPEG images, direct comparison can be done on the features that are extracted from the DCT blocks [9]. The major concern to the DCT domain analysis is the lack of spatial information contained in the DCT coefficients. Therefore, it is not easy to visualize shape, such as edges, in the DCT domain. To tackle this problem, A. Vellaikal et al. [4] proposed a joint spatial-spectral indexing method to explore the spatial information from the spectral content by organizing the DCT blocks in a tree structure. C.W. Ngo et al. proposed exploring features in DCT blocks from quad-tree structure [9,56]. DCT blocks are grouped to form subsets and embedded into the leaves in the quad-tree. Features are then extracted directly from the DC and AC coefficients from the quad-tree leaves. The quad-tree structure can effectively reduce the indexing time because the retrieval result is refined progressively by global features from the tree root followed by finer details from the leaves. Statistical measurement such as energy histogram, mean and variance are used to study the statistical features of JPEG images [17]. To make the retrieval system rotation invariant, joined feature vector including color histogram and moment invariants are proved to provide good retrieval results [9]. Texture is also taken into consideration since it is an important feature in most images. D. S. Wu et al. proposed reordering all the DCT coefficients into wavelet subband structure so as to capture spatial-spectral texture features in images for retrieval and classification [13].

2.8 Retrieval systems in JPEG2000 compressed domain

JPEG2000 is a new image compression standard as an improvement on the JPEG image compression standard. It compresses images using wavelets. Owing to the well localization in both time and frequency, wavelets can produce a multiple resolution view of an image. Several indexing techniques have been proposed to extract features in the WT subbands. Subband energy, subband magnitudes, salient points, the number of significant coefficients and statistical measures are proposed for indexing [22,73,40,62,52,41]. The multiple resolution property of WT allows indexing to be started at the lowest resolution level. The indexing results are then progressively refined using higher resolution coefficients [14]. K.C. Liang et al. suggested extracting four features from the wavelet domains, which are: frequency of dominant coefficients in each subband, luminance histogram of successive quantized coefficients, binary quantization map of the coefficients in the lowest frequency subband, and color histogram of coefficients. To make the indexing results to be rotation, translation and scaling invariant, they propose extracting features from subbands [42] by using normalized central moments [46].

2.9 Chapter Summary

In this chapter, components that compose grayscale and color images are discussed. Numerous coloring schemes are proposed for different applications and result in various color formats. Owing to the constraints of transmission and storage, images are usually in compressed form. In spatial domain, images may be in color formats other than YCbCr. But they are usually converted to YCbCr color scheme during image compression. This can unify all compressed images in the same color domain for further processing.

Two image compression schemes are briefly introduced in this chapter. They are the commonly used JPEG and the newly developed JPEG2000 schemes. Due to the shortcomings of JPEG especially at high compression ratio, JPEG2000 is developed to provide a high quality compressed image while maintaining a good compression ratio. As images are commonly compressed in these two domains, searching images in compressed domains becomes an important application in multimedia systems and the World Wide Web. Recently, many image indexing algorithms are proposed to operate directly in JPEG and JPEG2000 compressed domains. However, they function in single domain only and thus are unable to search images in multiple domains.

As JPEG2000 is released as an improvement of JPEG, it is regarded as the next main-stream compression scheme. Currently, JPEG has been in existence for more than a decade so that most of the existing information is stored in JPEG format. To search for images in these two compressed domains by decompressing all information back to the spatial domain is extremely inefficient. It is most desirable that indexing can be performed in their original compressed domains. Therefore, direct indexing in multiple compressed domains becomes an important issue in modern multimedia applications. Indexing in these domains is possible only when similarity exists between them. Therefore, similarities between JPEG and JPEG2000 will be examined in the next chapter.

Chapter 3

Analysis on spectrum characteristics of DCT and WT

3.1 Introduction

Currently, retrieving compressed images in multiple compressed domains requires processing in the uncompressed spatial form in which common features can be extracted for indexing. However, many pre-processing operations are incurred in decompression, especially for large image archives. On the contrary, directly using the compressed data for feature characterization can undoubtedly reduce the computational complexity. Since there are a variety of compression techniques, features that can be extracted from the compressed domains depend greatly on the compression techniques that are used. With reference to our objective of studies, a retrieval system compatible to different compressed domains is desirable. As JPEG and WT-based compression techniques are the two most popular techniques in use nowadays, our studies will focus on image indexing in these two domains.

In chapter 2, we have briefly discussed two compression standards, i.e., the commonly used JPEG and the newly developed JPEG2000 schemes. By comparing the two coding algorithms, we notice that their difference starts from their transformation methods. JPEG and JPEG2000 employ dissimilar block-based DCT and WT as their transformation schemes respectively such that their compressed outputs seem to be unrelated. To investigate whether similar features exist in these two domains, images in the JPEG compressed domain are compared with that in the JPEG2000 compressed domain. The aim of this chapter is therefore to find a unified framework to study JPEG and JPEG2000 so that same features can be extracted directly from these domains without decompressing the compressed images back onto the spatial domain.

In this chapter, we will focus on the spectrum characteristics of DCT and WT. To facilitate our discussion, we will briefly discuss filter characteristics first. Then, a common filter framework is developed so that a direct comparison between DCT and WT is possible. Three aspects of comparisons are carried out to verify our derived filtering framework. They are the passband region, power spectrum and energy preservation which all are important characteristics in signal analysis. From our studies, it is proved that although DCT and WT belong to different transformation schemes, their filtering operations are very similar in some aspects.

3.2 Characteristic of filters

In this section, we will define some parameters which are commonly used to evaluate filter characteristics. It is well known that filter analysis is often carried out in the frequency domain instead of the spatial domain [16]. In the frequency space, it is easier to infer their degree of stability and to visualize their frequency response. Some important characteristics of filters are, magnitude spectrum, passband region and energy spectrum.

In the following filter analysis, the z-domain is used. The relationship between input *X(z)* and output *Y(z)* is

$$Y(z) = H(z)X(z) \tag{3.1}$$

In z-domain, we take

$$z=e^{j\omega} \tag{3.2}$$

where ω denotes the frequency and is given by the expression

$$\omega = \frac{2\pi k}{N} \tag{3.3}$$

where $k \in [0, N-1]$. Assume that *H(z)* is complex such that it has both real and imaginary parts. Therefore,

$$H(\omega) = Re(\omega) + jIm(\omega) \tag{3.4}$$

In the following, we will study the characteristics of filer *H(z)* in three aspects: magnitude spectrum, passband region and energy spectrum.

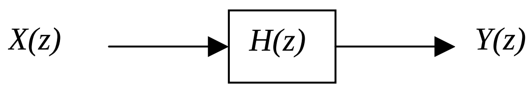


Figure 3.1: A filtering system with a transfer function H(z), input X(z) and output Y(z) in the z-domain.

3.2.1 Magnitude spectrum

Frequency response determines the output of a particular input. It consists of magnitude and phase spectra. Magnitude spectrum is the absolute amplitude of the transfer function, i.e. $|H(z)|$, which shows its response to an input signal at different frequency. Phase spectrum is denoted by $\angle H(z)$. Since $z=e^{j\omega}$, the phase spectrum is indeed

$$\phi(\omega) = \tan^{-1}\frac{\mathrm{Im}(\omega)}{\mathrm{Re}(\omega)} \tag{3.5}$$

We can verify the filter stability either by plotting $\phi(\omega)$ against frequency ω or by the poles and zeros in the z-plane.

Magnitude response provides valuable information about the filter such as passband region whereas phase response indicates filter stability. As different filter kernels are used in DCT & WT, one should compare their difference as well as similarities.

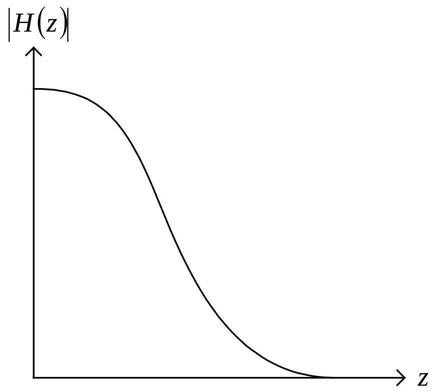


Figure 3.2: Magnitude spectrum of a lowpass filter.

3.2.2 Passband Region

Magnitude spectrum contains information such as filters passband and stopband regions. To determine these regions, one has to find out the 3-dB cutoff frequency f_c which is defined as,

$$f_c = 0.707 \times max(|H(z)|) \tag{3.6}$$

Frequencies whose magnitudes lie above f_c define the passband of a filter. From the location of the passband regions, we can distinguish whether filters are lowpass, bandpass, bandstop or highpass. Figure 3.3 shows the passband region of a transfer function to be [0, f_c] which belongs to a typical lowpass filter.

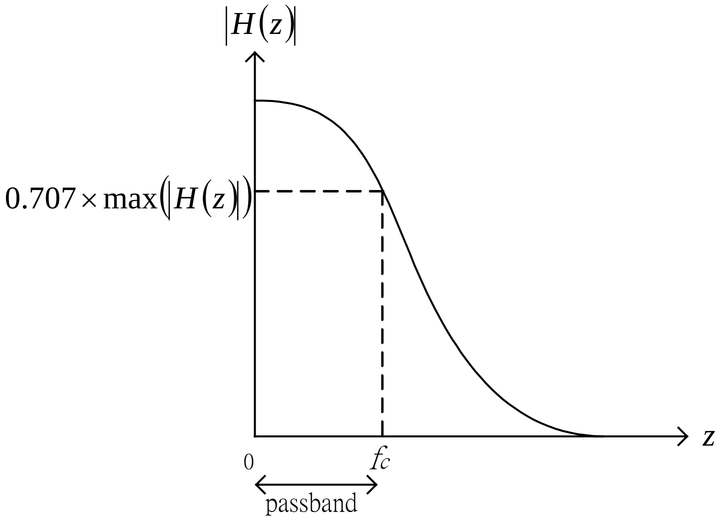


Figure 3.3: Passband and cutoff frequency f_c of a lowpass filter

3.2.3 Energy spectrum

Besides magnitude spectrum and bandwidth of filters, the energy spectrum can also show the energy preservation of a filter. The energy spectrum $E(z)$ is defined as

$$E(z) = |H(z)|^2 \tag{3.7}$$

However, since the maximum magnitudes of the energy spectrum differ from one filter to another, it is inaccurate to compare their energy preservation by just looking at their spectrum magnitude. To make a fair comparison, one need to use the normalized energy spectrum $E_N(z)$ defined as,

$$E_N(z) = \left(\frac{|H(z)|}{\max(|H(z)|)} \right)^2 \tag{3.8}$$

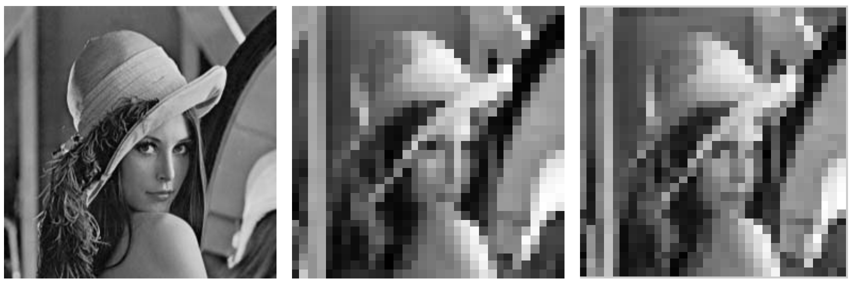
The magnitude of the normalized energy spectrum lies between 0 and 1, which provides a fair comparison between filters.

3.3 Output signal similarity between DCT and WT

In section 3.2, we have defined three filter characteristics, namely magnitude spectrum, passband region and energy spectrum. We can use them to compare the filtering operations of BDCT and WT. Before we look at the filter characteristics of the BDCT in JPEG and WT in JPEG2000, we would like to examine whether similarity really exists between the BDCT and WT operations.

Both JPEG and JPEG2000 present the information contents of images in the spatial-frequency domain. In accordance with the equation of BDCT in Eqn 2.1, all $X_{m,n}(0,0)$ are the DC coefficients (average) of the BDCT blocks. Concatenating all the DC coefficients, we can obtain the average pixel intensity at all blocks. The concatenated DC coefficients give a coarse or rough sketch of an image as can be seen in Figure 3.4(b). JPEG2000 uses WT to map images from spatial domain onto spatial-frequency domain. As shown in figure 2.8 in section 2.5, the low-low band gives coarse information of the original image (Fig. 3.4(c)). Referring to figure 3.4(b) and 3.4(c), a small blurred version of the original image can be reconstructed from either concatenating the DC coefficients from all DCT blocks or the low-low subband coefficients in WT. As in figure 3.4, spatial and frequency information regarding the original image are preserved in both transformed domains. Though Fig.3.4(b) and Fig.3.4(c) are obtained by non-similar BDCT and WT schemes respectively, their highly similar appearance indicates that the two transformation schemes may be partially similar to each other. Therefore, the relationship between DCT and WT are worthy of studies. By analyzing the filtering operations employed in the transformation schemes in BDCT and WT, retrieving JPEG and JPEG2000 compressed images may be simplified without decompressing all the images.

In the following two sections, a mathematical formulation regarding the BDCT and WT coarse image is formed which is used to theoretically study the similarity between the low frequency outputs from these two transformations.



(a) *(b)* *(c)*

Figure 3.4: Comparison between (a) Uncompressed image and blurred images formed by (b) DC coefficients of BDCT and (c) lowpass band of WT

3.3.1 DC signal from BDCT

In the following, we will investigate the DC image formed by DC coefficients of the BDCT. Since the 2D transformation is obtained from 1D row transformation followed by 1D column transformation, the mathematical formulation is done in 1D for simplicity.

The DC coefficient in each block is concatenated together to form a blurred version of the original image/signal. It is called the DC signal in the following discussion. Let x_α be the original signal, of length N and $\alpha \in [0,N-1]$, then the DC signal becomes,

$$\begin{bmatrix} X_0(0) \\ X_1(0) \\ \vdots \\ X_{\frac{N}{8}-1}(0) \end{bmatrix}_{\frac{N}{8}x1} = \begin{bmatrix} A_{1x8} & \mathbf{0}_{1x8} & \cdots & \mathbf{0}_{1x8} \\ \mathbf{0}_{1x8} & A_{1x8} & \cdots & \mathbf{0}_{1x8} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{1x8} & \mathbf{0}_{1x8} & \cdots & A_{1x8} \end{bmatrix}_{\frac{N}{8}xN} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}_{Nx1} \tag{3.9}$$

where $X_n(0)$ is the DC coefficient at the nth block and

$$\mathbf{A}_{1x8} = \frac{1}{2\sqrt{2}}(1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1) \tag{3.10}$$

$$\mathbf{0}_{1x8} = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0) \tag{3.11}$$

Using z-transform, the DC and the original signals can be expressed respectively as,

$$\tilde{X}_{DC}(z) = \sum_{n=0}^{\frac{N}{8}-1} X_n(0) z^{-n} \tag{3.12}$$

$$\tilde{X}(z) = \sum_{\alpha=0}^{N-1} x_\alpha z^{-\alpha} \tag{3.13}$$

From Eqn. 3.9, we can see that,

$$X_n(0) = \frac{1}{2\sqrt{2}} \sum_{i=0}^{7} x_{8n+i} \tag{3.14}$$

Using Eqn. 3.9 to Eqn. 3.14, it can then be shown that,

$$\tilde{X}_{DC}(z) = \tilde{F}_{DC}(z)\tilde{X}(z) \quad \downarrow 8 \tag{3.15}$$

where $\downarrow 8$ denotes a down-sampling by 8 operation and,

$$\tilde{F}_{DC}(z) = \frac{1}{2\sqrt{2}} \sum_{i=0}^{7} z^{-i} \tag{3.16}$$

Eqn.3.15 shows that the DC signal can be obtained by employing a filtering operation followed by a sub-sampling operation. The characteristic of the DC signal is thus dependent on the filter $\tilde{F}_{DC}(z)$. Figure 3.5 shows the resultant implementation. This reformulation helps us to compare the DC signal and the lowpass signal from the WT under the same framework.

$$\tilde{X}(z) \longrightarrow \boxed{\tilde{F}_{DC}(z)} \longrightarrow \boxed{\downarrow 8} \longrightarrow \tilde{X}_{DC}(z)$$

Figure 3.5: The relationship between the DC signal and the original signal from the perspective of a filtering operation.

3.3.2 Lowpass signal from WT

The lowpass signal from WT is obtained by a three-level decomposition of the original signal as shown in figure 2.6. In each decomposition level, the signal is lowpass and bandpass filtered which is followed by a sub-sampling by 2 operation. Three-level of decomposition is chosen so that the size of the lowpass signal is same as that of the DC signal. Mathematically, in each decomposition level, the wavelet filter, F_{WT}, can be constructed as,

$$F_{WT} = \begin{bmatrix} h_0 & h_1 & \cdots & & & & h_{N-2} & h_{N-1} \\ g_0 & g_1 & \cdots & & & & g_{N-2} & g_{N-1} \\ h_{N-2} & h_{N-1} & h_0 & h_1 & & & & \vdots \\ g_{N-2} & g_{N-1} & g_0 & g_1 & & & & \\ \vdots & & & & \ddots & & & \\ & & & & & & & \\ h_2 & h_3 & \cdots & h_i & \cdots & h_{N-1} & h_0 & h_1 \\ g_2 & g_3 & \cdots & g_i & \cdots & g_{N-1} & g_0 & g_1 \end{bmatrix}_{N\times N} \tag{3.17}$$

where h_i and g_i denote respectively the lowpass and bandpass filters. To obtain the lowpass signal, the original signal is lowpass-filtered followed by a subsampling by two operation three times as those in figure 3.6. Equivalently, the filtering and the subsampling operations can be reversed in order which results in the implementation as shown in Figure 3.7. Note that the filter F_{LOW} can be written as,

$$F_{Low} = F_L\Big|_{N=64} F_L\Big|_{N=128} F_L\Big|_{N=256} \tag{3.18}$$

where

$$F_L = \begin{bmatrix} h_0 & h_1 & \cdots & & h_{N-1} \\ h_{N-2} & h_{N-1} & h_0 & & \vdots \\ \vdots & & & \ddots & \\ h_2 & \cdots & & & h_1 \end{bmatrix}_{\frac{N}{2}\times N} \tag{3.19}$$

In z-domain, the lowpass signal can be expressed as,

$$\tilde{X}_{Low}(z) = \tilde{F}_{Low}(z)\tilde{X}(z) \quad \downarrow 8 \tag{3.20}$$

where the lowpass filter can be constructed as,

$$\tilde{F}_{Low}(z) = \tilde{F}(z)\tilde{F}\left(z^2\right)\tilde{F}\left(z^4\right) \tag{3.21}$$

and

$$\tilde{F}(z) = \sum_{n=0}^{N-1} h_n z^{-n} \tag{3.22}$$

By using this formulation, the characteristic of the lowpass signal depends solely on the filter $\tilde{F}_{Low}(z)$.

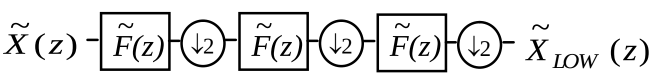


Figure 3.6: Lowpass signal from three-level wavelet transform.

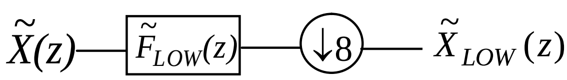


Figure 3.7: The relationship between the lowpass signal and the original signal from the perspective of a filtering operation.

3.3.3 Lowpass similarity between BDCT and WT

In subsection 3.2.1 and 3.2.2, a mathematical formulation is derived to obtain the DC and lowpass signal from BDCT and WT. From the above formulation, we show that both DC and lowpass images can be obtained by passing the signal through a filter and then a downsampling by 8 operation afterwards. Therefore, to compare the DC signal and the lowpass signal, we need to study the lowpass filters $\tilde{F}_{DC}(z)$ and $\tilde{F}_{Low}(z)$. To generalize our formulation, a mathematical framework is derived in the next section to analyze the characteristics of both the lowpass and bandpass filters that are employed in BDCT and WT.

3.4 Mathematical framework for common feature analysis in BDCT and WT

BDCT and WT are two different transformation schemes. The similarity between the outputs of this two schemes greatly depends on their filtering operations. In the following sub-sections, detailed mathematical formulation of BDCT and WT is derived. It is desirable that a similar filtering model can be derived to facilitate similarity comparison.

3.4.1 BDCT

As x_α is the original signal of $\alpha \in [0,N\text{-}1]$, the 1D BDCT coefficients $X_m(u)$ can be written as,

$$X_m(u) = \frac{1}{2} C(u) \sum_{i=0}^{7} x_{8m+i} \cos\frac{(2i+1)u\pi}{16} \quad \text{for } u = \{0,1,\ldots,N-1\} \tag{3.23}$$

where *C(u)* equals $1/\sqrt{2}$ for *u*=0 and 1 otherwise. Those BDCT coefficients with same *u* can be concatenated together so as to provide the spatial-frequency information, i.e,

$$\begin{bmatrix} X_0(0) \\ X_1(0) \\ \vdots \\ X_{\frac{N}{8}-1}(0) \end{bmatrix}_{\frac{N}{8}x1} = \begin{bmatrix} A_{1x8} & 0_{1x8} & \cdots & 0_{1x8} \\ 0_{1x8} & A_{1x8} & \cdots & 0_{1x8} \\ \vdots & \vdots & & \vdots \\ 0_{1x8} & 0_{1x8} & \cdots & A_{1x8} \end{bmatrix}_{\frac{N}{8}xN} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}_{Nx1} \tag{3.24}$$

where $\mathbf{0}_{1x8}$ is a 1x8 zero vector and

$$\mathbf{A}_{u,1x8} = \frac{C(u)}{2}\left(\cos\frac{u\pi}{16} \quad \cos\frac{3u\pi}{16} \quad \cos\frac{5u\pi}{16} \quad \cos\frac{7u\pi}{16} \quad \cos\frac{9u\pi}{16} \quad \cos\frac{11u\pi}{16} \quad \cos\frac{13u\pi}{16} \quad \cos\frac{15u\pi}{16}\right) \tag{3.25}$$

Following similar formulation in section 3.1.1, we found that the concatenated BDCT coefficients in Eqn. 3.24 are obtained by the filtering and then a sub-sampling operation,

$$\tilde{X}_{u,DCT}(z) = \tilde{F}_{u,DCT}(z)\tilde{X}(z) \quad \downarrow 8 \tag{3.26}$$

where $\tilde{F}_{u,DCT}(z)$ is formed using Eqn. 3.23 and Eqn. 3.24 as

$$\tilde{F}_{u,DCT}(z) = \frac{C(u)}{2} \sum_{i=0}^{7} \cos\frac{(2i+1)u\pi}{16} z^{-i} \tag{3.27}$$

The spatial-frequency characteristics of the concatenated BDCT coefficients are dependent on the filter $\tilde{F}_{u,DCT}(z)$. To compare the spatial-spectral characteristics of BDCT with that of WT, filtering formulation of WT will be derived in section 3.4.2.

3.4.2 Wavelet Transform

In each decomposition level, the signal is lowpass and bandpass filtered which is followed by a sub-sampling by 2 operation as shown in figure 2.6. Assuming a three-level decomposition, the lowpass and bandpass signals can be written respectively as,

$$\begin{aligned} \tilde{X}_{0,WT}(z) &= \tilde{F}_{0,WT}(z)\tilde{X}(z) \quad \downarrow 8 \\ \tilde{X}_{i,WT}(z) &= \tilde{F}_{i,WT}(z)\tilde{X}(z) \quad \downarrow 2^{4-i}, i = 1,2,3 \end{aligned} \tag{3.28}$$

where

$$\tilde{F}_{i,WT}(z) = \begin{cases} \tilde{H}(z)\tilde{H}(z^2)\tilde{H}(z^4) & i = 0 \\ \tilde{G}(z) & i = 3 \\ \tilde{G}\left(z^{2^{3-i}}\right)\prod_{j=1}^{3-i}\tilde{H}(z^j) & i = 1,2 \end{cases} \tag{3.29}$$

$$\tilde{H}(z) = \sum_{n=0}^{N-1} h_n z^{-n} \quad \tilde{G}(z) = \sum_{n=0}^{N-1} g_n z^{-n} \tag{3.30}$$

and h_i and g_i denote the lowpass and bandpass filters respectively.

3.4.3 Commonality between BDCT and WT

The filtering structure derived above facilitates our comparison between BDCT and WT. Both lowpass and highpass outputs of BDCT and WT are obtained by a filtering operation and a down-sampling operation afterwards. $\tilde{F}_{u,DCT}(z)$ is followed by a sub-sampling by 8 operation while the sub-sampling of $\tilde{F}_{i,WT}(z)$ depends on the decomposition level. Therefore, further investigation is required to analyze common features between $\tilde{F}_{u,DCT}(z)$ and $\tilde{F}_{i,WT}(z)$ such that the similarity between BDCT and WT outputs may be found. From Eqn. 3.30, we can see that the transfer function used in the lowpass and bandpass filtering are not defined exactly. This is because WT in JPEG2000 provides a high flexibility on kernel selection. To make our comparison realistic, four commonly used wavelet kernels are studied in the following section before further comparison on BDCT and WT can be carried out.

3.5 Commonly used WT kernels in image processing

As mentioned in section 3.4.2, the filtering operation in WT are composed of highpass and lowpass filters. However, no transfer functions are defined exactly for the lowpass and highpass filters. This is because JPEG2000 provides flexibility for the usage of the filters. As stated in JPEG2000 Part-I, it is recommended to use floating-point wavelet (9,7) and integer wavelet (3,5) for lossy and lossless compression schemes respectively [49,14]. JPEG2000 Part-II allows users to define their wavelets and the use of multiple wavelets. Numerous researchers analyze many other kernels in WT filtering and evaluate their performances in feature extraction [30,49,3]. The Haar and Daubechies families are commonly selected as the wavelet kernels for image processing. Haar is the simplest kernel with a low computational complexity [72]. Due to the high regularity of Daubechies' wavelet, smooth output images with less edge artifacts are obtained [24, 30]. According to the specification of JPEG2000 and results of past investigation by researchers, Haar, Daubechies 4 (DB4), Biorthogonal 93 (B93) and Biorthogonal 97 (B97) kernels are selected.

3.6 Comparison between the filtering operations of BDCT and WT

In this section, a comparative study is performed on BDCT and WT to find out their similarities at different passband regions, i.e., at all lowpass and highpass regions. Once the wavelet kernels are defined, characteristics of BDCT and WT can be compared. Eqn. 3.26 shows that the BDCT output is obtained by passing the input through a filter

$\widetilde{F}_{u,DCT}(z)$ followed by a down-sampling by 8 operation. Eqn. 3.28 shows that the WT output is obtained by passing the input through a filter $\widetilde{F}_{i,WT}(z)$ followed by a down-sampling operation. To sum up, for both BDCT and WT, the transformed outputs are always obtained by filtering and a down-sampling operation afterwards. This common subband model implies that spatial-spectral information is preserved by the two transforms, which forms a foundation for our theoretical analysis. By comparing the filters $\widetilde{F}_{u,DCT}(z)$ and $\widetilde{F}_{i,WT}(z)$, one can find out their spectral characteristics and then determine their spectral similarity. Three measures, passband region, filter similarity and energy spectrum, are considered. With the knowledge of filter similarity, we can examine the similarity of their outputs.

3.6.1 Frequency partition of BDCT and WT

In this section, the passband of BDCT and WT is the first characteristic that we would like to investigate. The magnitude spectra of $\widetilde{F}_{u,DCT}(z)$ and $\widetilde{F}_{i,WT}(z)$ are compared. It is found that $\widetilde{F}_{u,DCT}(z)$ partitions the frequency spectrum uniformly over the frequency range [-π,π]. For *u* from 0 to 7, the passband region of $\widetilde{F}_{u,DCT}(z)$ are (0-0.11)π, (0.09-0.26)π, (0.18-0.39)π, (0.30-0.51)π, (0.42-0.63)π, (0.54-0.76)π, (0.68-0.90)π and (0.85-1)π. In contrast, the width of the passband region of $\widetilde{F}_{i,WT}(z)$ increases with *i*. This implies that a particular wavelet filter $\widetilde{F}_{i,WT}(z)$ might share similar characteristics with a number of BDCT filters. To complicate the issue, $\widetilde{F}_{i,WT}(z)$ depends on the choice of the wavelet kernel. As can be seen in Table 3.1, Haar, DB4, B93 and B97 kernels give slightly different passband regions.

	$\tilde{F}_{0,WT}(z)$	$\tilde{F}_{1,WT}(z)$	$\tilde{F}_{2,WT}(z)$	$\tilde{F}_{3,WT}(z)$
Haar	0-0.11π	0.09π-0.29π	0.19π-0.61π	0.5π-π
DB4	0-0.12π	0.11π-0.27π	0.22π-0.56π	0.5π-π
B93	0-0.14π	0.16π-0.29π	0.30π-0.56π	0.64π-π
B97	0-0.13π	0.14π-0.28π	0.27π-0.55π	0.55π-π

Table 3.1: Passband regions for WT filters

3.6.2 Lowpass Filters Analysis

This subsection will focus on lowpass filtering studies. The formulation in sections 3.3.1 and 3.3.2 (Figure 3.5 and Figure 3.7) provides a foundation for a theoretical comparison between the DC signal in BDCT and the lowpass signal in WT. As can be seen in eqn.3.27, the expression of $\tilde{F}_{0,DCT}(z)$ is fixed. In contrast, the expression of $\tilde{F}_{0,WT}(z)$ depends on the choice of the wavelet kernel (Eqn. 3.28). Owing to their bandwidth difference, a number of BDCT fitlers may be concatenated together to be comparable to the WT filter.

In this sub-section, the magnitude spectra of $\tilde{F}_{0,DCT}(z)$ and $\tilde{F}_{0,WT}(z)$ are compared to observe their magnitude responses at all frequencies. Since $\tilde{F}_{0,DCT}(z)$ and $\tilde{F}_{0,WT}(z)$ are already defined by BDCT in JPEG and WT in JPEG2000 respectively, their filter responses are stabile. Thus, we will not compare their filter stabilities, i.e., their phase response $\phi(\omega)$. Their lowpass and bandpass filtering will be studied as follows.

3.6.2.1 Lowpass Filter For The Haar Kernel

Theoretically, from Eqn. 3.25 and Eqn. 3.29,

$$\widetilde{F}_{0,WT}(z) = \frac{1}{2\sqrt{2}} \sum_{i=0}^{7} z^{-i} = \widetilde{F}_{0,DCT}(z) \qquad (3.31)$$

With reference to Eqn. 3.27 and Eqn. 3.29, for $i=0$ and $u=0$, S equals to 1 in the Haar kernel, i.e., an exact match.

It can be seen that $\widetilde{F}_{0,DCT}(z)$ and $\widetilde{F}_{0,WT}(z)$ from Haar kernel are the same. Therefore, we can conclude that the DC image from BDCT and the lowpass image from WT using Haar kernel are exactly the same. In other words, same features can be found from these two compressed domains.

3.6.2.2 Lowpass Filters For Other Kernels

Besides the Haar kernel, we need to study other commonly used wavelet kernels in compression as discussed in section 3.5. Comparison will be done on the DB4, B93 and B97 wavelet kernels.

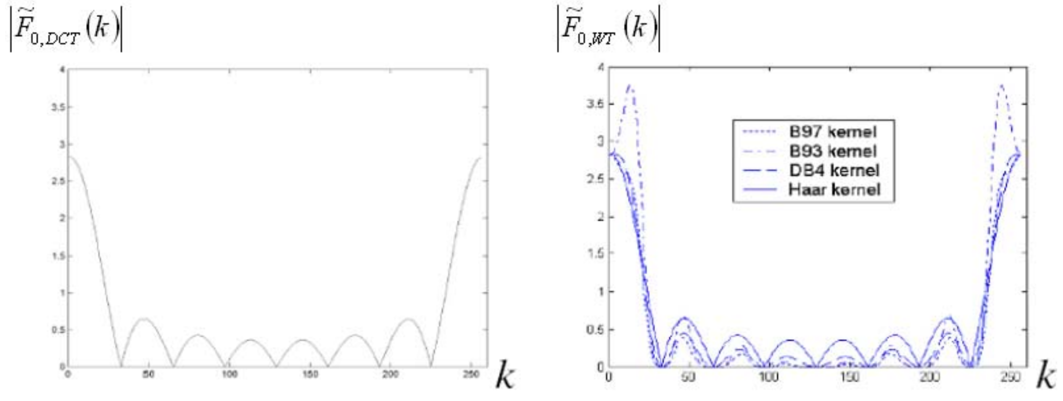
Due to the long length of these filters, it is inappropriate to write down all analytic expressions for $\left|\widetilde{F}_{0,WT}(z)\right|$. Instead, with $z = e^{j\frac{2\pi k}{N}}$ and $k \in [0,255]$, $\left|\widetilde{F}_{0,WT}(k)\right|$ for these kernels are plotted at all frequencies from 0 to 2π and compared with $\left|\tilde{F}_{0,DCT}(k)\right|$. Figure 3.8(a) shows the plot of $\left|\tilde{F}_{0,DCT}(k)\right|$ while Figure 3.8(b) shows that of $\left|\tilde{F}_{0,WT}(k)\right|$ using various wavelet kernels. In general, we can see that all of them match very well with each other at all frequencies.

To quantify their similarity, one can use the similarity measure [10] which is defined as,

$$S = \max_{j} \frac{\sum_{k=0}^{N-1} |A(k)||B(k-j)|}{\left(\sqrt{\sum_{k=0}^{N-1} |A(k)|^2}\right)\left(\sqrt{\sum_{k=0}^{N-1} |B(k)|^2}\right)} \tag{3.32}$$

for functions *A* and *B.* The similarity measure tries to find out how similar *A* and *B* are. Its value lies between 0 and 1. Large value means the shapes show a good match while a low value means they are not similar.

As can be seen in Figure 3.8, $\tilde{F}_{0,DCT}(z)$ from BDCT and $\tilde{F}_{0,WT}(z)$ from WT using Haar kernel are exactly the same. For other kernels, the similarities between $\tilde{F}_{0,DCT}(z)$ and $\tilde{F}_{0,WT}(z)$ are 0.98, 0.95 and 0.97 respectively for DB4, B93 and B97 kernels. Thus, they all have a large *S* value which means that filters are similar with a similar passband region (Table 3.1). In summary, the DC image from BDCT and the lowpass image from WT should have similar spectral content which can directly be used in indexing. In the next section, we will evaluate whether high similarity still hold in the bandpass filters.



(a) *(b)*

Figure 3.8: Plots of (a) $\left|\tilde{F}_{0,DCT}(k)\right|$ *and (b)* $\left|\tilde{F}_{0,WT}(k)\right|$ *for various wavelet kernels.*

3.6.3 Bandpass filters Analysis

In this subsection, the similarity between BDCT and WT bandpass filters will be investigated. Using the subband filtering model in section 3.4, we can see that $\tilde{F}_{1,WT}(z)$ can be directly compared to $\tilde{F}_{1,DCT}(z)$ since both involve a sub-sampling by 8 operation. The passband region of $\tilde{F}_{1,DCT}(z)$ is (0.09-0.26)π where the average of $\tilde{F}_{1,WT}(z)$ from table 3.1 is (0.125-0.28)π. The similarity measures of $\tilde{F}_{1,DCT}(z)$ and $\tilde{F}_{1,WT}(z)$ are 0.94, 0.94, 0.89 and 0.9 for Haar, DB4, B93 and B97 kernels respectively. The average value over the four kernels is 0.92. This means that $\tilde{F}_{1,DCT}(z)$ and $\tilde{F}_{1,WT}(z)$ are very similar.

For *i*=2 and 3, a few DCT filters should be combined to match the WT filters due to their difference in passband width. As the bandpass regions of $\tilde{F}_{2,DCT}(z)$, $\tilde{F}_{3,DCT}(z)$ and $\tilde{F}_{4,DCT}(z)$ overlap with those of $\tilde{F}_{2,WT}(z)$, all combinations of these three BDCT filters are examined with reference to the similarity measures. In results, the resultant filter composing of $\tilde{F}_{2,DCT}(z)$ and $\tilde{F}_{4,DCT}(z)$ shows a good match to $\tilde{F}_{2,WT}(z)$ in terms of both passband region and filter similarity. The passband region of this resultant BDCT filter is (0.18-0.64)π and is comparable to the average passband (0.25-0.57)π from table 3.1. The similarity measures of these two filters are 0.97, 0.96, 0.96 and 0.95 for Haar, DB4, B93 and B97 kernels respectively, which gives an average of 0.95. As compared with the passband region of $\tilde{F}_{2,WT}(z)$ in table 3.1, the similar passband region and high filter similarity indicate similar spatial-spectral contents are preserved. For *i*=3, the passband region of $\tilde{F}_{3,WT}(z)$ overlaps with that of $\tilde{F}_{4,DCT}(z)$, $\tilde{F}_{5,DCT}(z)$, $\tilde{F}_{6,DCT}(z)$, and $\tilde{F}_{7,DCT}(z)$. Thus, all combinations of the four BDCT filters are compared with $\tilde{F}_{3,WT}(z)$ in terms of bandpass regions and similarity measures. In our measures, $\tilde{F}_{3,WT}(z)$ shows a good match to the resultant filter from $\tilde{F}_{4,DCT}(z)$, $\tilde{F}_{5,DCT}(z)$, $\tilde{F}_{6,DCT}(z)$, and $\tilde{F}_{7,DCT}(z)$. The passband region from this resultant BDCT filters is (0.46-1)π which is comparable to that of the WT filter (Table 3.1). The similarity values are 0.98 for Haar, 0.99 for DB4, 0.98 for B93 and 0.98 for B97. The average similarity over the four kernels is 0.99. The comparison results are summarized in Table 3.2.

i	Name	Resultant DCT filter	Average similarity with $\tilde{F}_{i,WT}(z)$	Average passband region of $\tilde{F}_{i,WT}(z)$
0	Lowpass	$\tilde{F}_{0,DCT}(z)$	0.97	(0-0.125)π
1	Bandpass 1	$\tilde{F}_{1,DCT}(z)$	0.92	(0.09-0.26)π
2	Bandpass 2	$\tilde{F}_{2,DCT}(z)+\tilde{F}_{4,DCT}(z)$	0.95	(0.18-0.64)π
3	Bandpass 3	$\tilde{F}_{4,DCT}(z)+\tilde{F}_{5,DCT}(z)$ $+\tilde{F}_{6,DCT}(z)+\tilde{F}_{7,DCT}(z)$	0.99	(0.46-1)π

Table 3.2: Summary of resultant BDCT filters in term of similarity with $\tilde{F}_{i,WT}(z)$ and passband region.

3.6.3 Energy spectrum

Though $\tilde{F}_{2,WT}(z)$ and $\tilde{F}_{3,WT}(z)$ show a good match to the resultant filters from { $\tilde{F}_{2,DCT}(z)$ and $\tilde{F}_{4,DCT}(z)$ } and { $\tilde{F}_{4,DCT}(z)$, $\tilde{F}_{5,DCT}(z)$, $\tilde{F}_{6,DCT}(z)$ and $\tilde{F}_{7,DCT}(z)$ } respectively, the input energy preserved by them could be quite different. To investigate the difference in their energy spectra, one can use the spectra difference defined as,

$$D_{A,B}=\frac{\sum_{k=0}^{N-1}\left|\left(|A(k)|/\max(|A(k)|)\right)^2-\left(|B(k)|/\max(|B(k)|)\right)^2\right|}{\sum_{k=0}^{N-1}\left(|A(k)|/\max(|A(k)|)\right)^2}\times 100\% \qquad (3.33)$$

for functions *A* and *B*. The measure $D_{A,B}$ evaluates the energy difference preserved by the BDCT and the four WT kernels. A small $D_{A,B}$ is desirable because similar input energy is preserved. The percentage difference of lowpass, bandpass 1, bandpass 2 and bandpass 3 between resultant BDCT and WT filters are listed in table 3.3. The average percentage difference between lowpass ($\tilde{F}_{0,DCT}(z)$, $\tilde{F}_{0,WT}(z)$), bandpass 1 ($\tilde{F}_{1,DCT}(z)$, $\tilde{F}_{1,WT}(z)$), bandpass 2 ($\tilde{F}_{2,DCT}(z)$ + $\tilde{F}_{4,DCT}(z)$ and $\tilde{F}_{2,WT}(z)$) and bandpass 3 ($\tilde{F}_{4,DCT}(z) + \tilde{F}_{5,DCT}(z) + \tilde{F}_{6,DCT}(z) + \tilde{F}_{7,DCT}(z)$, $\tilde{F}_{3,WT}(z)$) are 0.20%, 0.26%, 0.17% and 0.13% respectively. The small percentage difference implies that the energy preserved by the resultant BDCT and WT filters are very similar.

	Haar	DB4	B93	B97
Lowpass	0.00	0.17	0.42	0.19
Bandpass 1	0.23	0.18	0.35	0.27
Bandpass 2	0.13	0.14	0.23	0.19
Bandpass 3	0.16	0.11	0.17	0.08

Table 3.3: Percentage error of energy spectra between resultant BDCT filters and WT filters for different wavelet kernels

3.6.4 Selection of resultant BDCT and WT filters

Owing to the difference in the passband regions between BDCT and WT as shown in subsection 3.6.1, several BDCT filters are combined to compare with the WT filters. In order to make the passband region of BDCT filters to be comparable to that of the WT filters, possible combinations of BDCT filters is examined in section 3.6.2. A similarity measure is used to find the most similar magnitude spectra between the resultant BDCT and WT filters. Finally, the difference between their energy spectra is evaluated in section 3.6.3. The small percentage difference between their energy preservation implies that they preserve common input characteristics. Results agree with our findings in section 3.6.2. Therefore, BDCT filters are combined to be comparable to WT filters for feature extraction in their transformed domains. The combination of BDCT filters is summarized in Table 3.4. To visualize the concatenation, figure 3.9 shows the 1D concatenation of BDCT filters to compare with WT filters.

Name	Resultant BDCT filter	Comparable WT filter
Lowpass	$\tilde{F}_{0,DCT}(z)$	$\tilde{F}_{0,WT}(z)$
Bandpass 1	$\tilde{F}_{1,DCT}(z)$	$\tilde{F}_{1,WT}(z)$
Bandpass 2	$\tilde{F}_{2,DCT}(z)+\tilde{F}_{4,DCT}(z)$	$\tilde{F}_{2,WT}(z)$
Bandpass 3	$\tilde{F}_{4,DCT}(z)+\tilde{F}_{5,DCT}(z)+\tilde{F}_{6,DCT}(z)+\tilde{F}_{7,DCT}(z)$	$\tilde{F}_{3,WT}(z)$

Table 3.4: Resultant combination of BDCT filters that will be compared with WT filters.

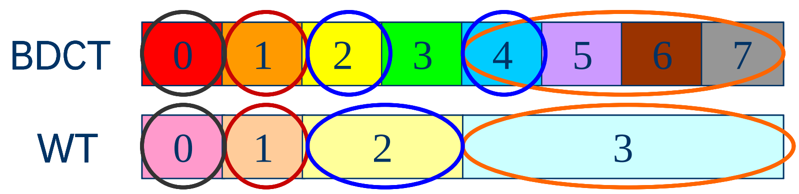


Figure 3.9: 1D concatenation of BDCT filters.

3.7 Chapter summary

The origins and the design of Block-based Discrete Cosine Transform (BDCT) and Wavelet Transform (WT) filters are quite different. Direct comparison on BDCT and WT outputs shows that their outputs are entirely dissimilar. However, our mathematical formulation shows that both can be studied through a subband filtering model, i.e., both consist of lowpass/highpass filtering followed by a downsampling operation.

JPEG2000 allows high flexibility on the usage of wavelet kernels. In our studies, four commonly used wavelet kernels (i.e., Haar, Daubechies 3, Biorthogonal 93 and Biorthogonal 97) are considered to complete our investigation. The lowpass filtering operation in WT is very similar to that of BDCT and they are indeed the same when Haar wavelet is used. Owing to the difference between passband regions in the bandpass filtering operation, direct comparison between bandpass WT and BDCT filters is inappropriate. In contrast, after concatenating several BDCT filters, the resultant BDCT filters can be very similar to the WT filters. Their resultant filtering models can be divided into four subband filters, namely lowpass, bandpass 1, bandpass 2 and bandpass 3.

We measure the resultant BDCT and WT filters in three aspects, i.e. passband regions, similarities between magnitude spectra and difference in energy preservation. It is proven that the resultant BDCT and WT filters (Table 3.4) are alike in terms of magnitude spectra and energy preservation. Therefore, similarity exists between the resultant BDCT and WT filtering outputs. To verify this, comparison on their uncompressed as well as compressed output signals will be done in the next chapter.

Chapter 4

Common feature analysis in JPEG and JPEG2000

4.1 Introduction

Images are stored in multimedia system in various forms, such as JPEG and JPEG2000 compressed formats. It is desirable that image indexing can be carried out in different compressed domains after partial decoding on the compressed data instead of full decompression. After partial decoding, image indexing is expected to be carried out on BDCT and WT transformed domains in JPEG and JPEG2000 respectively. Since BDCT and WT share common subband filtering models as derived in chapter 3, direct image indexing in the two transformed domains becomes possible as they bear similar filter characteristics. Similarity measure between their filtering operations found that at lowpass and lower bandpass frequency range (bandpass 1), similar filtering formulation is obtained such that comparison can be done directly on BDCT and WT filters. However, at higher frequency ranges, due to the large difference between the passband regions of a single BDCT and WT filters, more than one BDCT filters are combined to produce resultant BDCT filters which are comparable to WT filters. From the comparison of filtering characteristics in chapter 3, resultant BDCT and WT filters are found to be similar in three main filter characteristics, i.e. passband region, magnitude spectrum and energy preservation. As the filtering operations between BDCT and WT are similar, their outputs can also similar. If the outputs share common characteristics, then direct image indexing can be carried out at the two compressed domains without incurring full decompression. The outputs will be evaluated in the following sections.

In this chapter, focus is placed on output signal analysis of BDCT and WT filters. Before investigating 67 real 2D images, 1D signal extracted from Lena image is analyzed first. Lena is chosen as our pilot test image because it contains edges, texture and plain areas which are the main components of most images. Its output magnitude spectra and energy spectra in BDCT and WT domains will be studied. Then the 67 images from four different kinds of images are compressed by JPEG and JPEG2000 at different compression ratios. Comparison will be done on their compressed outputs to measure the output similarity between the resultant BDCT filter in JPEG and WT filter in JPEG2000. It is expected that if similarity exists between output signals from BDCT and WT, then common features can be extracted from the two compressed domains for image indexing.

4.2 1D output signal analysis in spectral domain of JPEG and JPEG2000

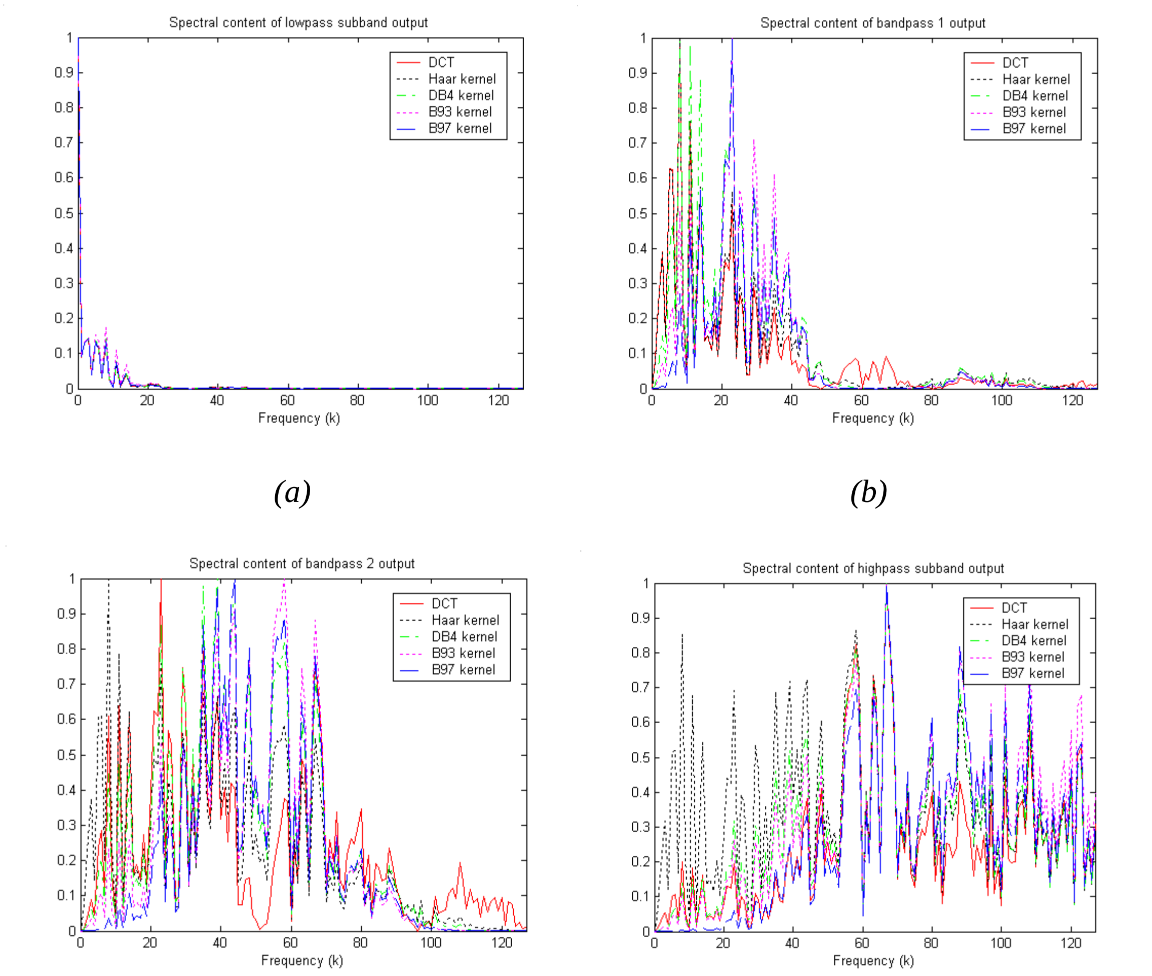
In chapter 3, it is proven that high similarity exists between the filtering operations in BDCT and WT. This implies that their filter outputs should be similar. In this section, we investigate if this reasoning holds. Firstly, the Lena image is used as our test image. The advantage of using the Lena image is that Lena contains plain areas, textures and edges which are common features of most of the images (Figure 4.1). In the following, experimental studies will be carried out on this 1D signal analysis in the spatial-spectral domain.

The resultant lowpass and bandpass filters in BDCT and WT are proven to be similar to each other theoretically in chapter 3. Outputs from them are further investigated to prove our derived subband filtering model. In this part, we will practically analyze the similarity between the resultant outputs $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$, which are the outputs from the resultant $\tilde{F}_{u,DCT}(z)$ and $\tilde{F}_{i,WT}(z)$ respectively. Using the rows of the Lena image as our 1D input signal, $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$ can be obtained by Eqn. 3.26 and Eqn. 3.28 respectively. 1D BDCT subband outputs are concatenated to compare with WT subband outputs as shown in Fig. 3.9. As similar filtering operations are performed, similar outputs should be obtained.

The 1D Lena image passes through the 1D resultant BDCT and WT filtering operations including the four wavelet kernels, i.e., Haar, DB4, B93 and B97 kernels. Outputs generated by the four sets of wavelet kernels should be different from one another. Therefore, the magnitude spectra from $\tilde{X}_{u,DCT}(z)$ should be different from each of the $\tilde{X}_{i,WT}(z)$ generated by the four wavelet kernels. Figure 4.2 shows the plots of $\left|\tilde{X}_{u,DCT}(z)\right|$ and $\left|\tilde{X}_{i,WT}(z)\right|$ using the four wavelet kernels with $z = e^{j\frac{2\pi k}{N}}$ and $k \in [0,127]$ to observe their frequency responses from 0 to π. Due to the variation of the magnitude between $\left|\tilde{X}_{u,DCT}(z)\right|$ and $\left|\tilde{X}_{i,WT}(z)\right|$, comparing their maximum and minimum magnitude cannot accurately reveal their spectra commonality. Thus, all plots are done based on the normalized magnitude spectrum. This makes a fair comparison since the magnitudes of all spectra lie within 0 and 1 at all frequency.

Note that the use of a specific wavelet kernel will not only affect filter length, but also the magnitudes of the subband wavelet coefficients. For example, consider the lowpass case, where the length of Haar is two and involves only 2 pixels in transformation but B97 involve 9 pixels. This gives quite different magnitudes in the transformed coefficients. The variation of the magnitude spectra provides important information about the preserved spectral content in the subbands. From the spectral contents, we can compare the information that is preserved by each kernel. Looking at the output spectra of all subbands in Figure 4.2, we observe that the output spectral content from the filtering operation between resultant BDCT and WT are in fact very similar. This implies that similar input characteristics are preserved by the resultant BDCT and WT filters.

Figure 4.1: Lena image contains plain areas, textures and edges.



(a) *(b)* *(c)* *(d)*

Figure 4.2: Plots of the magnitude spectra of $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$ with four different kernels. They are (a) lowpass, (b) bandpass 1, (c) bandpass 2 and (d) highpass outputs.

4.3 Spectral analysis on images

In section 4.2, spectral analysis is done by considering a 1D signal analysis on the Lena image. In this section, studies will be done on four different kinds of JPEG and JPEG2000 compressed images. Images are compressed and then partially decoded such that their BDCT and WT coefficients can be extracted for further comparison. These coefficients are indeed outputs of the resultant BDCT and WT filters which reveals spatial-spectral information. In order to measure the similarities between BDCT and WT outputs, experiments are done on their transformed domains. The aim of measuring the similarity between the two outputs generated by the two compression schemes is to verify whether common features can be extracted from the two schemes for image indexing after compression.

Experiments show that the output signals from the resultant BDCT and WT filters share great similarity. However, different images have varying spectral contents. For example, the spectral contents of natural images show importance in the low frequencies. Whereas the spectral content of texture images concentrate in the higher frequency range. Thus, difference in filters could be emphasized or de-emphasized when applied to images. Therefore, a similarity study will be carried out on various images. There are four kinds of test images as shown in Figure 4.3: 29 natural scenes, 11 structured texture patterns, 10 random texture patterns, and 17 man-made (artificial) object images. These four types of images are selected because each of them bears different frequency contents like plain areas in natural images, variation in textures and sharp edges in man-made objects.

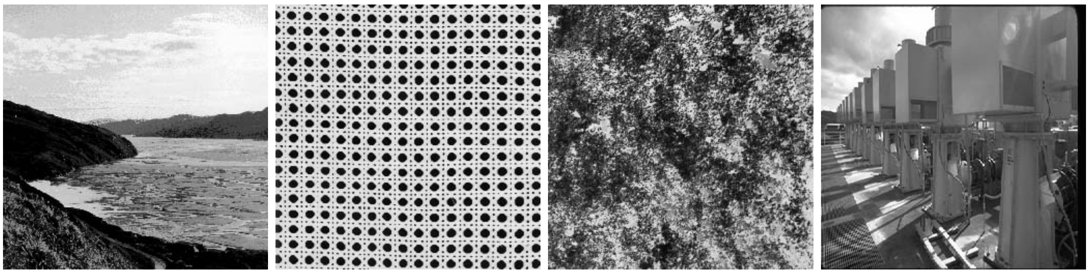
We consider the following four wavelet kernels again: Haar, DB4, B93 and B97. Average similarity is calculated on the four 2D image kinds (i.e., 67 images) to find out how similar the outputs, $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$, are. Table 4.1 summarizes the results.

	$\tilde{X}_{0,DCT}(z)$ vs $\tilde{X}_{0,WT}(z)$	$\tilde{X}_{1,DCT}(z)$ vs $\tilde{X}_{1,WT}(z)$	$\tilde{X}_{2,DCT}(z)+\tilde{X}_{4,DCT}(z)$ vs $\tilde{X}_{2,WT}(z)$	$\tilde{X}_{4,DCT}(z)+\tilde{X}_{5,DCT}(z)$ $+\tilde{X}_{6,DCT}(z)+\tilde{X}_{7,DCT}(z)$ vs $\tilde{X}_{3,WT}(z)$
Haar	1.0000	0.9793	0.9652	0.9077
DB4	0.9609	0.9392	0.9754	0.9705
B93	0.9515	0.8621	0.9244	0.9651
B97	0.9746	0.8461	0.9175	0.9896

Table 4.1: Average similarities between the BDCT and WT subband outputs for four different kernels.

From the similarity values shown in Table 4.1, most of the output images show a large similarity between the BDCT and WT. The high similarity values imply that the output spectral contents from the resultant BDCT filters are indeed very similar to that from WT filters. The only exception is for B93 and B97 in $\tilde{X}_{1,DCT}(z)$ and $\tilde{X}_{1,WT}(z)$. This is because the passband regions of $\tilde{F}_{1,DCT}(z)$ and $\tilde{F}_{1,WT}(z)$ are (0.09-0.26)π and in average (0.155-0.28)π respectively. The passband regsions are quite different near the low-frequency end. This thus amplifies the differences in the spatial-spectral information extracted by the BDCT and WT filters because many spectral characteristics are at the lower frequency end. Though outputs $\tilde{X}_{1,DCT}(z)$ and $\tilde{X}_{1,WT}(z)$ show small variations due to the differences in passband regions, the similarity of all the other outputs are close to 1.

The high similarity value supports our findings in section 4.2 that very similar spectral contents are preserved by the BDCT and WT filters. As a matter of fact, common features should be extractable from the BDCT and WT subband outputs for direct comparison. The energy preserved by them will be investigated.



(a) *(b)* *(c)* *(d)*

Figure 4.3: Sample images from (a) natural scene, (b) structured texture pattern, (c) random texture pattern, and (d) man-made object image.

4.4 Energy preservation

In section 4.3, the similarities of magnitude spectra between JPEG and JPEG2000 output signals are investigated. As a result of the passband difference in the resultant BDCT and WT filters, the output spectra of JPEG and JPEG2000 are also different. Besides magnitude spectrum, energy preservation is another important characteristic. If similar energy is preserved by the filters, common spectral content is preserved as well. In this section, the percentage difference of output energy preserved by the resultant BDCT and WT filters is studied. Calculation is similar to those in section 3.6.3. With reference to Eqn. 3.33, $A(k)$ and $B(k)$ are now the output energy preserved by filters for JPEG and JPEG2000. The average percentage difference $D_{A,B}$ of the subband filtering outputs of the 67 sample images is summarized in Table 4.2.

	Haar	DB4	B93	B97
Lowpass	0.00	0.04	0.22	0.04
Bandpass 1	0.09	0.23	0.43	0.37
Bandpass 2	0.09	0.08	0.17	0.16
Bandpass 3	0.19	0.10	0.11	0.06

Table 4.2: Percentage difference of energy preservation between resultant DCT filters and WT filters with different wavelet kernels.

Comparing Table 4.2 and Table 3.3, the energy difference between the outputs is much smaller than that of the filters besides bandpass 1. This may be due to the fact that spectral contents of the original image at a certain frequency range are very small, especially at those bandpass regions. This results in a small overall difference in the outputs. Also, it is obvious that a larger percentage difference exists in the bandpass 1 filtering operation. This is consistent with our findings in the passband difference between BDCT and WT filters as in Table 3.3. Table 4.2 shows a comparably larger percentage difference for the usage of the B93 wavelet kernel. The results agree with our investigation in section 4.3 that the smaller similarity of magnitude spectrum for B93 wavelet kernel generates larger percentage difference in energy preservation. Though a slightly larger percentage difference is obtained for the B93 kernel, the percentage differences for all kernels are indeed very small as all are less than 1%. Thus, the energy preserved by the resultant BDCT and WT filters at all passband are actually very similar.

4.5 Effect of compression

Compression is to preserve important information but discard unimportant fine details such that the file storage size can be reduced. Hence, the appearance of objects is preserved whereas texture and edges are blurred after compression. As a matter of fact, compression could affect the availability of features for direct feature extraction. Also, JPEG and JPEG2000 compression schemes employ different transformation (i.e., BDCT and WT respectively) and quantization schemes. Compression will certainly affect features remaining in the compressed images at high compression ratio even though similar energy spectrum are proven to be preserved in BDCT and WT domains. This section will study the similarity of the magnitude spectra between JPEG and JPEG2000 images under different compression ratios. Image indexing at different compression domains becomes possible only when a high similarity exists between their compressed outputs at high compression ratio.

4.5.1 Similarity analysis on different nature of compressed images

According to section 4.3, different image natures have dissimilar information contents, which are directly reflected in their frequency spectra. The similarity of information preserved by BDCT and WT filters indicate the amount of common features preserved by them. If a high similarity value is maintained at their filter outputs, indexing in different transformed domains becomes possible. To measure the similarity between JPEG and JPEG2000 compressed images, the 67 images from four kinds of images are compressed at six ratios: 6:1, 12:1, 18:1, 24:1, 32:1 and 44:1. Similarity will be measured on the filtering subband outputs of resultant BDCT and WT, i.e. lowpass, bandpass 1, bandpass

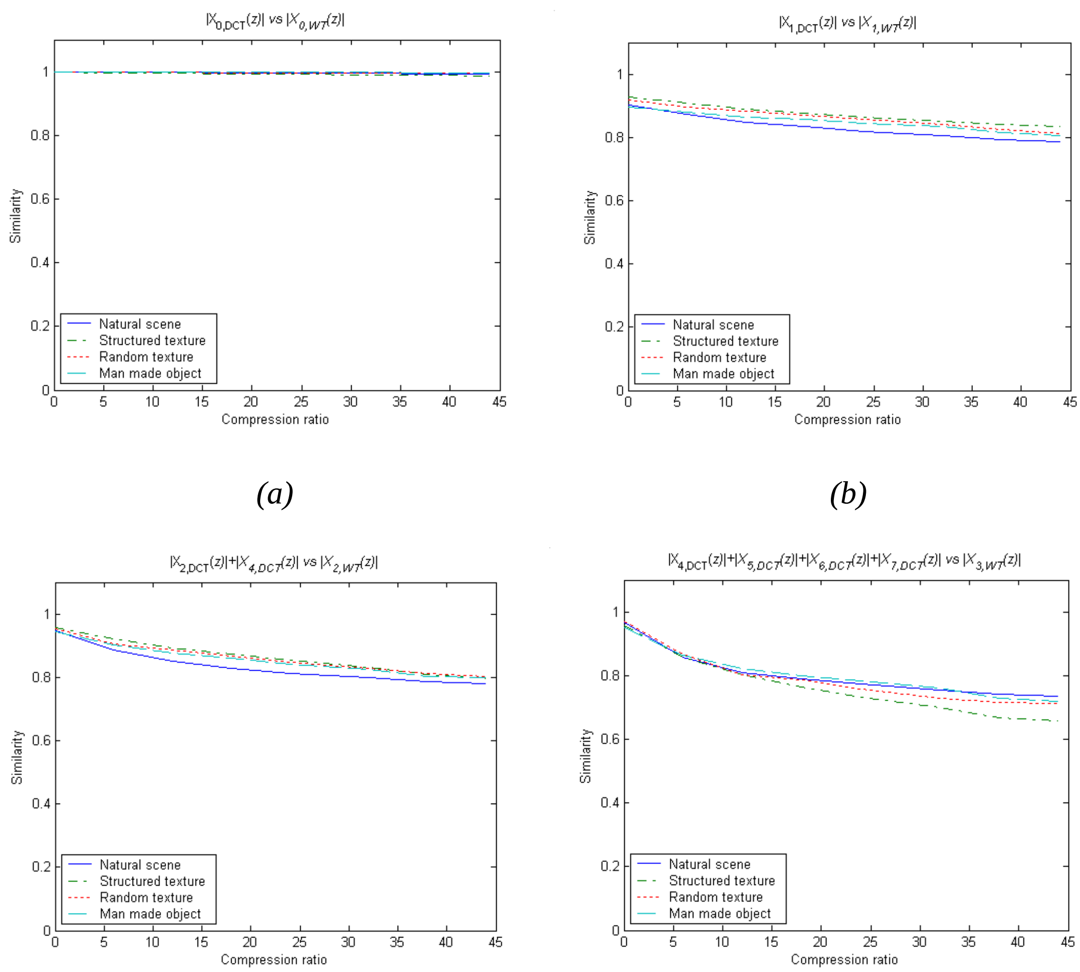
2 and bandpass 3 respectively. The similarity between magnitude spectra of resultant $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$ with different wavelet kernels (i.e., Haar, DB4, B93 and B97 kernels) versus compression ratio is plotted in Figure 4.4. Each figure plots the average similarity obtaining from all wavelet kernels at each subband.

Figure 4.4(a) shows that the similarity measure of compressed lowpass images from BDCT and WT is almost equal to 1. This means that the DC images match the lowpass images to a very high degree. This result matches our findings in section 3.6.2 and section 4.3. It shows that very high similarity exists not only between the filters $\tilde{F}_{0,DCT}(z)$ and $\tilde{F}_{0,WT}(z)$, but also between their output signals $\tilde{X}_{0,DCT}(z)$ and $\tilde{X}_{0,WT}(z)$. Therefore, despite high compression, the output similarity can still be high due to the large similarity between the BDCT DC and WT lowpass filters even when different wavelet kernels are used for compression.

Figure 4.4(b) indicates the comparison between BDCT and WT outputs at bandpass 1 region. We notice that even at low compression ratio, the similarity values for all wavelet kernels are no longer close to 1. This is due to the passband difference (section 3.6.1) and filter difference (section 3.6.2) between $\tilde{F}_{1,DCT}(z)$ and $\tilde{F}_{1,WT}(z)$. The comparably low output similarity also supports our findings in section 4.3 and 4.4 that the output spectra similarity greatly depends on filter similarity. It is obvious that the similarity of natural scene is the lowest among the other image kinds when compression ratio is larger than 4. This is because the natural scene contains mostly low frequency information but rare in high frequency content. Therefore, the spectral contents of natural scene are more significant in the lowpass band than in the bandpass region. For other kinds of images, i.e., textures and man-made objects, more significant spectral information is provided in the high frequency passband. In the compression process, spectral content in bandpass region will be discarded but lowpass spectral information is preserved. However, due to different coding schemes of JPEG and JPEG2000, the degree of ignorance of high frequency information is different for the two schemes even at the same compression ratio. JPEG2000 retains more high frequency information than JPEG at the same compression ratio, the different degree of ignorance of bandpass information in compression results in lower similarity value as well. Thus, the decline of similarity values between JPEG and JPEG2000 verse compression ratio is emphasized during the compression process.

As images contain less and less information at higher frequency bands such as in bandpass 2 and bandpass 3, fewer features are available for comparison. Based on the same analysis of image spectral contents in bandpass 1, the few but different spectral characteristics in the highpass subband result in low similarity value. Therefore, the overall similarity value of bandpass 3 is lower than that of bandpass 2. But we observe that at low compression ratios, the similarity value of bandpass 2 and bandpass 3 is higher than that of bandpass 1 among all the four wavelet kernels. This is due to the closer of the passband regions (section 3.6.1) and higher filter similarity (section 3.6.2) between resultant BDCT and WT filters in bandpass 2 and bandpass 3. Regarding bandpass 2, the filter output similarity of natural scene is lower than that of the other three image kinds as plotted in Figure 4.4(c). This is because natural images have low frequency information content, and the small amount of high frequency information is discarded by JPEG and JPEG2000, resulting in dissimilar subband output.

There is not much spectral information available at bandpass 3, no matter whether images are compressed at high compression ratios. Therefore, the similarity value of natural images is no longer the lowest in bandpass 3. On the contrary, for other image kinds such as texture, some significant characteristics are preserved in this bandpass. However, due to the different coding algorithm of JPEG and JPEG2000, the high frequency spectral characteristics are discarded to different extent. The ignorance of the high frequency characteristics causes low similarity value of texture images in bandpass 3 as illustrated in figure 4.4(d).



(a) *(b)* *(c)* *(d)*

Figure 4.4: Similarities in the spatial domain between the BDCT and WT subband images, (a) lowpass, (b) bandpass 1, (c) bandpass 2 and (d) bandpass 3 images, at different compression ratios.

4.5.2 Compression effects on images

Average similarity of the four image kinds versus compression ratio will be studied in this subsection. The average is obtained by averaging the similarity values from section 4.5.1. Figure 4.5 plots the average similarity measure over the 67 images from four image kinds at lowpass, bandpass 1, bandpass 2 and bandpass 3.

We can see that the similarity between $\tilde{X}_{0,DCT}(z)$ and $\tilde{X}_{0,WT}(z)$ continues at a very high value even under a high compression ratio. This means that very similar spectral information is preserved by BDCT and WT filters and can be extracted from this lowpass band. It is well known that most of the image information is concentrated at this lowpass subband. $\tilde{X}_{0,DCT}(z)$ and $\tilde{X}_{0,WT}(z)$ are thus ideal for image indexing since spectral characteristics are preserved in this subband even under high compression ratios. Therefore, retrieval performance would not be greatly affected by compression.

The similarity values of all of the three higher subbands under all compression ratios are lower than that of the lowpass subband. The similarity value of the $\tilde{X}_{1,DCT}(z)$ and $\tilde{X}_{1,WT}(z)$ is obviously the lowest when compression ratio is less than 5. This matches our findings in section 3.6.1 where the passband regions of filters $\tilde{F}_{1,DCT}(z)$ and $\tilde{F}_{1,WT}(z)$ do not show a good correspondence. As for the other two higher bandpass subbands, i.e., bandpass 2 and bandpass 3, similarity values are higher at low compression ratios. However, their values decrease rapidly with increases in compression ratio. This is due to the difference between the JPEG and JPEG2000 coding algorithms. It is well known that JPEG2000 retains more high frequency information than JPEG under the same compression ratio. Thus, spectral information that is retained in JPEG2000 cound be omitted by JPEG during compression. The difference in the spectral characteristics is thus revealed in their output similarity measure. As a result, the output difference between BDCT and WT highpass subband is emphasized in large compression ratio cases. This gives a rapidly declining similarity curve of higher subbands.

From our findings, we can see that the lowpass subbands provide the highest similarity compared with all the other subbands. As the spectral characteristics dominate in the lowpass, bandpass 1, bandpass 2 and bandpass 3 in descending order, more emphasis can be put on the lowpass subband and then the other bandpass subbands correspondingly in image indexing. The bandpass subbands can be regarded as an improvement during image indexing. Thus, the lower similarity between highpass subbands such as bandpass 1 would not affect the indexing results significantly. Subband selection and feature extraction from subbands for the purpose of image indexing will be discussed further in chapter 5.

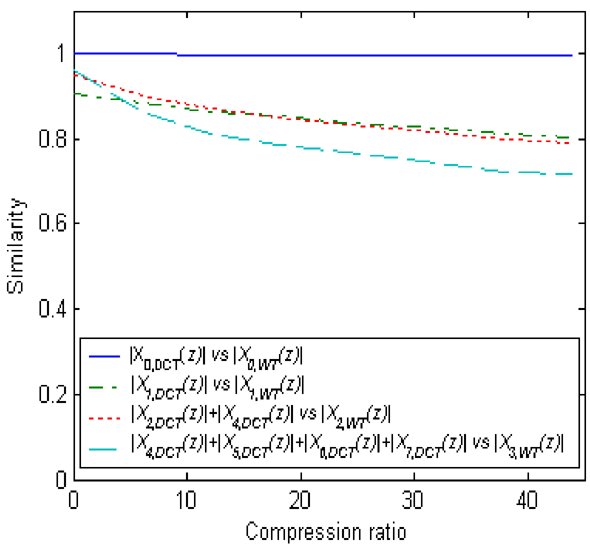


Figure 4.5: Similarity measure between the lowpass, bandpass 1, bandpass 2 and bandpass 3 subband images versus compression ratio.

4.6 Chapter summary

In this chapter, we analyze the common image characteristics preserved by JPEG and JPEG2000 with and without compression to validate the subband filtering models of Block-based Discrete Cosine Transform (BDCT) and Wavelet Transform (WT). First of all, the Lena image without compression is used to examine whether similar spectral information is preserved by the resultant BDCT and WT filters from JPEG and JPEG2000 respectively. It is found that similar spectral details are retained by the filters at all subbands. This shows that common image characteristics are preserved by the resultant BDCT and WT filters.

To confirm our findings, experiments are done on sixty-seven JPEG and JPEG2000 compressed images obtained from four image kinds. Our experiments prove that the output signals from the BDCT and WT filters are very similar in two aspects, namely spectral characteristics and energy preservation in four filter passband regions: lowpass, bandpass 1, bandpass 2 and bandpass 3. Simulation results show that common spectral characteristics and energy contents are found between all BDCT and WT subband outputs. In other words, very similar input characteristics are preserved by both BDCT and WT filtering operations.

The sixty-seven images are also compressed to six different compression levels ranging from 6:1 to 44:1. We find that compression has very small effect on the lowpass output similarities even under high compression ratio. The similarity values between all BDCT and WT subband outputs are close to 1 under high compression. This implies that they are indeed very similar. However, since JPEG2000 retains more high frequency information than JPEG even at the same compression ratio, their similarities at highpass subband reduce as compression ratio increases. Owing to the difference in the passband regions between resultant BDCT and WT filters, their bandpass outputs are no longer highly similar to each other. As a result, their similarities decline at high compression ratios. Though not very similar outputs are obtained from the resultant BDCT and WT filters, the similarity values are all over 0.7 which means they are indeed quite similar.

As the filtering operations of JPEG and JPEG2000 preserve similar spectral details, common image characteristics are available even under high compression. Direct image indexing may be performed to retrieve images from JPEG and JPEG2000 multimedia systems. To evaluate this, an image indexing system is established. Simulations will be done to estimate the retrieval performance in the two compressed domains. Details of the image indexing system will be discussed in depth in the next chapter where experimental results will also be provided.

Chapter 5

Feature extraction in JPEG and JPEG2000 domains

5.1 Introduction

In chapter 3, we have derived a subband filtering model relating BDCT and WT transformation. With the help of this subband model, we have shown that filters in BDCT and WT are very similar in three aspects, namely passband regions, filter similarity and energy preservation. In chapter 4, the outputs of the filters are compared to investigate whether similarity still hold at high compression ratios. It is found that similarity between JPEG and JPEG2000 subband outputs reduces as the compression ratio increases. This is due to the difference of compression algorithms between JPEG and JPEG2000 in that JPEG2000 retains more high frequency information than JPEG. Though similarity decreases with an increase in compression ratios, it is confirmed that similar spectral information is preserved at JPEG and JPEG2000 subband outputs even under high compression ratios. As similar information is preserved in both domains, image indexing in both domains become possible.

To investigate direct image indexing in different compressed domains, image features are extracted from JPEG and JPEG2000 subband outputs for simulation. As BDCT coefficients in JPEG are organized in BDCT blocks which are different from WT organization, the BDCT coefficients are rearranged with subband first. In fact, this re-organization is included in our subband feature model. After coefficient rearrangement, both BDCT and WT subband coefficients are in similar spatial-spectral structure such that features can be extracted from subbands directly. Energy, significance map and normalized central moment are extracted from these subbands as their features. Though features can be extracted from all BDCT and WT subbands, certain subbands that have low high frequency information can be neglected. To verify our analysis, a database containing one thousand and eight hundred images from nine image kinds is established to test the indexing results of our proposed retrieval algorithm. Precision and recall are used to evaluate the performance of the indexing results from the two compressed domains under different compression ratios. Details of our analysis and simulation results are discussed in this chapter.

5.2 Rearrangement of the BDCT coefficients into WT subbands structure

BDCT coefficients are organized in 8x8 blocks whereas WT coefficients are divided into subbands according to their spectral contents and decomposition levels. BDCT coefficients must be rearranged in a form similar to the WT subband structure before direct comparison can be done. For example, to construct a lowpass subband from BDCT coefficients similar to that in WT structure, the DC coefficients can be concatenated from all BDCT blocks according to their block order. Therefore, all BDCT coefficients at different block locations can be concatenated together according to their frequency characteristics. Thus, BDCT block coefficients that are corresponding to a particular frequency region can form subbands that are comparable to WT subband structure. This rearrangement makes further feature extraction more efficient because features can be extracted in a common subband form in both BDCT and WT domains.

Assume that the image size is $N \times N$ and the level of wavelet decomposition is three. Table 5.1 summarizes the rearrangement process and indicates the relationship between the concatenated BDCT block and its comparable WT subbands. The values of u and v indicate the coordinates of coefficients in the BDCT block $\tilde{X}_{m,n}(u,v)$ whereas m and n represent the BDCT block location. The coefficients from all BDCT blocks with the same values of u and v are concatenated to form a wavelet subband structure B_j. Therefore,

$$B_j \in \{X_{m,n}(\varepsilon_j) \mid \varepsilon_j \in (u,v), u \in n_{u_j}, v \in n_{v_j}\} \qquad (5.1)$$

where ε_j defines the coefficient coordinates, n_{u_j} and n_{v_j} are set of coordinate *(u,v)* defined in table 5.1. This re-organization is in fact included in our subband feature model.

For easier reference to the concatenated BDCT and WT subbands in latter sections, SB_j is a label given to the subbands for $j=\{0,1,2,\ldots,9\}$.

j	Subband name	Concatenated BDCT subband	Comparable WT subband	n_{u_j}	n_{v_j}
0	SB_0	B_0	LL	0	0
1	SB_1	B_1	LH_3	0	1
2	SB_2	B_2	HL_3	1	0
3	SB_3	B_3	HH_3	1	1
4	SB_4	B_4	LH_2	0,1	2,4
5	SB_5	B_5	HL_2	2,4	0,1
6	SB_6	B_6	HH_2	2,4	2,4
7	SB_7	B_7	LH_1	0,1,2,3	4,5,6,7
8	SB_8	B_8	HL_1	4,5,6,7	0,1,2,3
9	SB_9	B_9	HH_1	4,5,6,7	4,5,6,7

Table 5.1: Rearrangement of BDCT coefficients to form wavelet subband structure.

Example in Figure 5.1 shows that rearranging BDCT coefficients into a wavelet subband structure allows a direct comparison with WT subband.

Figure 5.1: BDCT coefficients rearranged into a wavelet subband structure

5.3 Feature extraction

After the rearrangement of BDCT coefficients, the structure of the resultant BDCT and WT subbands are in similar form such that features can be extracted directly from them for further comparison. Features are extracted from the BDCT and WT coefficients that are specified in table 5.1. As our re-arrangement is only a reference for our features extraction on the related BDCT and WT coefficients, no extra subbands are actually formed from the re-arrangement. Thus, no extra computation and storage are needed to reorder and combine our concatenated BDCT outputs. Three main features will be extracted from the subbands to characterize their features, they are energy, significance map and normalized central moments.

5.3.1 Energy of subbands

One of the important properties of BDCT and WT is the conservation of energy. Energy of images in the spatial domain is preserved after transformation. Also, it is well-known that a large portion of the energy contributes at the low frequency subband whereas the remaining portion is distributed in the remaining subbands according to their frequency contents. If images contain many edges or complex patterns, then more energy are distributed to the high frequency subbands. Therefore, the total energy in a subband indicates the importance of that subband in an image. Energy E_j can be calculated as the sum of the square of the *j*th subband coefficients, i.e.,

$$E_j = \sum_{x,y} SB_j(x,y)^2 \tag{5.2}$$

where $SB_j(x,y)$ is the coefficient of *j*th BDCT and WT subbands at coordinate (x,y). As significant subband gives higher value of E_j but those insignificant gives smaller, the energy value thus becomes a valuable feature of subband.

Note that energy will not be taken from all subbands to form our feature vector due to the distribution of spectral properties. The subbands at the lowest decomposition level, i.e., SB_7, SB_8, SB_9 store very high frequency components. However, most images do not have many edges or sharp changes. Thus, there are very few dominant high frequency components in these three subbands. Hence they contain low energy levels. Information in the high frequency subbands become insignificant when compared to that in the low frequency subbands. As a result, comparing the energy values in these subbands becomes meaningless and thus SB_7, SB_8, SB_9 will not be taken into consideration. In contrast, all subbands SB_j for $j=\{0,\ldots,6\}$ contain dominant frequency components and form valuable feature as can be seen in Equation 5.3.

$$f^{E} = \{E_0, E_1, ..., E_6\} \tag{5.3}$$

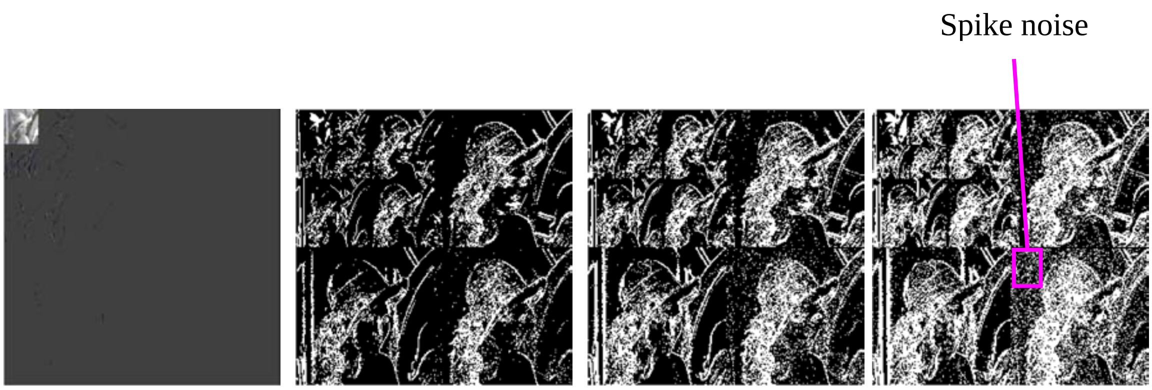
5.3.2 Significance map

As we can observe in Figure 5.1, some coefficients are dominant but some are insignificant. Dominant coefficients are mainly due to sharp changes and edges but do not exist frequently in most images. As dominant coefficients usually represent the overall shape of an image, they should be considered in feature extraction. Therefore, a significance map will be constructed to screen out all insignificant coefficients. To generate the significance maps of subbands, a predefined percentage of dominant coefficients in a subband is considered to be significant and these coefficients are marked as “1”. The other coefficients in the same subband are insignificant and marked as “0”. Thresholding is employed so as to preserve the significant coefficients and set those insignificant to zero. Mathematically, the significance map at coordinate *(x,y)* is defined as

$$M_j(x,y) = \begin{cases} 1 & \text{if } SB_j(x,y) > T_j \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

$SB_j(x,y)$ is the coefficient of *j*th BDCT and WT subbands at *(x,y)*. T_j is the threshold at the *j*th subband such that when $SB_j(x,y)$ is larger than T_j, it is considered as significant and marked as ‘1’. Before T_j can be defined, the acceptable percentage of significant coefficients has to be assumed first.

The significance map provides an outline on objects shape. To define the threshold T_j, the percentage of dominant coefficients should be carefully selected such that it can well describe the significant shapes in the transformed domain. To illustrate the effects of T_j, an example is given in Figure 5.2. Significance maps are generated by retaining the most significant 20%, 30% and 40% of the total coefficients in Figure 5.2(a). As we can see in Figure 5.2(b), when the percentage is small, some edges or sharp change are missed. However, a large value does not always generate proper significance map. Figure 5.2(d) shows an overly complicated significance map which induces too much details, especially in the high frequency subbands such as SB_j for j = {3,6,9}. Figure 5.2(c) shows an appropriate significance map which outlines the texture pattern with comparably low spike noise. When we focus on the SB_j subbands at j = {3,6,9}, Fig. 5.2(c) can outline the meaningful hair and shapes, but Fig. 5.2(b) is unable to outline the long curly hair and Fig. 5.2(d) contains too much spike noise. Experiment with different percentage values is done and shows that significance map with 30% dominant coefficients provides sufficient important information for further feature extraction. Thus, significance map with 30% of most significant coefficients is chosen as the best representative map.



(a) *(b)* *(c)* *(d)*

Figure 5.2: (a) A wavelet transformed Lena image and significance map of Lena image by retaining the most significant (b) 20%, (c) 30% and (d) 40% of the total coefficients.

5.3.3 Moment of significance map

Moment analysis provides a way to measure the center of mass, the biasing point, and the symmetrical property of an image. During retrieval, user may want to search for similar objects which may shift in position. Moments are greatly dependent on the position and orientation of the objects in the image. On the contrary, normalized central moments are calculated with respect to the centre of mass of the image. Therefore, the shift of centroid or objects across the image does not alter the normalized central moments. Also, the normalized central moments can eliminate the effect of scaling and rotation in images. To make use of the properties of shifting and scaling invariant, normalized central moments are used [46,47]. They are defined as,

$$\eta_{pq}^{j} = \frac{u_{pq}^{j}}{(u_{00}^{j})^{\gamma}} \tag{5.5}$$

where p and q define the order of the moments at the jth subband, and

$$u_{pq}^{j} = \sum_{y=1}^{M}\sum_{x=1}^{N}(x-\bar{x})^{p}(y-\bar{y})^{q} M_{j}(x,y) \tag{5.6}$$

$$\gamma = \frac{p+q}{2}+1 \tag{5.7}$$

$$\bar{x}_{j} = \frac{u_{10}^{j}}{u_{00}^{j}}\ ,\ \bar{y}_{j} = \frac{u_{01}^{j}}{u_{00}^{j}} \tag{5.8}$$

After generating the significance map, second and third order normalized central moments are calculated from the significance map in order to find out its variance and skewness in each subband SB_j. Variance indicates the spread of significant coefficients with respect to the centroid along the image. And skewness indicates the symmetry of the distribution, i.e., whether the coefficients are biased at any direction.

In each subband, three values from the second order moment and four values from the third order moment are obtained. The three second order normalized central moments are η_{20}^{j}, η_{11}^{j} and η_{02}^{j} which indicate the variance along the horizontal, diagonal and vertical direction respectively. Whereas, the four third order normalized central moments are η_{30}^{j}, η_{21}^{j}, η_{12}^{j} and η_{03}^{j} which measure the skewness of image along the horizontal, 30 degree, 60 degree and vertical directions respectively. These seven moment values form another feature $f_{SB_j}^{M}$, that is defined as,

$$f_{SB_j}^{M} = \{\eta_{20}^{j}, \eta_{11}^{j}, \eta_{02}^{j}, \eta_{30}^{j}, \eta_{21}^{j}, \eta_{12}^{j}, \eta_{03}^{j}\} \tag{5.9}$$

Regarding the variance and skewness calculation, three second and four third order normalized central moments are extracted from the significance map of each subband. In this case, we can measure the skewness and variance of the vertical, horizontal and diagonal components of the image. Our studies have found that the moment values at all decomposition levels are more or less the same due to the high correlation between frequency components across decomposition levels. Therefore, the moments at the highest decomposition level are already sufficient to characterize the variance and skewness of the significant coefficients. To optimize the size of the feature vector, normalized central moments are evaluated at SB_j for j = {0,…,3} at the highest decomposition level only. A new feature vector f^{M} concerning moments thus becomes

$$f^{M} = \{f_{SB_0}^{M}, f_{SB_1}^{M}, f_{SB_2}^{M}, f_{SB_3}^{M}\} \tag{5.10}$$

5.4 Feature vector

A resultant feature vector *f* is then defined by f^E and f^M as shown in Eqn 5.11.

$$f = \{ f^E, f^M_{SB_0}, f^M_{SB_1}, f^M_{SB_2}, f^M_{SB_3} \} \tag{5.11}$$

As f^E and f^M generate 35 values, the resultant feature vector *f* of 35 dimensions can be constructed to characterize the features of the whole image. The 35 dimensional (35D) feature vector is composed by seven energy values from the SB_j for $j=\{0,\ldots,6\}$ plus three second order and four third order moment values from SB_j for $j=\{0,\ldots,3\}$. This 35D feature vector can well represent the features of images at all horizontal, diagonal and vertical directions. To compare the features between two images, it can be easily be done by measuring the distance between their feature vectors.

5.5 Distance measure

After the characterization of images by their energy and the normalized central moments of the significance maps, we can retrieve images by using these features. To find out the most similar image to a query image in the database, distance measure between the query image and the indexing image is performed. Distance between two images is defined by measuring the vector distance in Euclidean space [10] as stated in Eqn.5.12.

$$D_{p,q} = \sum_{k=1}^{35} \left(f_k^q - f_k^p \right)^2 \tag{5.12}$$

where $D_{p,q}$ is the distance between the query image *q* and indexing image *p*. f_k^q and f_k^p is the k^{th} element in the feature vectors of *q* and *p* respectively.

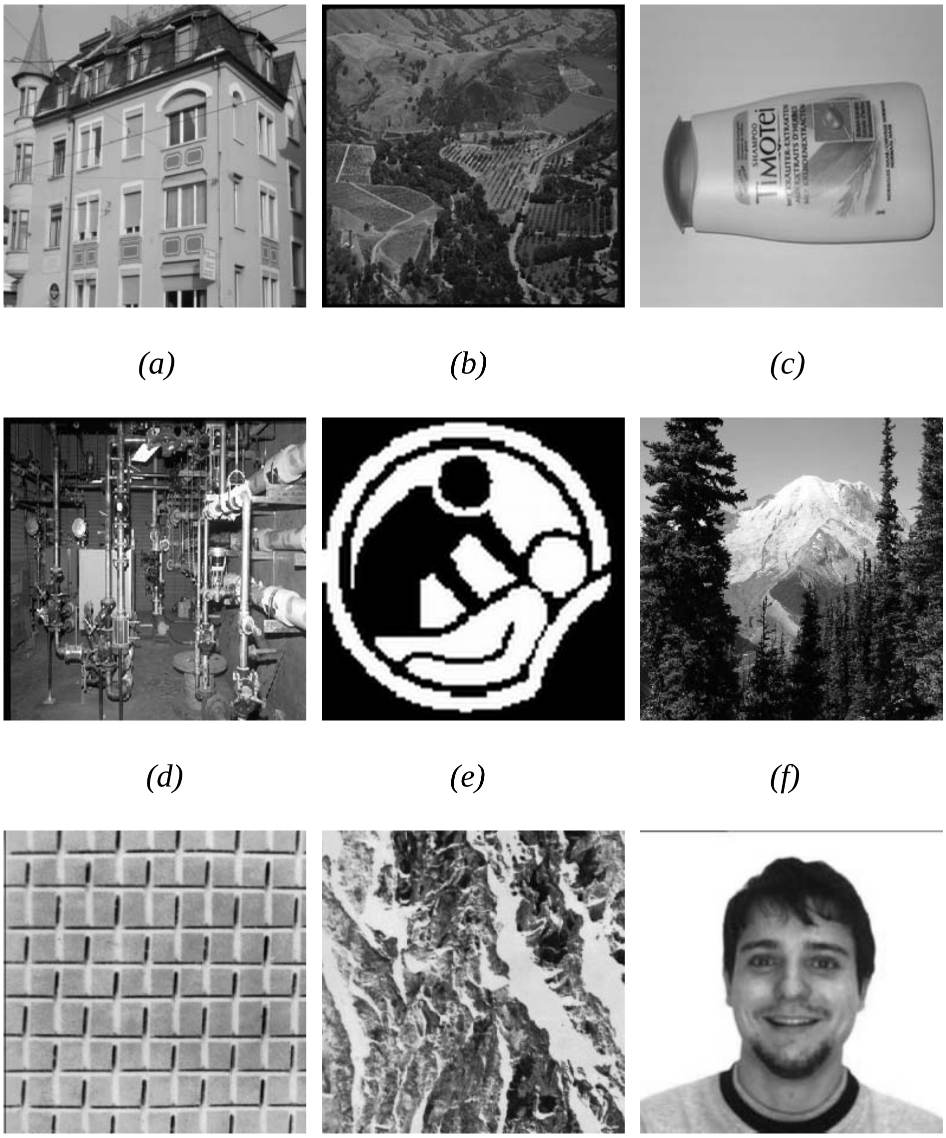
$D_{p,q}$ indicates how similar the images are. A small value indicates that the two images are very close in their feature vectors such that they bear common characteristics. On the contrary, if two images have non-similar characteristics, then the distance $D_{p,q}$ is large.

$D_{p,q}$ is a mathematical way to quantity image similarity. In practice, $D_{p,q}$ is usually non-zero. Thus, $D_{p,q}$ are sorted in ascending order to rank their similarity. Therefore, indexing candidates with shorter distance from the query image are placed in the front position whereas those with longer distance are put near the end of the list. The ranking process not only benefits user to find out the most similar indexing image to the query image, but also provides efficiency to evaluate retrieval performance. After evaluating all the images in the database, the indexing images with the smallest distances between the query images are selected and considered as the best matched images.

5.6 Selection of database

As this study aims to do direct image indexing in JPEG and JPEG2000 compressed domains, comparison between JPEG and JPEG2000 images has to be done. To study the indexing performance on different image kinds, nine image kinds, each with 200 images, are used. The nine image kinds consist of building, landscape, man-made single object, made-made objects, trade-mark, natural image, pattern texture, random texture, and the human face. These combinations include images with different nature, some have a lot of edges, some consists of plain areas, texture pattern, curves, etc. As we can see in figure 5.3 (a), building images contain a lot of edges and plain areas. Figure 5.3(e) shows that shape changes characterize a trade-mark image. Landscape and natural images (Fig.

5.3(b) and Fig. 5.3(f) respectively) show a mixture of curves and random texture pattern. Figure 5.3(g) and 5.3(h) are composed of repeated pattern and random changes to form pattern texture and random texture patterns respectively.



(a) *(b)* *(c)*

(d) *(e)* *(f)*

(g) *(h)* *(i)*

Figure 5.3: Images from nine image kinds. (a) Building; (b) landscape; (c) man-made single object; (d) man-made objects; (e) trade-mark; (f) natural image; (g) pattern texture; (h) random texture; (i) human face

5.7 Performance evaluation

To evaluate the performance of our proposed algorithm, precision and recall measures are used [66,47]. Precision is the percentage of correct retrieved images to the total number of retrieved images at a specific retrieval size, including all relevant and irrelevant images. Recall is defined as the percentage of correct retrieved images to the total number of relevant images from the database. Let n_c, n_m and n_f be the numbers of correct, missed and false candidates respectively. The precision p_q and recall r_q for a query image q are defined as,

$$p_q = \frac{n_c}{n_c + n_f} \times 100\% = \frac{n_c}{M} \times 100\% \tag{5.13}$$

$$r_q = \frac{n_c}{n_c + n_m} \times 100\% = \frac{n_c}{N_q} \times 100\% \tag{5.14}$$

Precision p_q is measured for the first M retrieved images. Our database contains a total of 1800 images, which are equally cataloged into nine classes each with 200 images. Thus, let the number of similar images N_q of each image kind be 200 and the size of the database be 1800. It is desirable that both p_q and r_q approach 100%. To obtain accurate precision and recall values to evaluate the performance of our proposed indexing algorithm, we calculate the values from each catalogue and then obtain average precision p_{av} and average recall r_{av} values from all images, i.e.,

$$p_{av} = \frac{1}{K} \sum_{q=1}^{K} p_q \tag{5.15}$$

$$r_{av} = \frac{1}{K} \sum_{q=1}^{K} r_q \tag{5.16}$$

where K is the number of query images involved in our experiment

The precision and recall figures reveal the retrieving performance of our proposed indexing algorithm when we are looking for JPEG2000 images that are similar to query JPEG images. From their values, we can evaluate the accuracy of our proposed image indexing algorithm in different compressed domains.

5.8 Indexing results at different compression ratios

Due to the differences in compression algorithms such as quantization, the preserved energy in JPEG and JPEG2000 are quite different even at the same compression ratios as shown in figure 5.4. JPEG2000 retains more information than JPEG and results in better quality. To test the indexing performance at different compression ratios, all images in the database are compressed with seven compression ratios, they are 1.6:1, 9.5:1, 19:1, 27:1, 37.5:1, 51.5:1 and 72:1. Therefore, JPEG compressed images will try to find the most similar JPEG2000 images from the database at the same compression ratios. According to the compression algorithm standardized in JPEG2000, users can choose to use B97 kernel or other user defined wavelet kernels. Thus, in this simulation, JPEG2000 images are compressed by Haar, DB4, B93 and B97 wavelet kernels. The query JPEG image will search for the most similar JPEG2000 images to evaluate the indexing performance from different compressed domains. The search will be carried out on each image kind for four times as JPEG2000 images are compressed by four different wavelet kernels. If the resultant indexed images come from the same catalogue as the query image, the retrieving is classified to be a correct case; otherwise a false case. As the retrieval is performed at different compression ratios across different kinds of images, average precision is calculated for each image kind. To evaluate the accuracy of our proposed feature extraction algorithm, overall precision across different image kinds is obtained at various retrieval sizes to estimate the retrieval performance. Figure 5.5 shows the interface of our image retrieval system. It shows the first 12 most similar images. It also allows the user to calculate the precision value from the user defined retrieval size. Figure 5.6 shows a flowchart of our proposed retrieval system.

(a) *(b)*

Figure 5.4: Lena image compressed by (a) JPEG and (b) JPEG2000 at compression ratio of 72:1

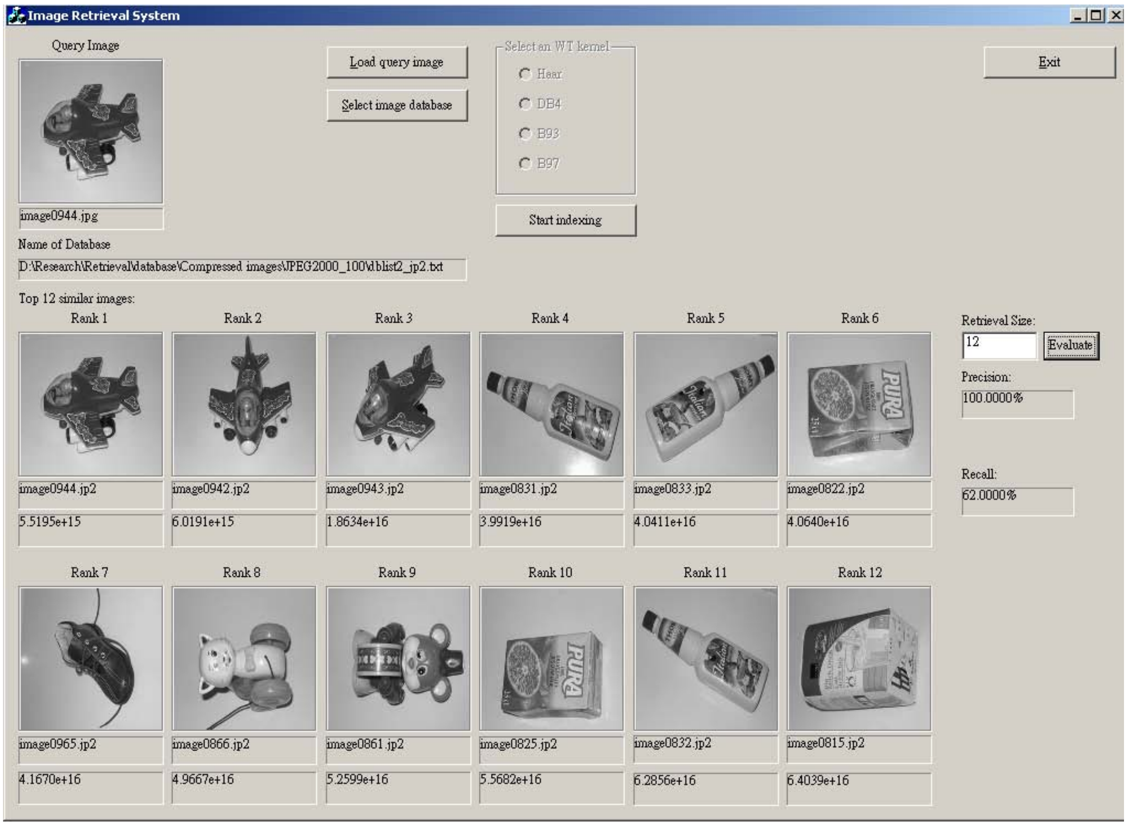


Figure 5.5: Interface of our proposed retrieval system to search images from different compressed domains.

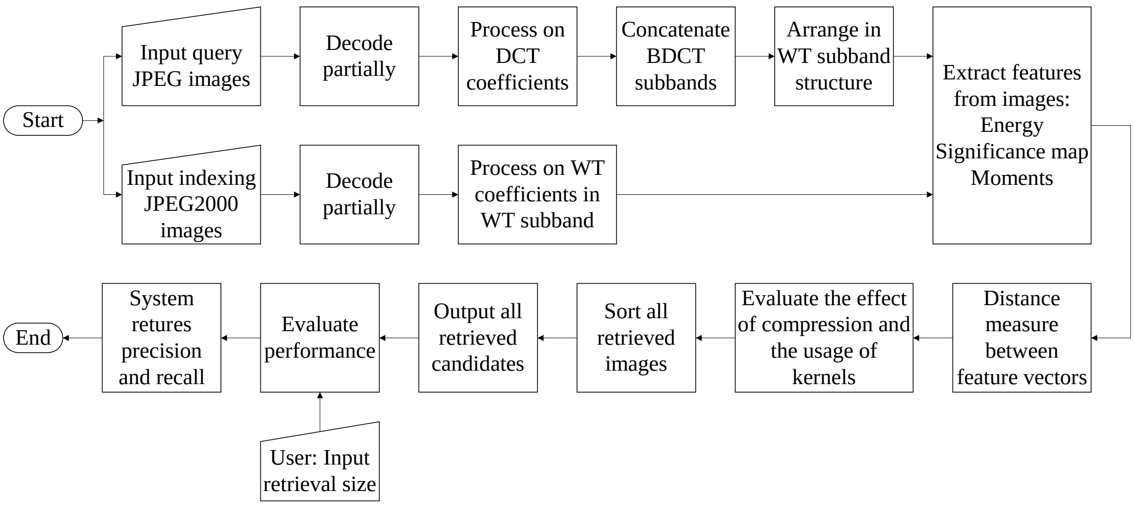


Figure 5.6: Flowchart of our proposed retrieval system.

5.8.1 Precision

5.8.1.1 Precision versus compression ratio

To investigate the effect of compression and wavelet kernel on image indexing, average precision of various image kinds that are compressed by different wavelet kernels at different compression ratios is studied. Figure 5.7 plots the average precision of retrieving JPEG images from JPEG2000 images versus different compression ratios when the retrieval size is 1. In other words, we consider the first rank of the indexing result only to see whether the query JPEG image and the most similar retrieved JPEG2000 image are from the same catalogue. This is because in practical image indexing, users want to search for images that are most similar to the query image but not necessarily the same one. Average precision is calculated from averaging the precision from the same image kind, i.e., K=200. As JPEG2000 images are compressed by four wavelet kernels in this simulation, the average precision of retrieving JPEG images from JPEG2000 images are plotted separately for each kernel, i.e. Haar, DB4, B93 and B97.

Figure 5.7 shows that the average precision of most image kinds slightly increases when compression ratio increases from 1 to 20 in Haar and DB4 cases and from 1 to 10 in B93 and B97 cases. It then drops when the compression ratio increases further. This figure shows that the selection of wavelet kernel greatly influences the average precision of a particular image kind. When JPEG2000 images are compressed by Haar and DB4 wavelet kernels, the average precision of most image kinds are over 50% even under high compression ratio. But when B93 and B97 are used, the average precisions are lower than that of Haar and DB4 cases. The precision of building and random texture pattern drop below 40% rapidly when the compression ratio is over 40. In particular, the average precision of landscape is always below 30% for B93 and B97 kernels no matter what the compression ratio is.

The above analysis shows that the indexing accuracy of the first rank indexing image greatly depends on the compression ratio and compression kernel. The effect of compression certainly influences the remaining ranks of indexing.

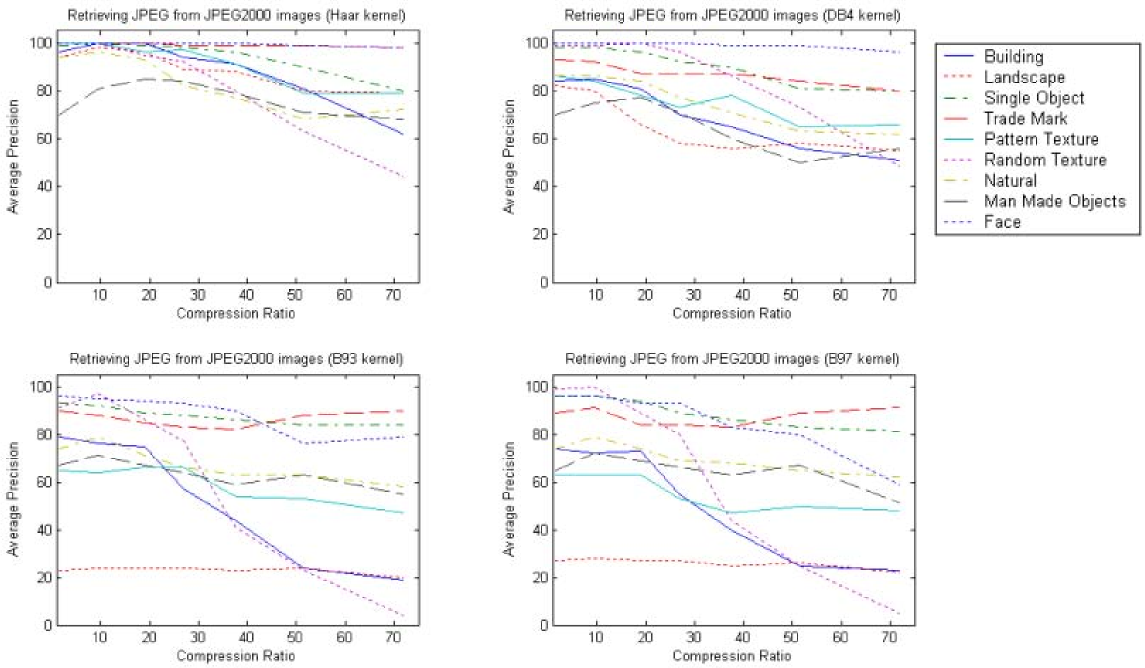


Figure 5.7: Average precision of each image kind versus compression ratio for Haar, DB4, B93 and B97 wavelet kernels

5.8.1.2 Precision versus retrieval size

Average precision versus retrieval size provides another important data for us to evaluate the proposed retrieval system. Retrieval size in fact considers the number of retrieved images that are taken in our measurement. Precision at various retrieval sizes tells us how much relevant information can be obtained from the system. Thus the precision value may change as retrieval size increases. Average precision is then obtained from the same image kind (i.e., K=200) no matter what kernel is used and what compression ratio it is. In figure 5.8, we plot the average precision for different image catalogues to observe how their image characteristics, together with the compression ratio, affect the indexing results at different retrieval sizes. In this figure, average precision is calculated from averaging precision at all compression ratios over the four kernels.

Besides the human face, the average precision of all other image kinds drops about 10% to 30% when retrieval size increases. This means more irrelevant images are retrieved when more images are considered in our calculations. More than half of the image kinds provide average precision of over 50% when retrieval size increases. In particular, human face images give outstanding indexing results, then followed by single object images. Unlike figure 5.7, average precision of landscape images is not the worst among others. The lowest average precision now is the pattern texture kind which drops rapidly when retrieval size is larger than 6. The average precision of random texture and pattern texture shows fluctuation when retrieval size is less than 10 and then drops more than 20%. In particular, the average precision of pattern texture drops more than 50% when retrieval size increases from 6 to 200.

In figure 5.8, we only observe the average precision of various image kinds over the four kernels. To observe the effects of the kernel versus retrieval size, figure 5.9 plots the average precision for similar JPEG2000 images that are compressed by different wavelet kernels. In this figure, the curve for a particular wavelet kernel is plotted by averaging all precision values from all image kinds (i.e., K=1800) at all compression ratios.

Average precision from indexing JPEG2000 images that are compressed by four different wavelet kernels drops when retrieval size increases. This matches our overall observation in figure 5.8. Results also confirm our similarity measure in chapter 4 that when JPEG2000 images are compressed by the Haar kernel, it shows the highest similarity with JPEG images such that more relevant images can be retrieved. Theoretically, the similarity between BDCT and WT outputs that we analyzed in chapter 4 are Haar, DB4, B97 and B93, in descending order. Same similarity order is obtained again in the average precision plot in figure 5.9.

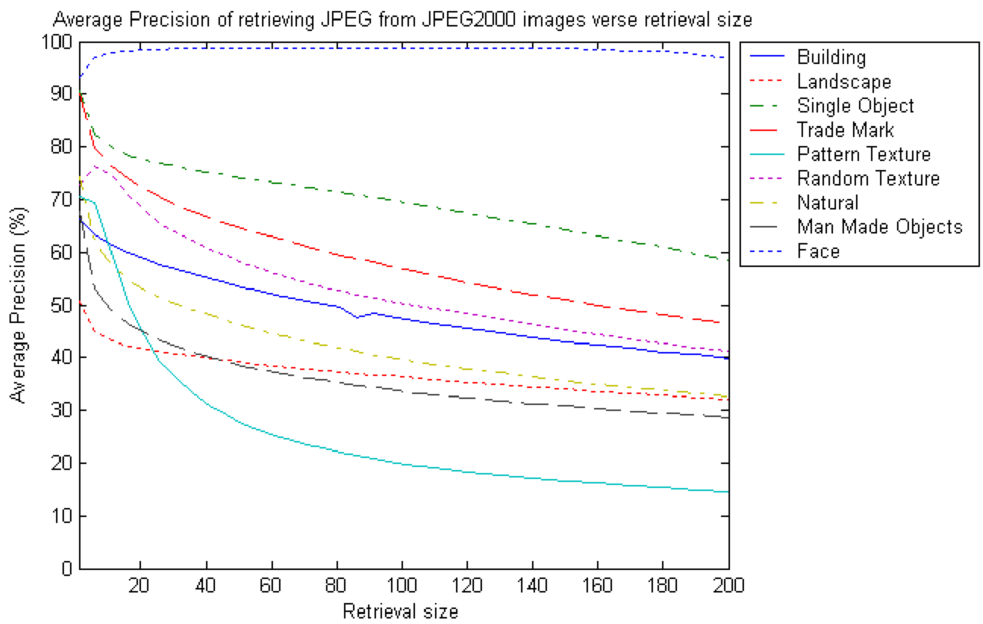


Figure 5.8: Average precision of each image kind versus retrieval size in ignorance of wavelet kernel and compression ratio.

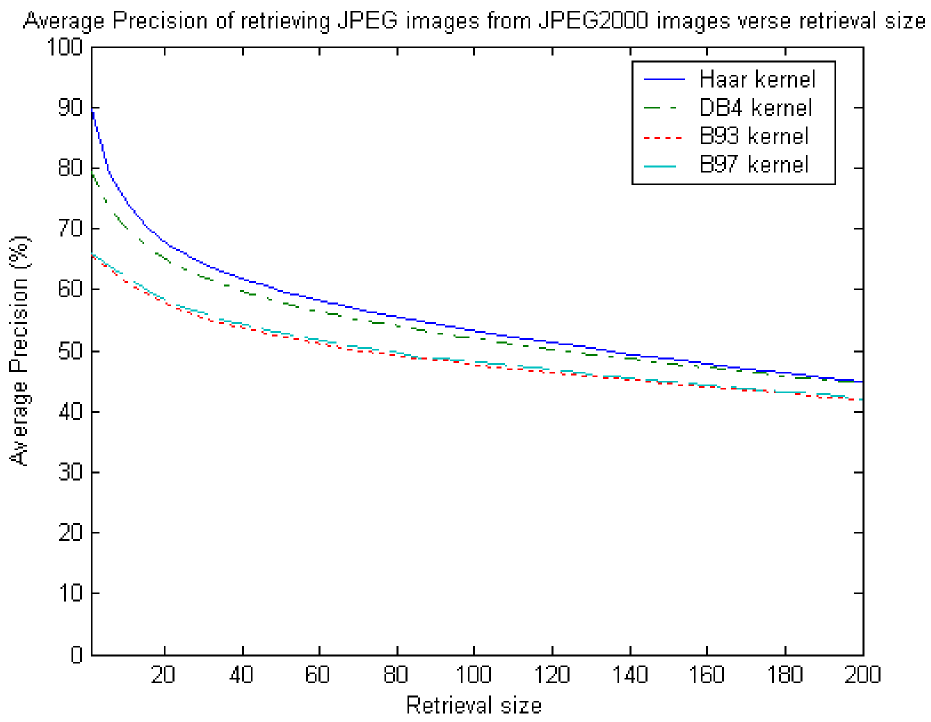


Figure 5.9: Average precision of all image kinds versus retrieval size when JPEG images is retrieved from JPEG2000 images with different kernels.

5.8.2 Recall

Recall is another parameter to estimate the ratio of correct retrieved images to the size of a particular image kind. With reference to equation 5.14, a large recall value, r_c, means that more relevant images (n_c), are retrieved from all relevant images. On the contrary, a small r_c is obtained if more relevant images are missed. Thus recall estimates how many images are correctly retrieved from each image kind of 200 images (N_q=200).

5.8.2.1 Recall versus compression ratio

Two figures (figure 5.10 and figure 5.11) of average recall versus compression ratio are plotted when the retrieval size is 200. As JPEG2000 images are compressed by four different wavelet kernels, figure 5.10 considers the retrieval results when images are compressed by each of the four kernels at different compression ratios. Average recall is then calculated by averaging recall from all images kinds (i.e. K=1800).

The average recalls for all kernels are below 50% and drops slightly when compression ratio increases. This means that nearly half of relevant images are retrieved no matter how big the compression ratio is. As average recall only decreases slightly even under high compression ratios, compression does not affect the number of correct and missed images when the retrieval size is as large as 200.

Information preservation of different image kinds is different after compression. In figure 5.11, we study the indexing results on each image kind at various compression levels. The average recall is obtained from the same image kinds (i.e. K=200) and also from the four JPEG2000 compression cases. This means that no matter what compressed kernels they employ, all recalls from the same image kinds are averaged. Figure 5.11 shows the average recall of each image kind versus compression ratio.

In the first 200 ranks, most image kinds miss nearly half of their relevant images. When compression ratio increases, average recall of all image kinds, besides building and random texture, fluctuate when the compression ratio is less than 20 and then slightly increase after this. Recall of building and random texture increase when compression ratio is less then 30 and 10 respectively and then drop more than 20% when compression ratio increases to 72. In contrast with the result in subsection 5.7.1, pattern texture misses the smallest number of relevant images.

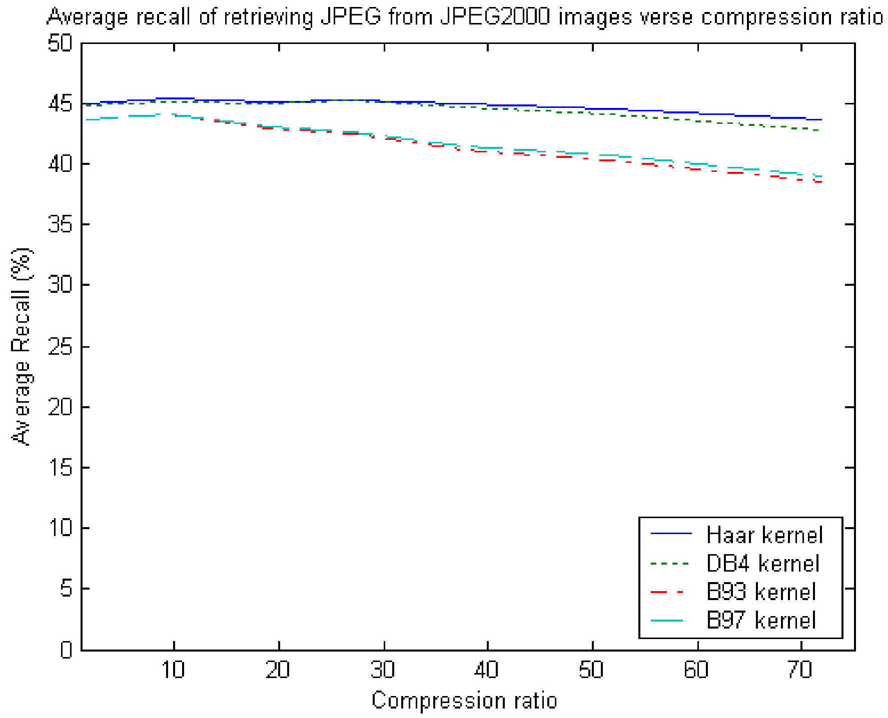


Figure 5.10: Average recall of all image kinds versus compression ratio is plotted when JPEG images are retrieved from JPEG2000 images with different kernels.

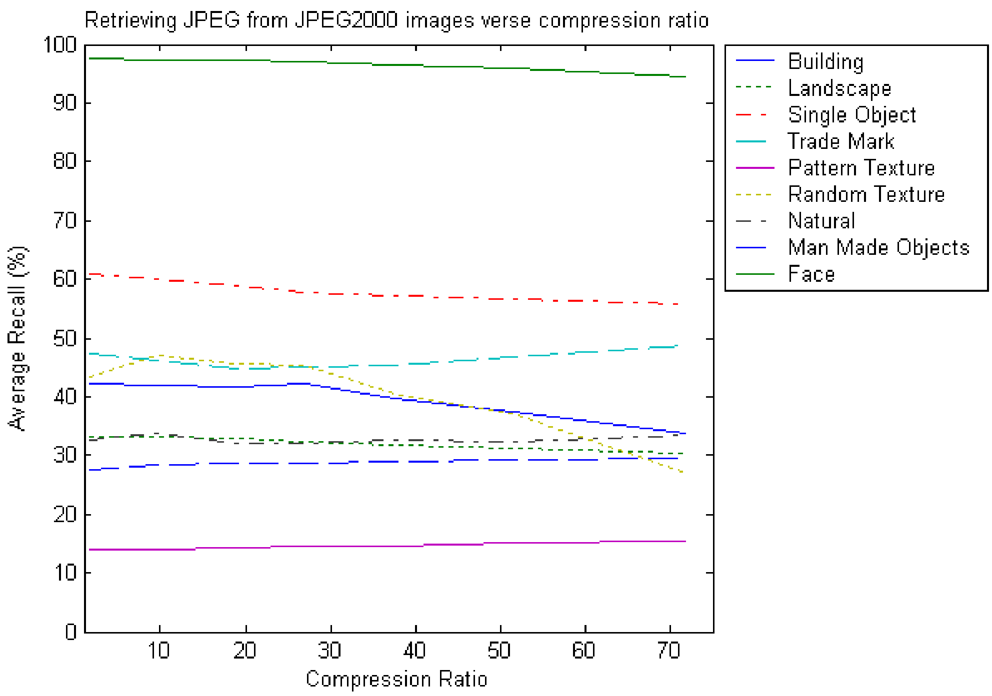


Figure 5.11: Average recall versus compression ratio of each image kind in ignorance the usage of compression kernel.

5.8.2.2 Recall versus retrieval size

Recall versus retrieval size provides another information to evaluate how many relevant images can be retrieved from all available images. Figure 5.12 plots the average recall versus retrieval size for different image kinds. The average recall is obtained by averaging the recall values from retrieving every image kind (i.e. K=200) over all compression ratios. Since each image catalogue contains 200 images, it is desirable that the recall approaches 100% as retrieval size tends to 200. In general, the average recall of all image kinds increase steadily as the retrieval size increases. In particular, the recall of face images performs the best; it rises to 96% when retrieval size is 200. Following human face, single object images give good recall results as it rises to 60%. Over half of the single object images can be retrieved successfully when 200 retrieved images are used. Besides human face and single object images, the average recall of all the other image kinds rises to 50% when retrieval size is 200. The average recall of pattern texture is the lowest among all image classes. This figure also illustrates that the average recall of landscape and natural images are very close to each other.

Figure 5.13 describes the average recall versus retrieval size of searching JPEG images from JPEG2000 images that are compressed by different wavelet kernels at different compression ratios. All of the four recall curves increase almost steadily from near 0% to over 40%. Searching JPEG images from JPEG2000 images that are compressed by the Haar kernel give the best recall results, which is followed by DB4, B97 and B93.

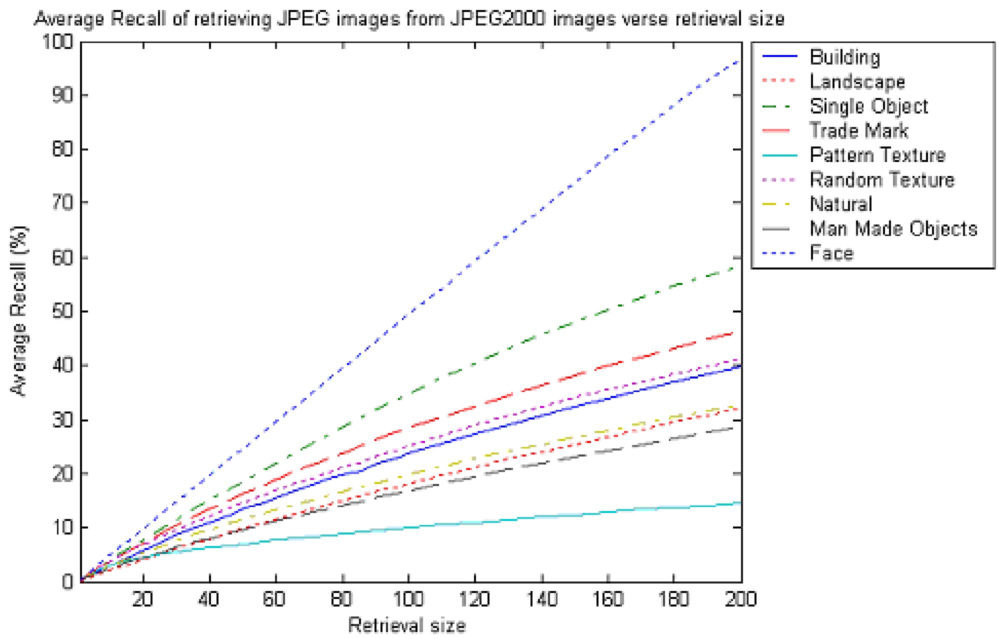


Figure 5.12: Average recall of each image kind versus retrieval size

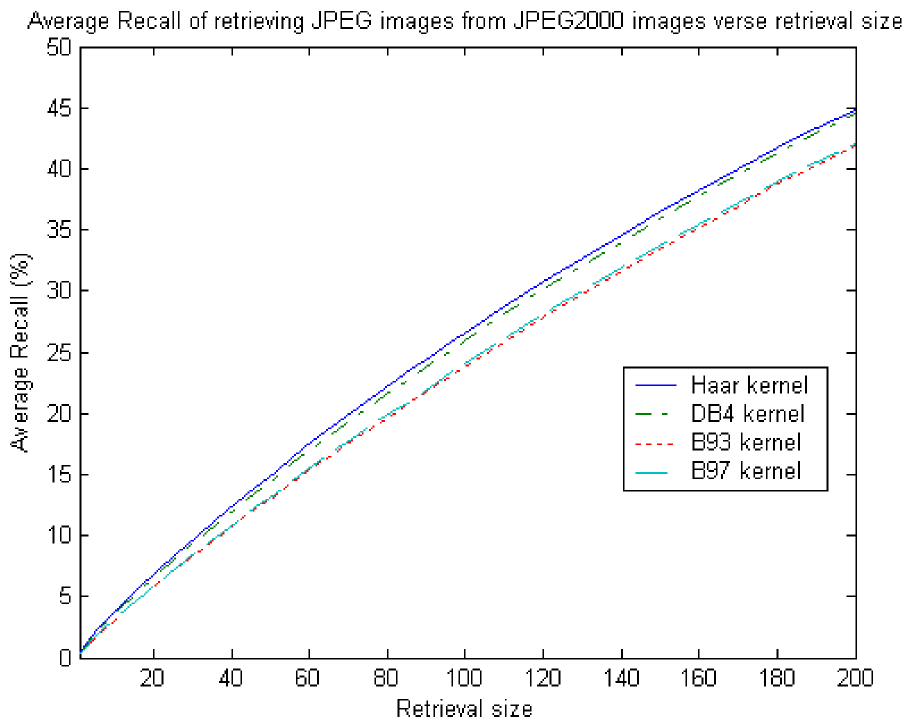


Figure 5.13: Average recall versus retrieval size of retrieving JPEG images from JPEG2000 images that are compressed by different wavelet kernels.

5.9 Chapter summary

Theoretically, Block-based Discrete Cosine Transform (BDCT) and Wavelet Transform (WT) outputs are proven to be very similar in chapter 4. Evaluation of our derived subband model in practical retrieval systems is performed in this chapter.

In accordance with JPEG and JPEG2000, BDCT coefficients are arranged in 8×8 blocks that are quite different from WT coefficients in subbands. This makes direct feature comparison to be difficult if it is done on BDCT blocks and WT subbands. To process feature extraction smoothly and efficiently, the BDCT coefficients are concatenated into WT subband structure. This concatenation is in fact included in our subband filtering model. This places both sets of coefficients in the same arrangement to facilitate direct feature extraction.

Three characteristics of images, which are energy, significance map and normalized central moments of the significance map, are then defined and obtained from the resultant BDCT and WT subbands to form a feature vector. Based on the degree of significant information preserved in BDCT and WT subbands, energy will be obtained from seven subbands that contain most of the energy. Thresholding technique is applied to the subbands to construct their significance maps by retaining their dominant coefficients. Then second and third order normalized central moments are calculated from the significance maps of the subbands to measure their variance and skewness with respect to their centers of mass. This makes the proposed feature vector translation and rotation invariant. Since the variance and skewness are more or less the same across all decomposition levels in the same frequency orientation, the size of the feature vector is reduced as normalized central moments are calculated from all subbands at the highest decomposition level only. Finally, the three characteristics generate a 35 dimensional feature vector. Euclidean space measure is then used to quantify the distance between query and indexing images.

A retrieval system is established to search for JPEG and JPEG2000 images to examine our image indexing in different compressed domains. One thousand and eight hundred images from nine catalogues are compressed at seven compression ratios ranging from 1.6:1 to 72:1. These nine kinds of images consist of plain areas, edges, texture and man-made graphics, which contain both low and high frequency contents. Precision and recall are plotted against compression ratios and retrieval sizes to evaluate our proposed retrieval system. It is found that different image kinds give quite contrasting retrieval performances. Compression certainly affects the retrieval accuracy. When a large retrieval size is considered, precision decreases, whereas, recall increases. Moreover, the indexing performance depends on the usage of WT kernels, i.e., Haar, Daubechies 4, Biorthogonal 93 and Biorthogonal 97. Detail discussion on the retrieval results will be made in chapter 6.

Chapter 6

Discussion on retrieval performance

6.1 Introduction

A subband filtering model was developed in chapter 3 to provide a theoretical framework for comparing the BDCT and WT operations in JPEG and JPEG2000 respectively. Using this model, results in chapter 4 confirm the high similarity between their outputs even under high compression ratios. Thus, common features do exist between JPEG and JPEG2000 in their respective domains so that full decompression can be avoided.

In chapter 5, a simulation is done to evaluate our proposed image searching algorithm by the use of the subband filtering model. A feature vector that characterizes the image features is defined to find relevant images in these two domains. Precision and recall of the proposed retrieval system are plotted to measure the accuracy of our proposed indexing algorithm. As illustrated in the figures, precision and recall depend on the compression ratios, retrieval sizes and image contents greatly. Different image kinds thus give quite different precision and recall values as compression ratios and retrieval sizes change. In this chapter, a detailed discussion on the simulation results will be given.

6.2 Precision

Precision reveals the number of retrieved images that are visually similar to the query images. In this subsection, we describe the effect of compression ratio, retrieval size and wavelet kernels on precision.

6.2.1 Discussion on precision versus compression ratio

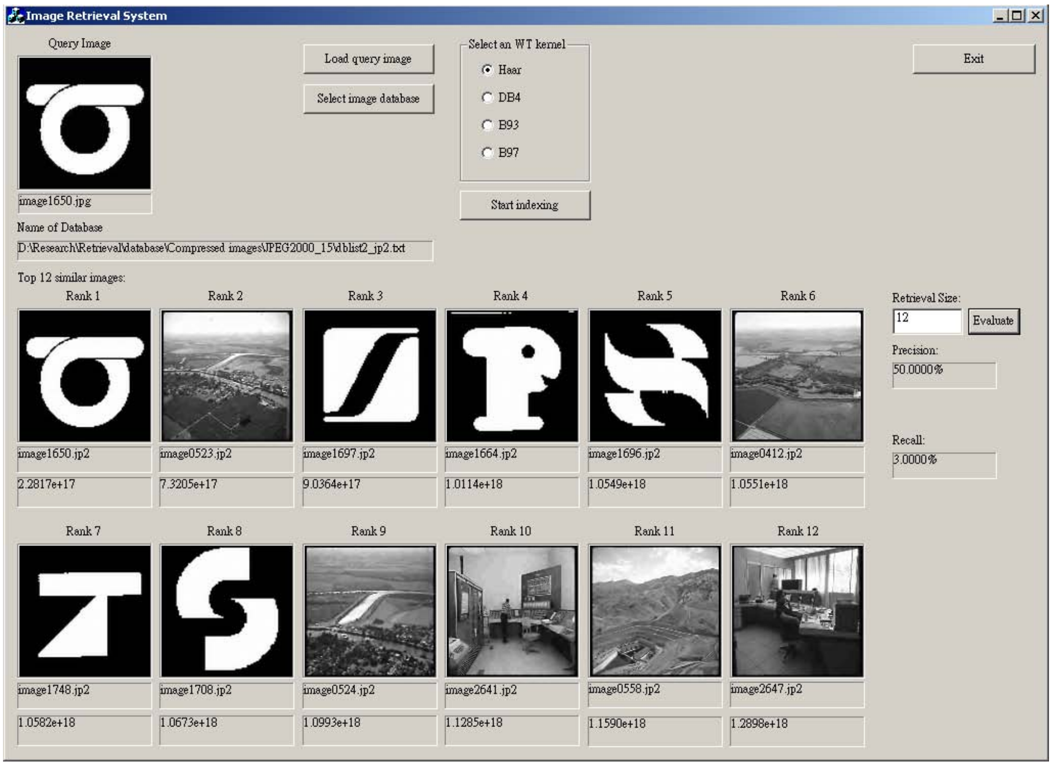
Figure 5.7 shows a plot of the average precision of retrieving JPEG images from JPEG2000 images versus the compression ratio when the retrieval size is 1. This figure plots the precision for each image kind against individual wavelet kernels. By comparing the curves between the four kernels, it is found that Haar and DB4 kernels comparably give higher values than B93 and B97 kernels. This confirms our analysis in chapter 3 and 4 that the filtering operations of Haar and DB4 employed in JPEG2000 are more similar to the BDCT operation in JPEG, thus give more similar compressed outputs than that of B93 and B97 kernels. Their different filtering operations thus generate quite different outputs. Thus, feature vectors extracted from the two transformed domains would be quite different. The output difference between BDCT and the four WT kernels is further emphasized by the different compression schemes in JPEG and JPEG2000. Though images belong to the same catalogue, they are far more apart in the distance measure after transformation and compression. This explains the wide distribution of the precision curves of various image kinds from 0% to 100% at very low compression ratios. When we look at the precision curves of Haar and DB4 kernels, the curves describing individual image kinds are clustered at the range from 70% to 100% at low compression ratios but 40% to 60% at high compression ratio. In contrast, curves for B93 and B97 kernels are widely distributed at the range from 20% to 99% and 3% to 90% at low and high compression ratio respectively.

Hence, when the compression ratio increases, all curves decline with fluctuations. This fluctuation is due to the difference between JPEG and JPEG2000 compression schemes, such as quantization. The compression schemes blur information in the higher frequency subbands which store fine details of the images. The fine details in images such as texture patterns and edges may be lost, and sharp changing areas may look like plain areas after compression. For example, the curly hair of Lena (Fig. 6.1(a)) lost its texture pattern after compression (Fig. 6.1(b)). This thus affects the features that can be extracted from these higher frequency subbands, including the energy, variance and skewness of the significance map. The originally unlike images become similar after compression which leads to false retrieving. But when compression increases, more insignificant changes are lost while important sharp changes are preserved. This makes images from different kinds distinguishable from one another again. This would cause fluctuation in the similarity between JPEG and JPEG2000 images after compression. An example is shown in figure 6.2(a) where the compression ratio at 19:1 blurs the fine details in trade-mark images and result in six false candidates. But when the compression ratio is as large as 72:1, more correct images are retrieved (Fig. 6.2(b)). Thus, the retrieval accuracy of the first rank is greatly dependent on the compression ratios and compression kernels. The accuracy of individual image kind versus retrieval size is worthy of study.

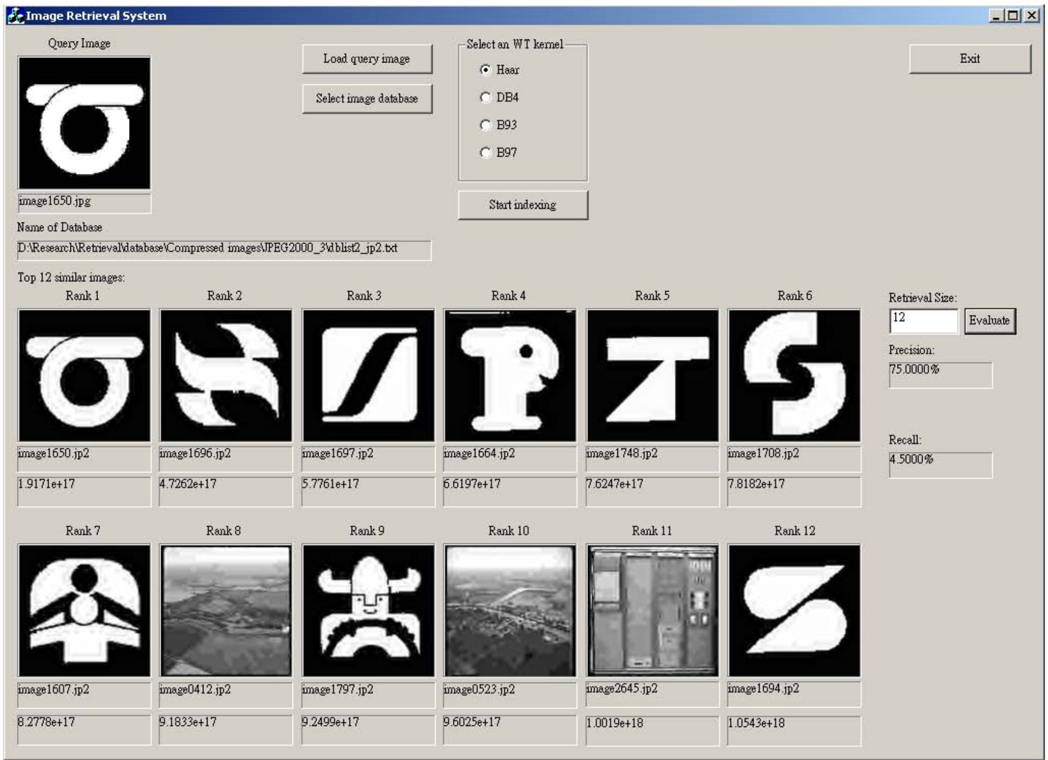
With reference to Fig. 5.7, landscape images give the lowest precision at the cases of B93 and B97 kernels. Landscape images contain several main edges and texture areas, which are bandpass and highpass information respectively. In accordance with the characteristics of B93 and B97 kernels, they preserves more spectral information in the bandpass and highpass regions than Haar and DB4 kernels. However, even B93 and B97 keep more information in the bandpass, they are discarded by compression during quantization. Thus, the main edges in landscape images become important features during retrieval. But these edges confuse with the edges in man-made objects and buildings. Retrieving landscape images give man-made objects and buildings as shown in Fig. 6.3. As a result, landscape images give the worse retrieval results when they are compressed by B93 and B97 kernels.

(a) *(b)*

Figure 6.1: Lena images (a) without any compression; (b) compressed by JPEG2000 at 72:1

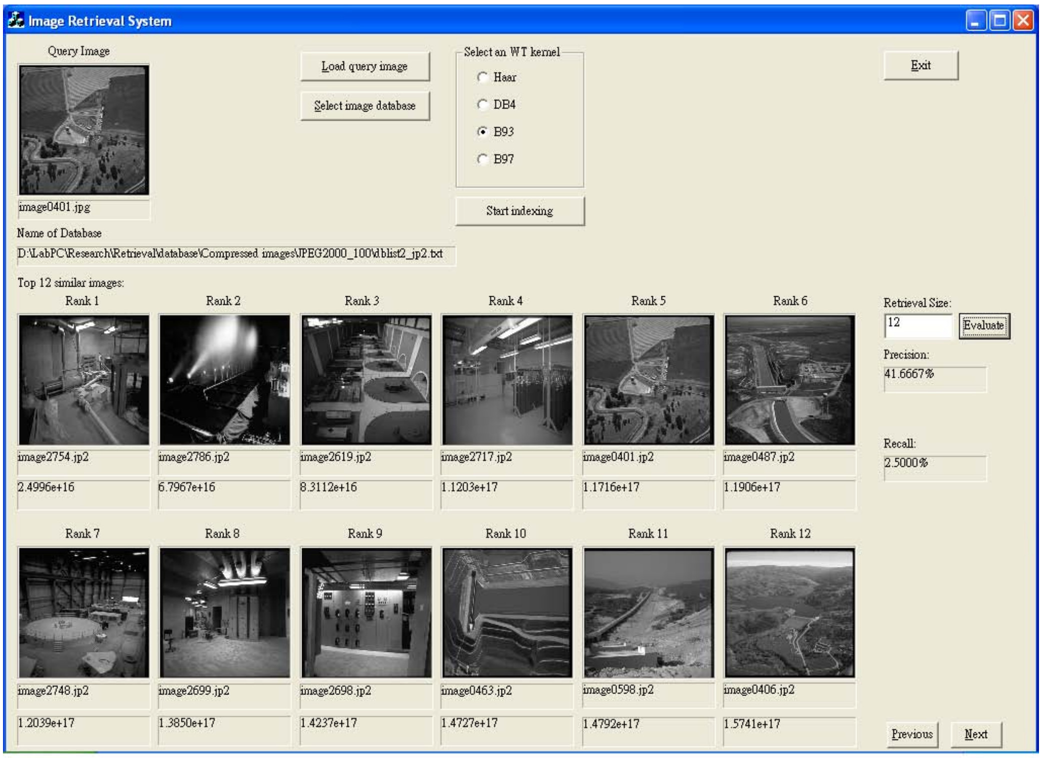


(a)

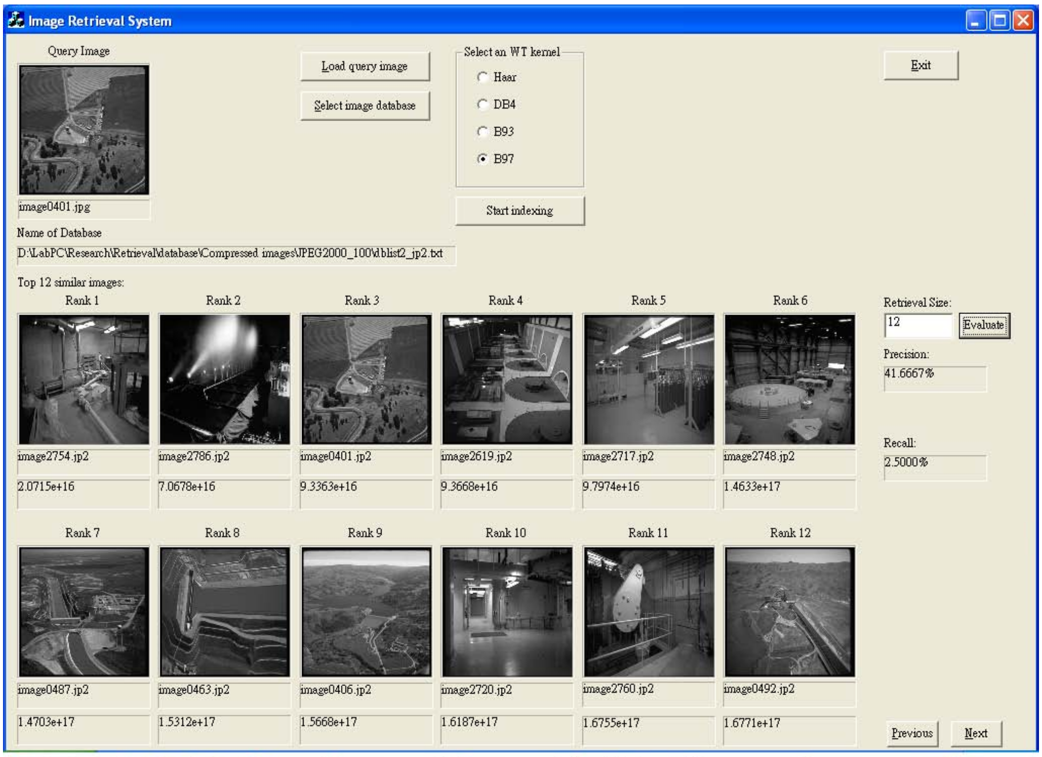


(b)

Figure 6.2: Retrieval results of a trade-mark image when compressed at (a) 19:1 and (b) 72:1



(a)



(b)

Figure 6.3: Retrieving of a landscape image when it is compressed by (a) B93 and (b) B97 kernels.

6.2.2 Discussion on precision versus retrieval size

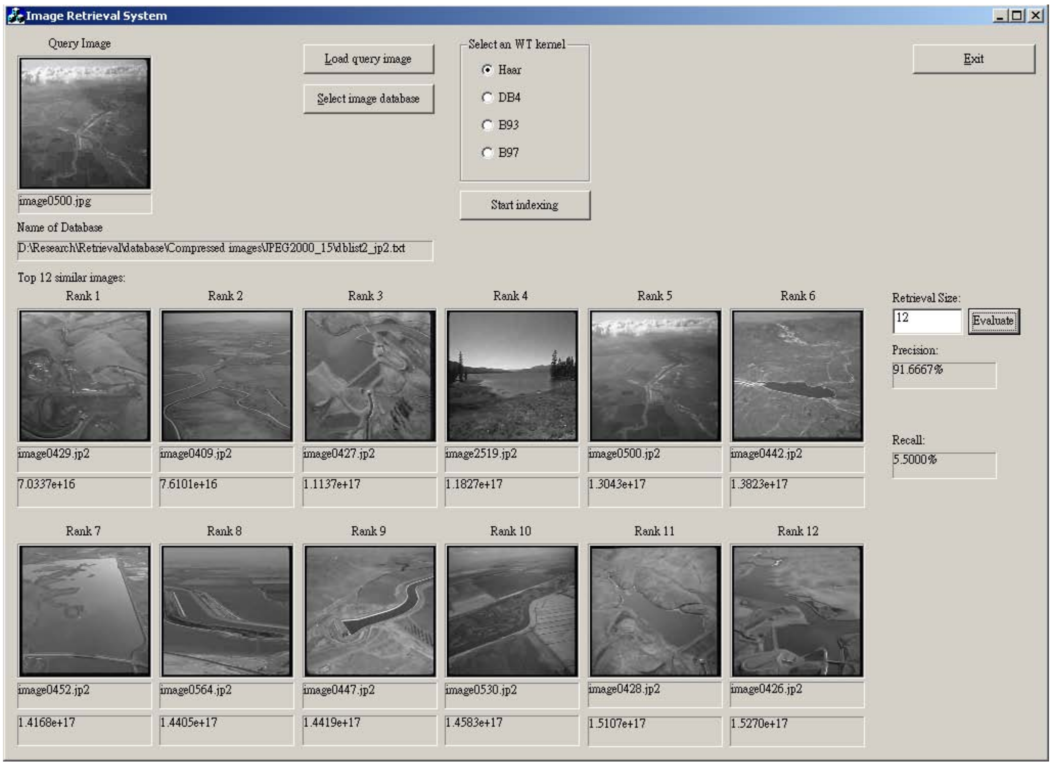
The analysis in section 6.2.1 shows that the indexing accuracy of the first retrieved image greatly depends on the compression ratio and compression kernel. The effect of compression certainly influences the remaining ranks of indexing. From the value of average precision versus retrieval size, we can evaluate how much relevant information can be searched from the proposed retrieval system. Average precision is obtained by averaging the precision of retrieving JPEG images from JPEG2000 images that are compressed by the four compression kernels and at all compression ratios. It helps us to observe how their image characteristics affect the indexing results along different retrieval sizes.

Besides the human face, the average precision of all other image kinds drops when the retrieval size increases. This means more irrelevant images are retrieved when more ranks are considered. As we look back at the analysis in section 6.2.1, compression reduces the retrieval precision due to the loss of high frequency information. Thus, the average precision from averaging all results at all compression ratios will be lower than the highest precision (99%) that we obtained in section 6.2.1. Also, sometimes compression blurs fine details so that different image kinds might look similar. For example, the appearance of landscape and natural looks similar after compression. Figure 6.4(a) shows that when a landscape image is compressed at 19:1, it retrieves mostly landscape images together with some natural images. But when it is compressed at 72:1, more landscape images are retrieved. From the results, we observe that the number of false candidates increases with the retrieval size. Compression makes landscape and natural images to be similar in appearance and thus reduces the average precision with retrieval size increases. Due to the similar appearance between landscape and natural images, similar indexing results are obtained when natural image becomes the query image (Figure 6.5).

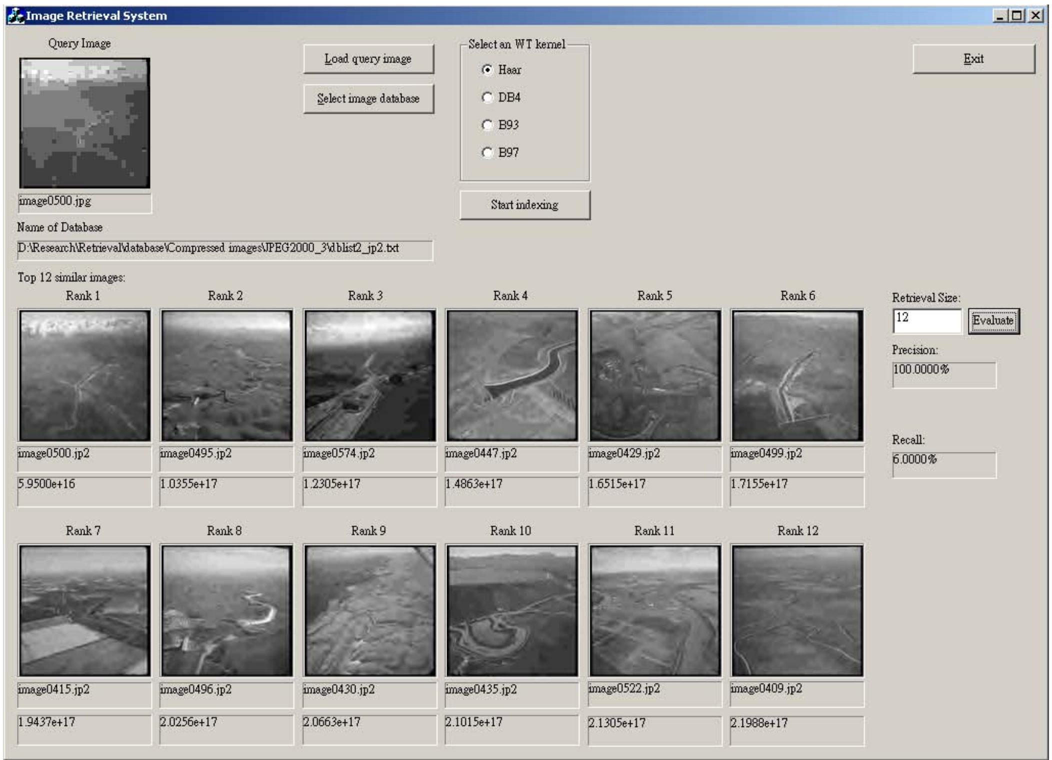
For other image kinds such as the human face, trade-mark and single object, since their special characteristics can be easily extracted even under high compression ratios, excellent retrieving results are obtained. For example, as all human face images contain mainly the human face, they can be easily identified even under high compression (Figure 6.6). Considering the trade-mark image in figure 6.7, all of them are man-made graphical images with sharp edges and boundaries, thus giving high retrieving accuracy under compression. As shown in figure 6.8, single object images also provide good retrieval results.

Half of the image kinds provide average precision of over 50% when the retrieval size increases. This means that about half of the images kinds preserve common characteristics after JPEG and JPEG2000 compression. Unlike the situation in figure 5.7, average precision of landscape is no longer the lowest. The lowest average precision now becomes the texture pattern which drops rapidly from 70% to 20% when the retrieval size is larger than 6. The false retrieving is caused by the loss of repeated high frequency characteristics under compression. The retrieval results as shown in figure 6.9 illustrate that buildings and other man-made graphics are indexed by this texture at a larger retrieval size. This is because texture pattern losses its pattern under compression and thus becomes similar to buildings or other man-made images. The retrieval precisions of landscape and natural scene are quite close to each other. If we refer back to figure 5.7, we found that the fluctuation of landscape and natural scenes always follows the same pattern with their precision, and rise up and drop down almost at the same compression ratios. This is because compression influences these two image kinds and they become indistinguishable.

In summary, the average precisions of different image kinds are quite different. The different compression algorithms in JPEG and JPEG2000 affect the retrieval results when more images are used. As most image kinds provide an average precision over 50%, over half of relevant images can be retrieved by the proposed image indexing algorithm.

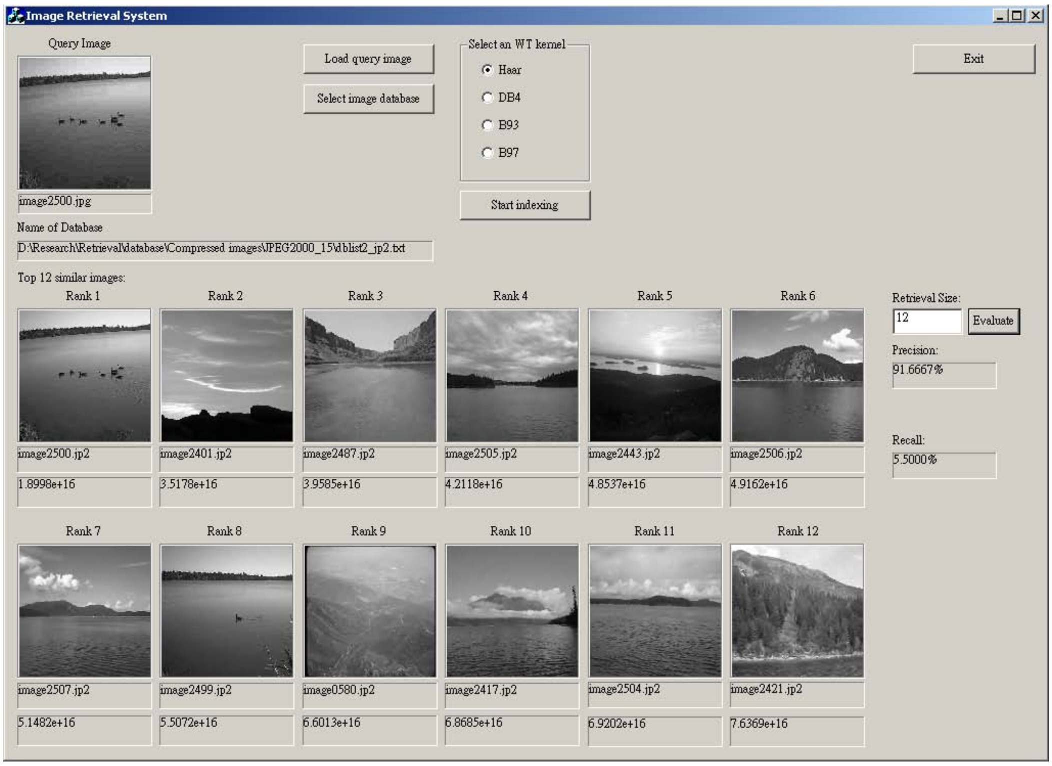


(a)

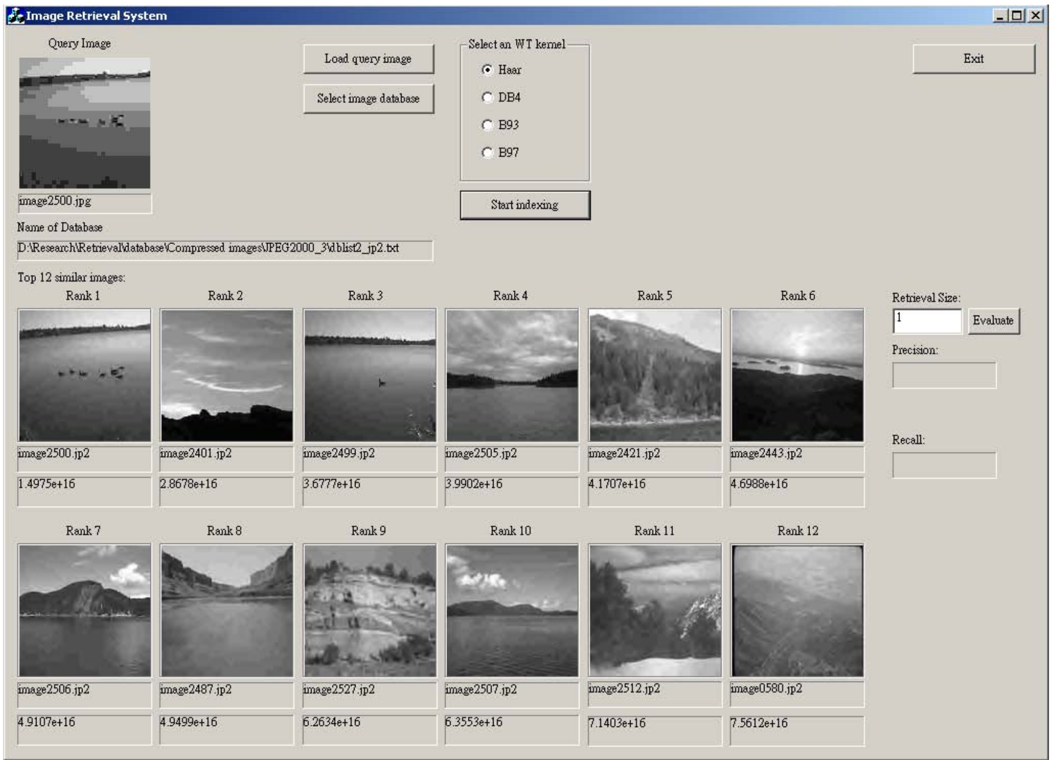


(b)

Figure 6.4: Retrieving results of compressed landscape images at (a) 19:1 and (b) 72:1



(a)



(b)

Figure 6.5: Retrieval results of compressed natural images at (a) 19:1 and (b) 72:1

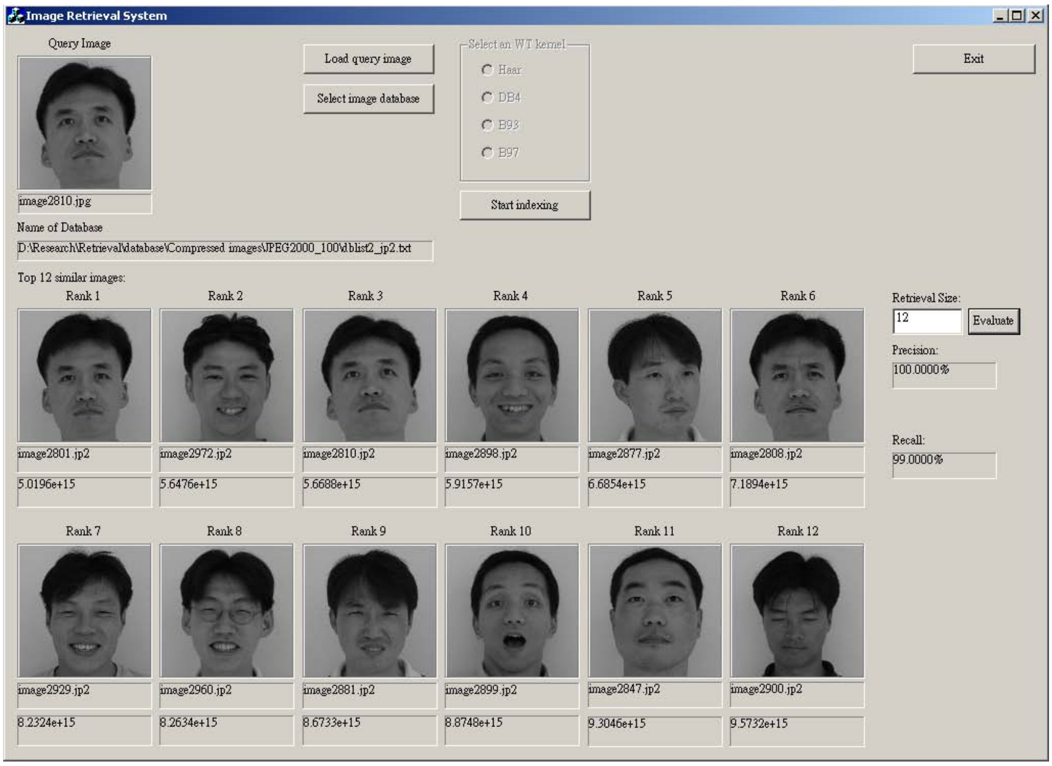


Figure 6.6: Retrieval result of a human face image.

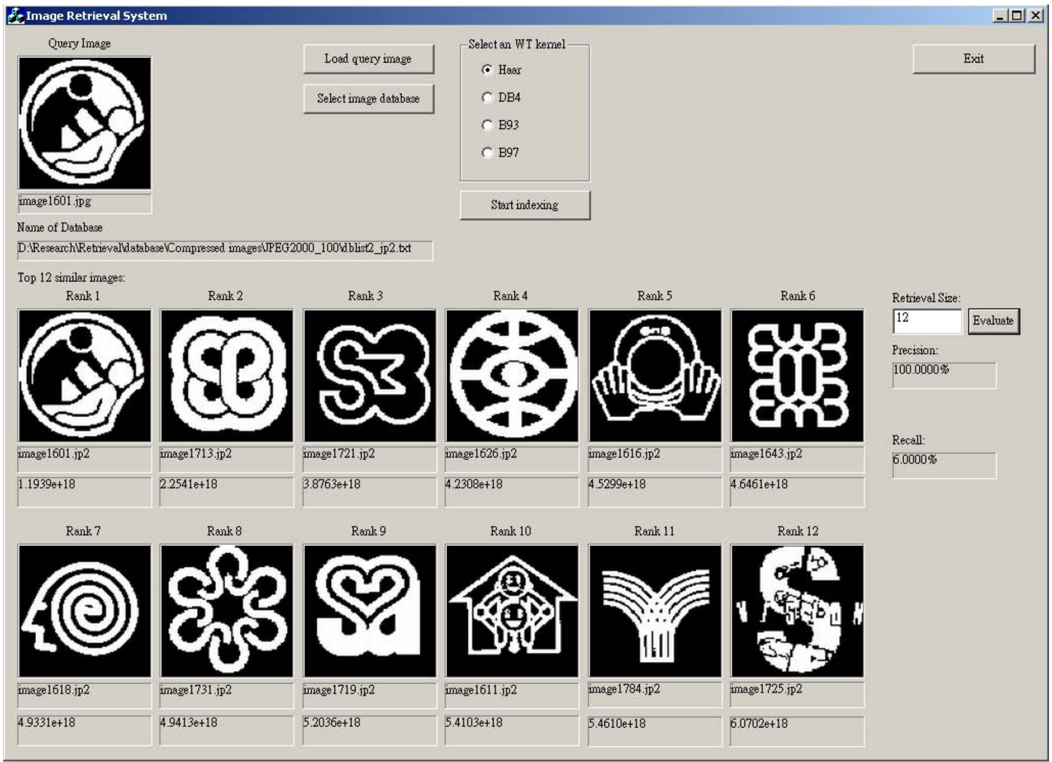


Figure 6.7: Retrieval result of a trade- mark image.

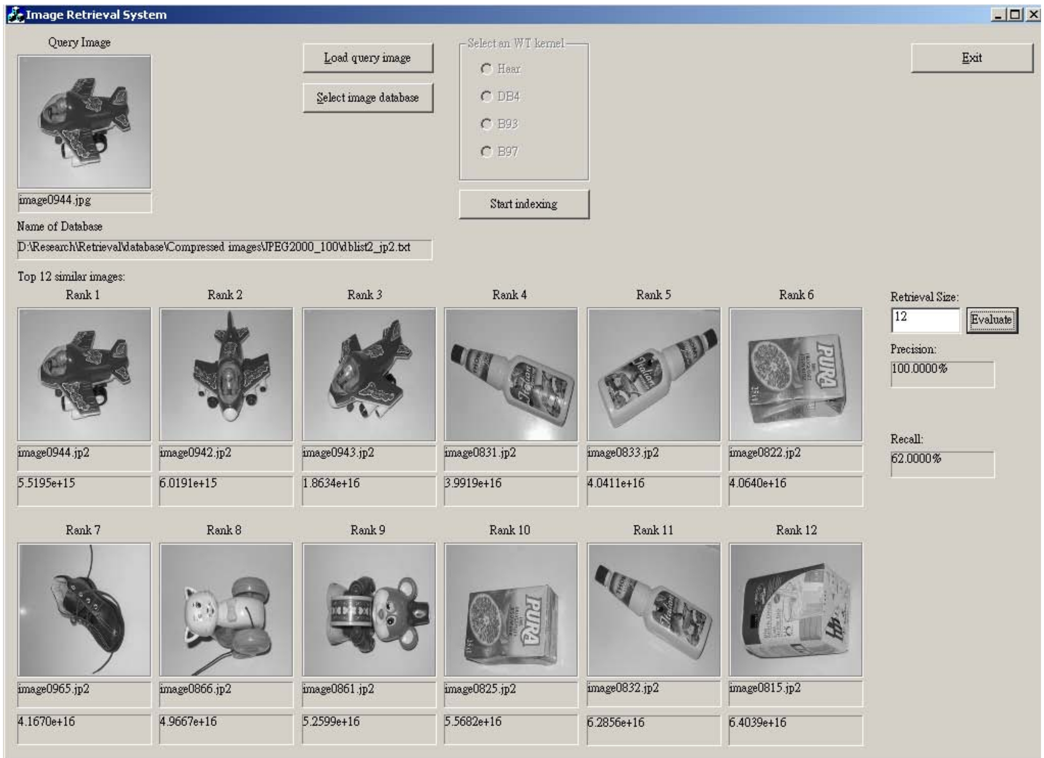


Figure 6.8: Retrieval result of a single object image.

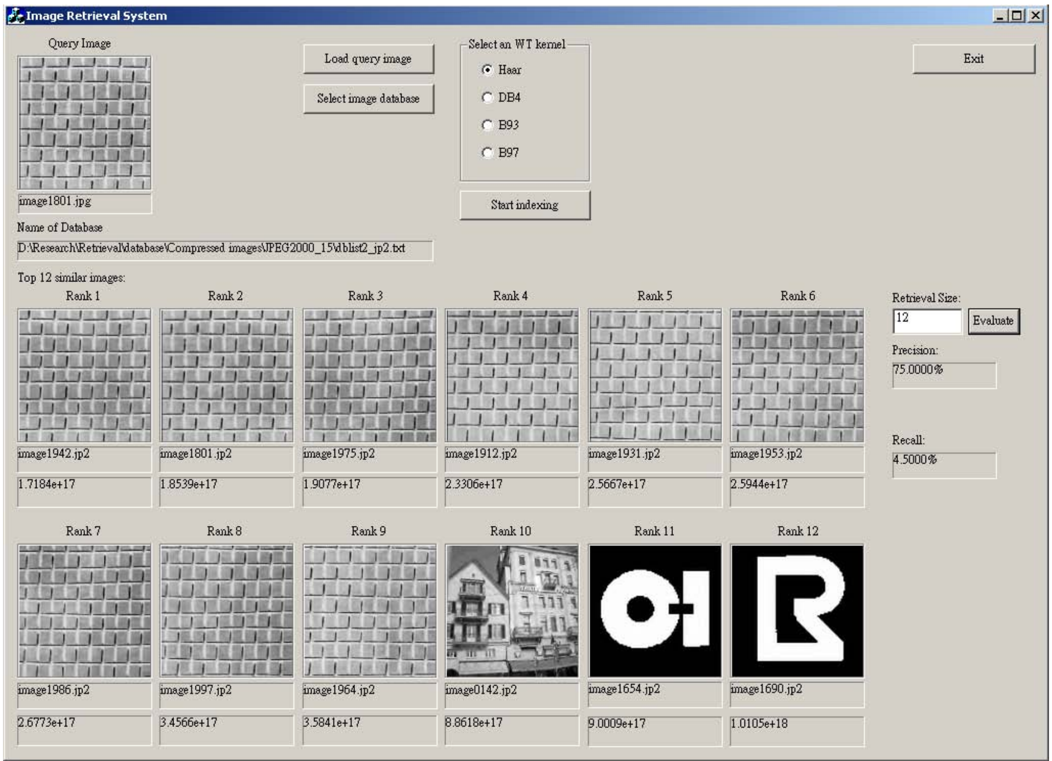


Figure 6.9: Retrieval result of a pattern texture image.

6.2.3 Discussion on precision versus retrieval size in usage of kernels

To observe the effects of wavelet kernels on precision, figure 5.9 plots the average precision versus the retrieval size of searching JPEG images from JPEG2000 images with different compression kernels. This figure aims to illustrate the precision of searching for relevant JPEG2000 images when they are compressed by different wavelet kernels. It can be seen that the Haar kernel gives the highest precision, followed by DB4, B97 and B93 kernels in descending precision. This confirms our analysis in chapter 3 that BDCT and Haar filtering operation are very similar and exactly the same in the lowpass subband. The higher similarity between transformed outputs imply that more common image characteristics are preserved and can be extracted in subsequent analysis. Our similarity measure in chapter 4 shows that B93 and B97 outputs are less similar to the BDCT outputs, especially at the higher bandpass bands. The lower similarity value means that information preserved in B93, B97 and BDCT are not the same. Thus, the retrieving results greatly depend on the usage of kernel.

There is a trend that precision decreases (from 90% to 40%) as retrieval size increases (from 1 to 200). Most of the retrieved images are relevant at the first 12 outcomes. More than half of the retrieved images are relevant when as large as 100 retrieved images are considered. The decline of precision along retrieval size is caused by compression. Though BDCT and WT have very similar filtering operations, we prove that they still preserve different image characteristics. In addition, a high similarity is found between BDCT and WT in both filtering operations and together in their outputs, compression degrades their similarities. This is proven in chapter 4.5. The situation becomes worse when the compression ratio increases. As our average precision values are obtained by averaging the precision values at all compression levels, the average precision is degraded by compression. Compression leads to losses in high frequency details. Losing the high frequency details in compression, natural scenes might not look similar to each other but images from other classes might become indistinguishable. The rate of decline in precision is very slow when the retrieval size is over 40 which means that a steady amount of relevant images is searched.

In summary, if we want to find JPEG2000 images that are relevant to the JPEG query image, nearly half of the retrieved images are relevant no matter they are compressed by JPEG2000 Part I or II. Regarding the direct image indexing in different compressed domains, the Haar kernel gives the best retrieval results.

6.3 Recall

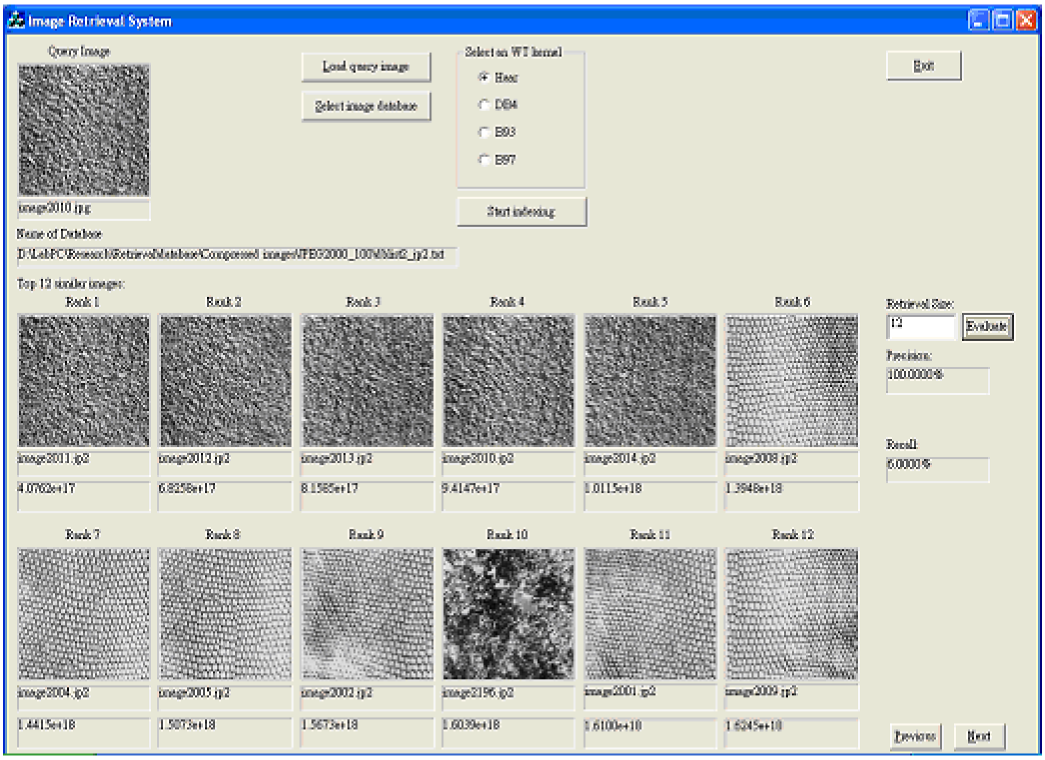
Recall defines the proportion of correct retrieval to the total relevant information. It illustrates how much relevant information are retrieved and lost in the proposed image indexing algorithm.

6.3.1 Discussion on Recall versus compression ratio

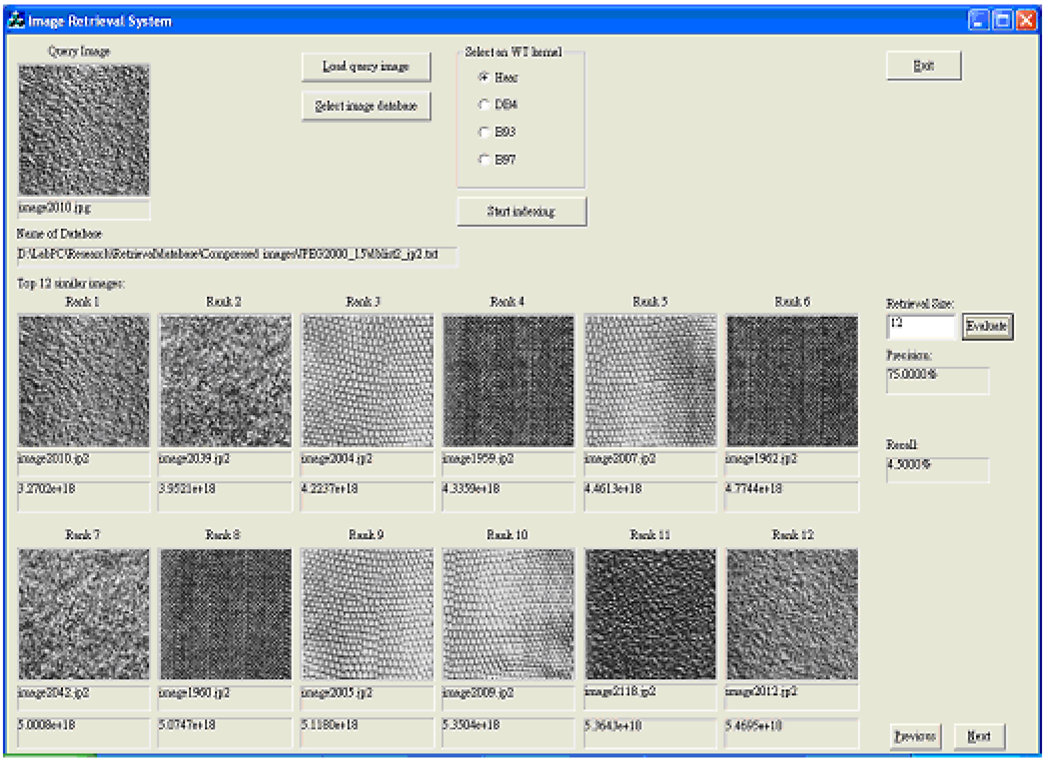
Figure 5.10 shows the recall versus compression ratio when searching JPEG images from JPEG2000 images which are compressed by different kernels. Average recall of the first 200 resultant ranks is calculated by averaging the recall of all images kinds (i.e. K=1800) for a particular wavelet kernel. As each image kind contains 200 images, it is desirable that the recall reaches 100% at all compression ratios. As figure 5.10 shows, the average recall is all below 50% at low compression ratios and then drops slightly to 40% when the compression ratio increases to 72:1. As recall reveals the ratio of correct images to the total number of relevant images, this means about half of relevant images are retrieved under all compression ratios. As average recall decreases only slightly even under high compression ratio, compression does not affect the number of correct and missed images seriously when the retrieval size is as large as 200.

However, figure 5.10 shows the average recall on all image kinds only. In fact, the recall performance of different image kinds is varying. Figure 5.11 plots the average recall of each image kind individually. The average recall is obtained by averaging the values from the same image kinds (i.e. K=200) with different WT kernels. This means that no matter what kernels they employ, recalls from the same image kinds are all averaged.

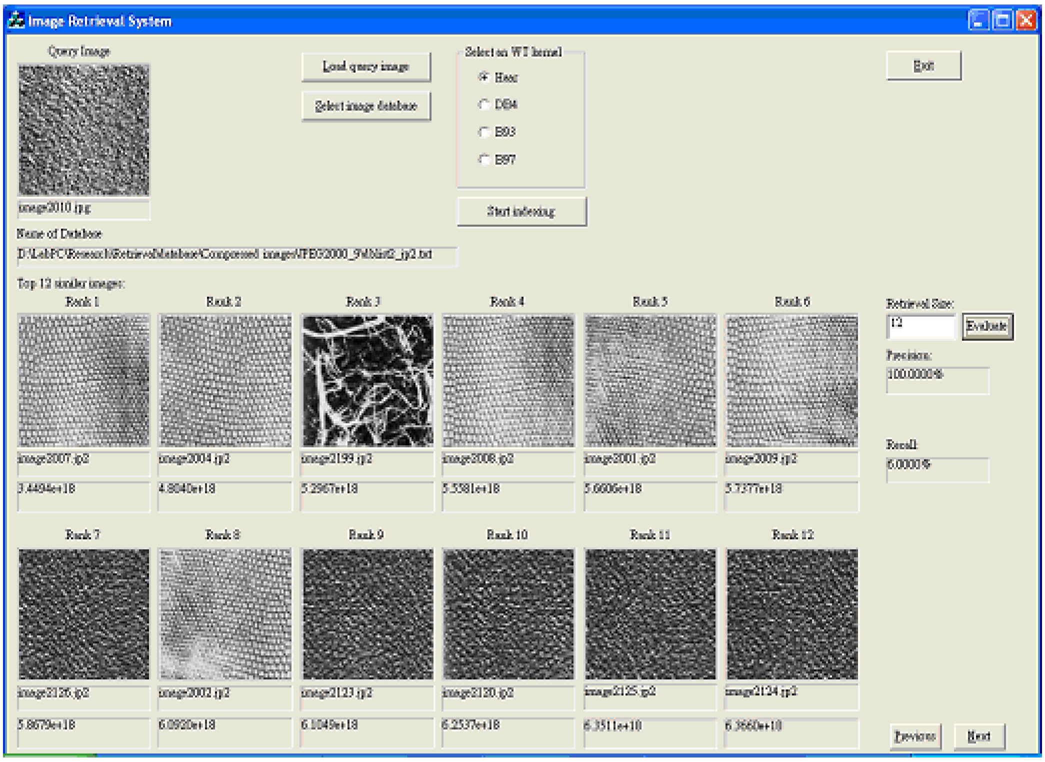
In the first 200 ranks, most of the image kinds miss more than half of their relevant images at small compression ratios. Besides buildings and random texture, average recalls of all other image kinds fluctuate when the compression ratio is less than 20 and slightly increases afterwards. In contrast, average recalls of building and random texture increase when the compression ratio is less than 30 and 10 respectively and then drop for more than 20% when the compression ratio is 72. The indexing result of random texture is provided in figure 6.10. Fluctuation is mainly due to compression. When it is small, images from these kinds may become indistinguishable. When it is large enough, details of each image kind are lost seriously which cause those image kinds to be distinguishable to other kinds. Same as the analysis in precision, human face and single object images give outstanding recall performance. This is due to their special image characteristics so that their feature vectors can easily be identified. Pattern texture gives the worst recall results because compression blurs its characteristics seriously. In summary, the special characteristics of images result in different retrieval accuracy under different compression ratios.



(a)



(b)



(c)

Figure 6.10: Retrieval results of random texture images when compressed at (a) 2:1, (b) 19:1 and (c) 27:1.

6.3.2 Discussion on Recall versus retrieval size

Figure 5.12 plots the average recall of each wavelet kernel versus retrieval size. It is obtained from averaging the recall from searching every image kind (i.e. K=200) and over all compression ratios. Since each image catalogue contains 200 images, it is desirable that recall approaches 100% as the retrieval size tends to 200. In general, the average recall of all image kinds increase steadily as retrieval size increases. In particular, the recall of face image out-performs other image kinds, which rises to 96% when retrieval size is 200. This is because the special characteristics of human face images are preserved even under high compression ratios, as proven in subsection 6.3.1. Thus, almost all of the human face images can be retrieved successfully. Besides human face, single object images give excellent recall results that rises to 60%. Over half of the single object images can be retrieved successfully when the retrieval size is 200. Besides human face and single object images, the average recalls of the other image kinds approach 50% when retrieval size is 200. This means that different compression schemes such as JPEG and JPEG2000 affect the common characteristics that can be preserved. Particularly, the average recall of texture pattern is the lowest. Fine details of texture pattern are blurred by JPEG and JPEG2000 seriously. The average recall of landscape and natural images are very close. Due to the loss of their fine details, they become indistinguishable. This agrees with our findings in subsection 6.2.2 that the precision of these two image kinds is very close.

6.3.3 Discussion on recall versus retrieval size in usage of kernels

Figure 5.13 describes the average recall versus retrieval size using different wavelet kernels. All the four recall curves increase steadily from nearly 0% to over 40%. This means that slightly less than half of the relevant images are retrieved by the proposed indexing algorithm when 200 images are considered. Haar kernel gives the best results, followed by DB4, B97 and B93. The results confirm our analysis in section 6.2.3 about the precision performance of the four wavelet kernels.

6.4 Comparison between proposed and existing retrieval systems

Our proposed retrieval system use a 35D translation and rotation invariant feature vector to characterize the features of the images in three aspect, they are: energy, significance map and normalized central moments. To quantify the efficiency of our proposed retrieval system, we will compare our retrieval system with other existing retrieval algorithm in this sub-section. Xiong proposed a subband-based retrieval system which extracts features from image subbands to form its feature vector [73]. In Xiong's system, compressed images are partially decoded where variances are extracted from their subband coefficients. Mean *E[Y]* and variance *Var(Y)* of the *j*th subband are defined as,

$$E[Y_j] = \frac{\sum_{i=1}^{N_j} y_i^j}{N_j} \tag{6.1}$$

$$Var(Y_j) = \frac{\sum_{i=1}^{N} (y_i^j - E[Y_j])^2}{N_j} \tag{6.2}$$

where N_j is the number of coefficients in the *j*th subband.

To compare our proposed system with Xiong's system, features are extracted from the concatenated BDCT and WT subbands since both systems work on subbands. In our analysis, three-level decomposition is used thus results in 10 subbands. Thus, Xiong's algorithm generates a 10D feature vector *f* as,

$$f' = \{Var(Y_j)\} \qquad \text{for } j = \{0,1,\ldots,9\} \tag{6.3}$$

Same distance measure $D_{p,q}$ as stated in section 5.5, is used to measure the distance between the query and indexing images.

$$D_{p,q} = \sum_{k=1}^{10} (f'^{q}_{k} - f'^{p}_{k})^2 \tag{6.4}$$

To show the structural difference between our proposed and Xiong's retrieval sytem, figures 6.11 and 6.12 outline their processes. The only difference between them are their feature extraction algorithms. Therefore, our and Xiong's systems generate 35D and 10D feature vectors respectively. In our simulation, same image database is used, which consists of 1800 images from 9 image kinds compressed at the same set of compression ratios. JPEG query images will look for the most similar JPEG2000 images as what we did in sections 5.6 and 5.7. Average precision and recall are calculated to compare the two systems, where the average values are obtained from all compression ratios and with all kernels. The average precision and recall of the two systems are plotted in figure 6.13 and 6.14 respectively.

From the two figures, we observe that our proposed retrieval system gives better retrieval performance than Xiong's retrieval system. Both precision and recall of our proposed system are always 15% above that of Xiong's system. Thus, more relevant images can be searched by using our set of feature vector. The main reason is that our feature vector is translation and rotation invariant, which consists of energy and normalized central moments. Therefore, even images bear similar contents but shift in position, our feature vector can still identify the similar features between them. However, Xiong proposed variance as its feature vector, which is a translation and rotation variant feature. This is because variance is the moment that is not calculated with respect to its center of mass. Therefore, it suffers from the shift of centroid and results in false indexing candidates.

In summary, a comparison is done between our proposed and an existing image retrieval system. The two systems extract different features from the same WT subband structure. Experimental results show that our feature vector can efficiently search for more relevant images as it uses translation and rotation invariant feature vector.

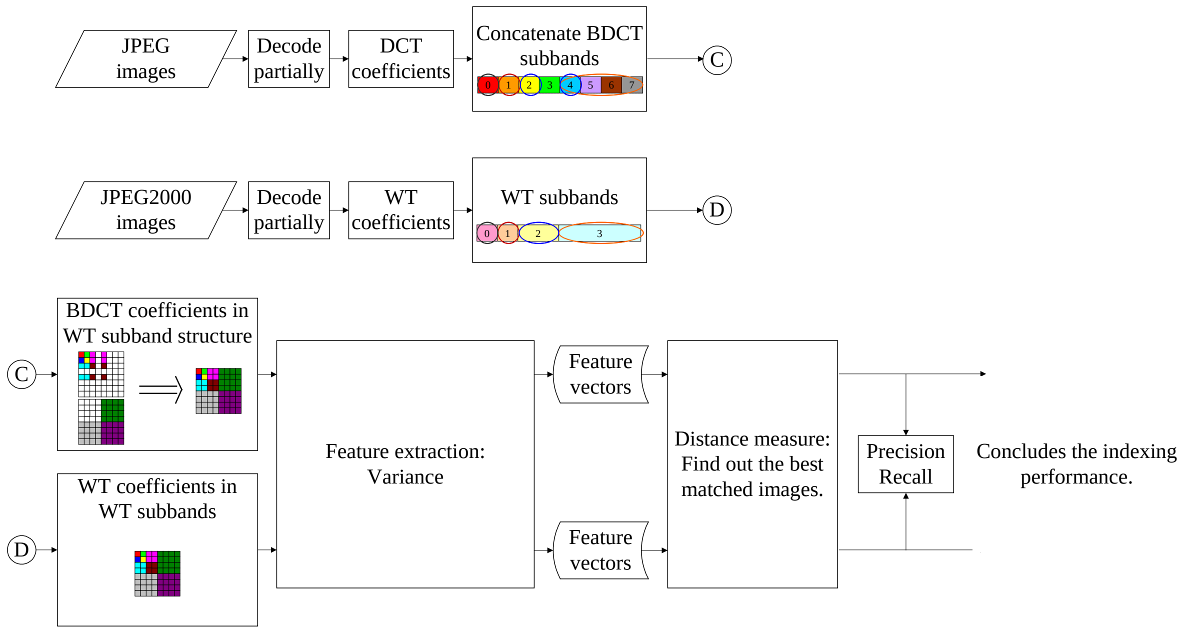


Figure 6.11: Flowchart of our proposed retrieval system.

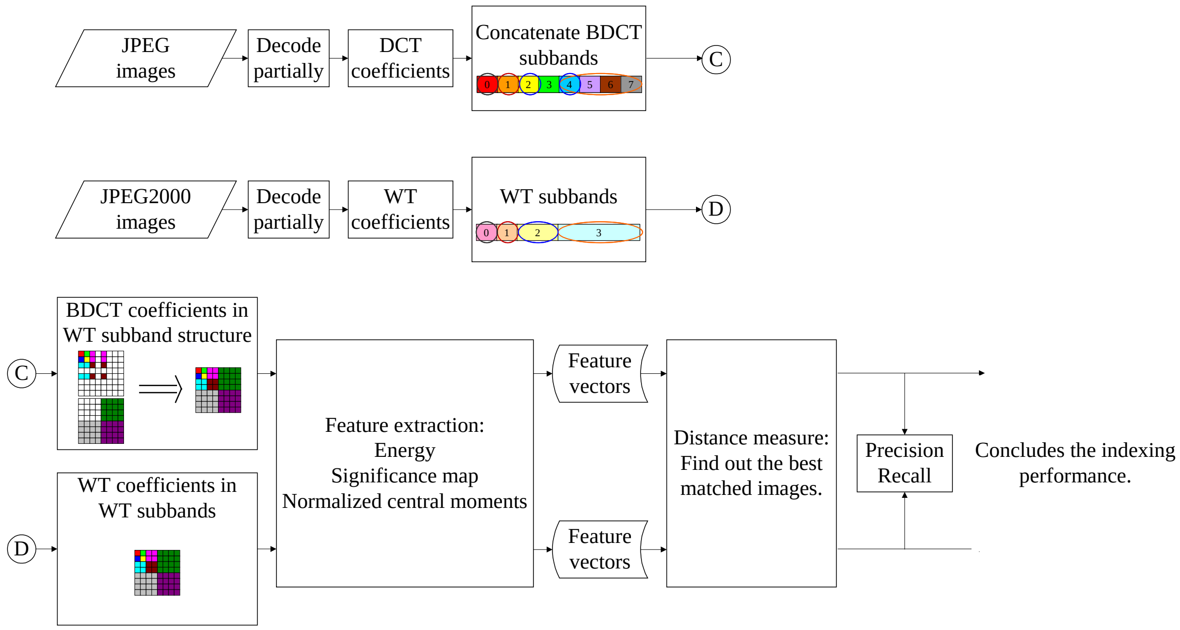


Figure 6.12: Flowchart of Xiong's retrieval system.

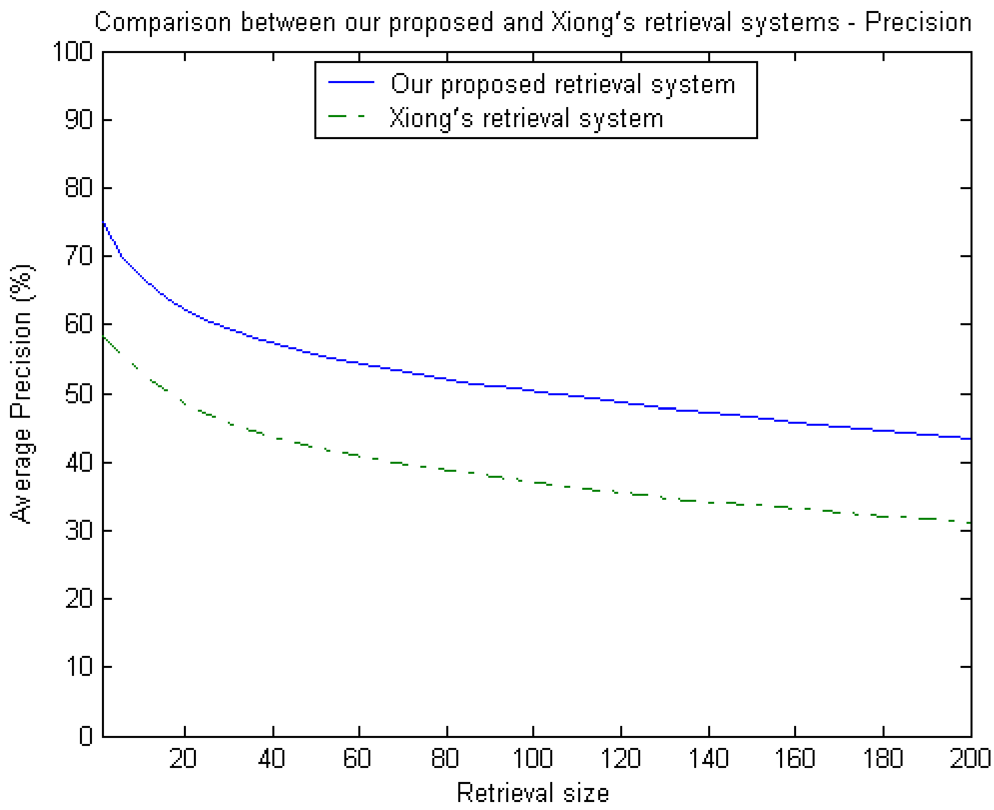


Figure 6.13: A plot of precision between our proposed and Xiong's retrieval systems.

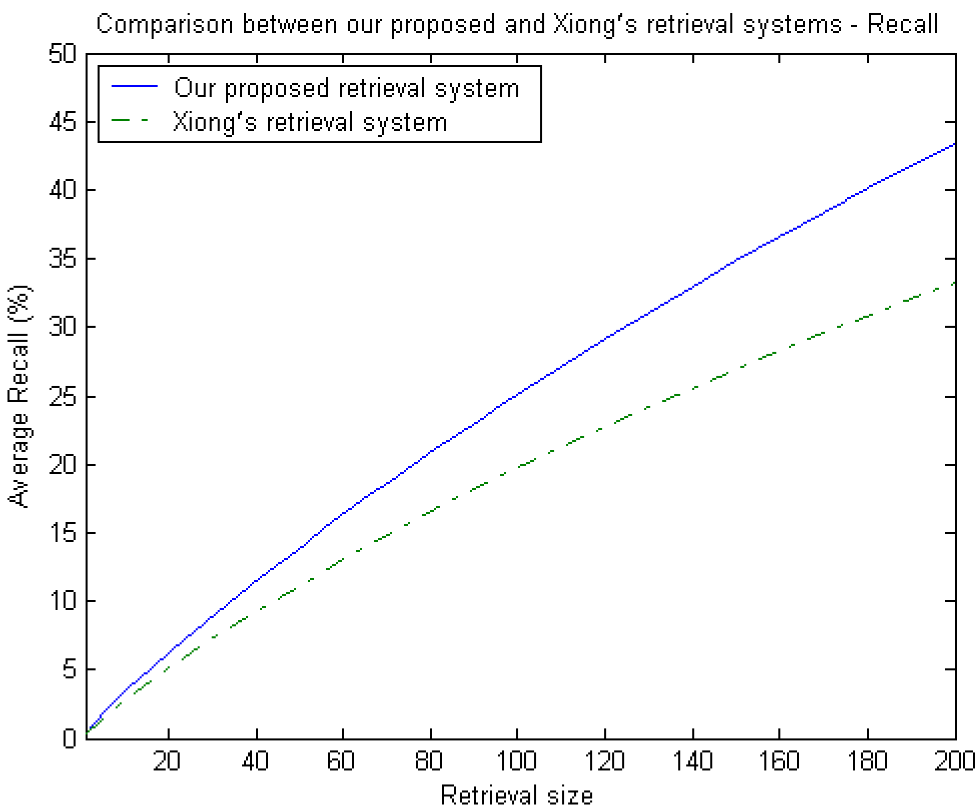


Figure 6.14: A plot of recall between our proposed and Xiong's retrieval systems.

6.5 Chapter summary

In this section, simulation results of our proposed image indexing algorithm in different compressed domains is discussed. Retrieving performance versus compression ratio, retrieval size and wavelets kernels have been analyzed. Figure 6.15 concludes our analysis in Chapter 5 and 6.

Since JPEG and JPEG2000 compression schemes are different from transformation to quantization, we study how compression influences our retrieval performance. Precision and recall versus compression ratio conclude that compression reduces the number of correctly retrieved images. Owing to the special characteristics of various image kinds such as human face, single objects and trade-mark images, they provide excellent retrieval accuracy at all compression ratios. Since JPEG2000 preserves more spectral characteristics than JPEG, their outputs become dissimilar after compression. Thus, high compression ratio degrades the retrieval accuracy as some image kinds are indistinguishable after compression. When compression ratio further increases, fine details in images are seriously blurred so as to make them distinguishable from one another again. This causes fluctuation in precision and recall values. Although compression influences the retrieval precision, nearly half of relevant images can be retrieved at all compression ratios.

To quantify how much relevant information can be retrieved, retrieval performance versus retrieval size is studied. Precision and recall versus retrieval size suggest that JPEG images can search for similar JPEG2000 images when either one of the four commonly used JPEG2000 wavelet kernels, i.e., Haar, Daubechies 4, Biorthogonal 93 and Biorthogonal 97 kernels, are used. For some special image kinds such as human face, single object and trade-mark images, excellent retrieval performance can be achieved. For other image kinds, good retrieval results can be obtained even under high compression ratio.

The usage of wavelet kernels in JPEG2000 also influences the retrieval accuracy of searching JPEG images. Particularly, JPEG2000 images that are compressed by Haar kernel can retrieve a large number of similar JPEG images. Even for images compressed by the JPEG2000 Part I algorithm (i.e., Biorthogonal 97 kernel) at high compression ratio, nearly half of relevant JPEG images can be retrieved. It also reveals that, no matter at what compression ratio, most image kinds can accurately search for more than 100 relevant images out of the 200 retrieved images.

The simulation results conclude that common features can be extracted in JPEG and JPEG2000 compressed domains such that relevant images can be searched from the two compressed domains. Thus, our proposed image indexing algorithm proves that image searching in different compressed domains is possible without a full decompression.

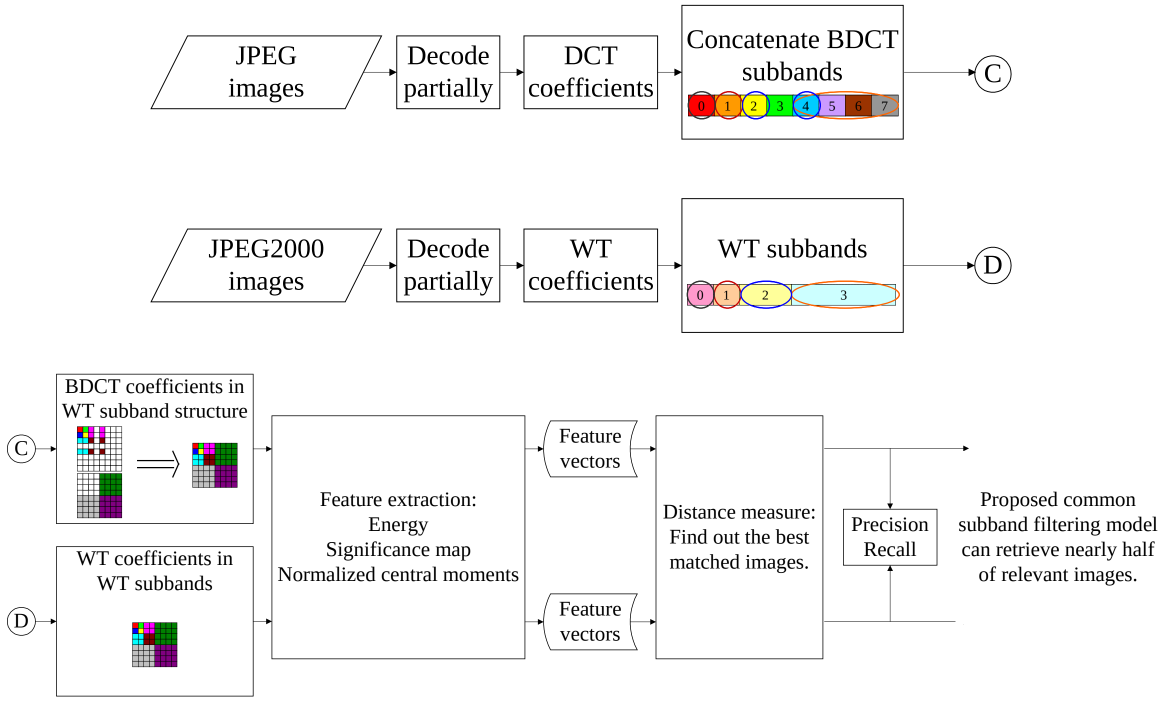


Figure 6.15: Work flows and summaries of our proposed image retrieval system.

Chapter 7

Conclusion

7.1 Introduction

A large amount of visual information requires correspondingly a large storage space. As a result, information is usually compressed to reduce the storage requirement. Many image compression schemes have been proposed in the past decade. An overview of some existing compression schemes is given. JPEG is the most commonly used compression scheme nowadays. JPEG2000 is newly released as an improvement of JPEG so as to tackle the artifacts of JPEG at high compression ratios. JPEG2000 retains better image quality and provides stronger compression power than JPEG. It is forecast that JPEG2000 will be widely applied in multimedia applications.

Image retrieval is a multimedia requirement for retrieving relevant images from large image archives. Text-based image retrieval methods are proposed in the early stage, but unreliable keywords searching greatly affects their accuracy. Content-based image retrieval systems in both spatial and compressed domains are then investigated to search for relevant candidates by their image characteristics. However, most of these systems retrieve images in single domain, in other words, query images can only search for candidates in the same image format. In modern multimedia systems, images in various domains are stored inside the same system. To find out relevant candidates in different domains, it is inefficient to reconvert all images into the same domains, especially for large image archives. Existing problems of current image retrieval systems in multiple domains are addressed in chapter 1. As a matter of effective image indexing in different domains, retrieval systems working in several domains are most desirable.

As JPEG and JPEG2000 are the current and future main-stream image compression schemes respectively, our focus is put on image indexing in these two domains. Both compression schemes transform images from spatial to spatial-frequency domains. In the course of compression, JPEG and JPEG2000 employ dissimilar Block-based Discrete Cosine Transform (BDCT) and Wavelet Transform (WT) to extract their frequency characteristics as described in chapter 2. Indexing in the two compressed domains becomes time consuming if all images are decompressed to their original spatial domain. To cope with this problem, a novel image retrieval system in different domains is proposed in this study. In the following section, we will summarize the similarities between JPEG and JPEG2000 compression schemes. We will also conclude our findings of common feature extraction between these two domains. Finally, we will describe some future directions in multimedia applications.

7.2 Conclusion on current works

The difference between JPEG and JPEG2000 begins from their transformation schemes. The origins and the design of BDCT and WT filters employed by JPEG and JPEG2000 are quite different. Direct comparison on their filter outputs cannot expose their relationship. By employing our derived subband filtering model in chapter 3, filters in BDCT and WT can be compared directly. Our mathematical derivations show that their filtering operations can be constructed by a subband filtering model, i.e., both consist of lowpass / bandpass filtering following by a downsampling operation. Following the JPEG2000 Part I and II algorithms, commonly used wavelet kernels (i.e., Haar, Daubechies 4, Biorthogonal 93 and Biorthogonal 97) are used to investigate the WT subband filtering model.

To verify our theoretical model, we use three filter characteristics, i.e., magnitude spectrum, passband region and difference in energy preservation, to analyze whether common features exist in the BDCT and WT domains. Our comparison shows that lowpass filtering in WT and BDCT are very similar and indeed the same when the Haar wavelet is used. Due to the passband difference in the bandpass filtering, direct comparison is inappropriate. In contrast, after concatenating several BDCT filters, the resultant BDCT filters become comparable to WT filters. Their resultant filtering models can be divided into four subband filters, namely lowpass, bandpass 1, bandpass 2 and bandpass 3. Although tiny passband difference is observed between Biorthogonal 93 and the resultant BDCT filter at bandpass 1, a high similarity value is obtained between them. Our similarity measures on magnitude spectrum and energy preservation prove that their subband models are in fact extremely similar.

To validate BDCT and WT subband filtering models, their outputs are further analyzed in chapter 4. Four kinds of images, which consist of characteristics such as texture, edges and plain areas, are used to examine the compression effects on outputs. The difference between JPEG and JPEG2000 compression schemes certainly degrades their output similarity. Our investigation thus involves compressed natural scene, man-made object, structured texture and random texture images at six compression ratios ranging from 6:1 to 44:1. To be more specific, JPEG2000 images are compressed by Haar, Daubechies 4, Biorthogonal 93 and Biorthogonal 97 wavelet kernels, and then partially decoded. Similarity measure on the resultant BDCT and WT output spectra proves that similarity values slightly decrease as compression ratio increases. This is because more information is retained by JPEG2000 than that by JPEG at high compression ratios, resulting in reduction of similarity. Even though compression reduces the similarity between JPEG and JPEG2000, they preserve quite similar spectral characteristics. In particular, extremely high similarity is observed at the lowpass output for all kernels.

Evaluation of our derived subband filtering model in a practical retrieval system is performed in chapter 5. Features are extracted from JPEG and JPEG2000 subband outputs for image indexing. In accordance with JPEG and JPEG2000 compression algorithms, their spectral outputs are organized in 8×8 blocks and subbands respectively. In order to extract common features from their spectral outputs for further feature extraction, the BDCT coefficients are concatenated into WT subband structure. This concatenation is in fact included in the proposed subband filtering model. The resultant BDCT subband structure is comparable to the WT subband structure, and thus three common characteristics can be extracted from the subbands, namely energy, significance map and normalized central moments of significance map. Energy reveals subband importance. Thresholding technique, which aims to screen out insignificant coefficients but retains those dominant, constructs the significance map effectively. Applying second and third normalized central moments on the significance map can extract translation and rotation invariant features properly. Finally, the three characteristics provide a 35 dimensional feature vector for further image retrieval.

To examine our compressed domains retrieval algorithm, a user-friendly image retrieval system and a database is established. Our database consists of one thousand and eight hundred images from 9 catalogues and are all compressed by JPEG and JPEG2000 at seven compression ratios ranging from 1.6:1 to 72:1. These nine kinds of images consist of plain areas, edges, texture and man-made graphics, which contain both low and high frequency contents. 35 dimensional features are extracted from these images. To search for relevant images with respect to the query, Euclidean space measure is used to quantify the distance between their feature vectors. The distance values between query and retrieved candidates are sorted in descending order and displayed on the system interface. By the use of precision and recall, we can accurately evaluate the retrieval performance of our proposed system. Our simulation results show that the compression ratio slightly influences the retrieval accuracy. In addition, the proportion of relevant candidates varies as the retrieval size increases. Furthermore, the indexing performance depends on the usage of WT compression kernels, i.e., Haar, Daubechies 4, Biorthogonal 93 and Biorthogonal 97 wavelet kernels.

Simulation results are discussed in chapter 6. Owing to the special characteristics of various image kinds such as human face, single object and trade-mark images, they provide excellent retrieval accuracy at all compression ratios. Since JPEG2000 preserves more spectral characteristics than JPEG, their outputs become dissimilar after compression. Thus, high compression ratio degrades the retrieval accuracy as some image kinds are indistinguishable after compression. When compression ratio further increases, fine details in images are seriously blurred so as to make them distinguishable from one another again. This causes fluctuation in precision and recall values. Although compression influences the retrieval precision, nearly half of relevant images can be retrieved at all compression ratios. No matter images are compressed by JPEG2000 part I (i.e., Biorthogonal 97 kernel) or part II algorithms (other kernels) at all compression ratios, they can provide a retrieval of a large number of similar JPEG images.

Particularly, a great number of relevant JPEG images can be searched by JPEG2000 images when the Haar kernel is used. A comparison between our proposed and existing retrieval systems is done. Our simulation shows that using our set of feature vector, better retrieval results can be obtained than other retrieval systems.

Our throughout investigation concludes that common features can be extracted in JPEG and JPEG2000 compressed domains. Relevant images can be searched from the two compressed domains no matter which compression kernel is used and at what compression ratio. Both of our theoretical and experimental studies prove that direct image indexing in different domains is possible without a full decompression. Our derived filtering model can be applied to Discrete Cosine Transform and Wavelet Transform based compression schemes for retrieval application. Applications and future development of our proposed retrieval algorithm in different compressed domains will be presented in the next subsection.

7.3 Future development

In this subsection, we will present the future development of our research. Our future development will focus on three areas: color image retrieval systems for video applications and medical diagnosis in different domains. Retrieval system in different domains for color images and video sequences provide convenience for end-users to search relevant information from different sources. Retrieval system for medical diagnosis can bring valuable contribution to medical treatment. These systems will be briefly described in the following sub-sections.

7.3.1 Color image retrieval system in different domains

Our proposed image retrieval algorithm in different domains can search for relevant JPEG and JPEG2000 images from each other. To look for similar color images, we need to consider both luminance and chrominance components. Although no color scheme is specified in JPEG and JPEG2000, color images other than YCbCr format are usually transformed to this color scheme for further compression. Thus, JPEG and JPEG2000 images are already in the same color planes, they do not undergo any other color transformation. Many color retrieval systems have been proposed [32,45,5]. However, they are sensitive to illumination changes than those in chrominance, and usually emphasize the luminance characteristics [45,5]. Moreover, they discard their vector dimensionality [32] and regional spatial correlation between pixel colors [44,34]. As a result, the retrieval results sometimes cannot search images bearing the same chrominance characteristics.

To cope with this, two different feature descriptors can describe the partially decompressed luminance and chrominance information. In order to extract luminance characteristics effectively, our proposed feature descriptor consisting of energy and normalized central moments can be applied. To extract the chrominance feature, fuzzy color histogram (FCH) can be applied to the Cb and Cr plane respectively [39]. However, the proposed FCH algorithm works in the pixel domain to find out similarity between color pixels by spreading the contribution of each pixel to all histogram bins. The effectiveness and robustness of this algorithm in BDCT and WT transformed domains require further investigation. In our future development, a new subband level FCH could be developed to take into account of coefficients at all subbands. The new color descriptor should be able to extract translation and rotation invariant color features and form a small size feature vector. The resultant feature descriptors can provide a fast computation and a robust color image retrieval system for real time applications.

7.3.2 Motion JPEG and JPEG2000 retrieval systems

Owing to the advance in technologies, consumer electronics such as digital camera can capture video conveniently. JPEG is widely applied in current electronic devices to compress images and videos into JPEG and Motion JPEG respectively. Motion JPEG indeed contains a sequence of JPEG images to form a video sequence [15]. Example application includes real-time Motion JPEG transmission for medical images[11]. Due to the strong compression power and good compressed quality, JPEG2000 is forecast as the next main-stream image compression scheme. Similar to Motion JPEG, Motion JPEG2000 will be integrated into multimedia systems to compress sequences of JPEG2000 images [60]. Thus, systems may store a large amount of videos in various format, such as Motion JPEG and Motion JPEG2000. Users may want to search for video sequences with relevant contents. A retrieval system in different video compressed domains is then desirable.

A sequence of Motion JPEG and Motion JPEG2000 consists of several or even more images. Video is captured in a constant frame rate (for example, 15 frames/second). During indexing, frame-by-frame comparison incurs a huge computational time. Also, the high temporal redundancy between frames expends unnecessary computational resources. To speed up the process, a key-frame can be selected to represent a sequence of images. As high temporal redundancy exists between frames in a particular duration, a number of key-frames can be extracted to provide coarse information about the whole video sequence. To retrieve relevant video sequences, our proposed image retrieval algorithm can be applied to the key-frames. Luminance and chrominance features can be extracted from the key-frames for similarity measures. By comparing their distances, relevant video sequences are sorted according to their similarity. Similar Motion JPEG and Motion JPEG2000 videos can then be retrieved.

7.3.3 Retrieval on medical images

Medical imaging requires a large storage of high quality clinical data such as sonogram, angiogram, computed tomography and X-ray images. For practical diagnosis, the quality of the compressed medical images should be acceptable even under the constraint of transmission bandwidth and storage [54]. Over the past decade, most medical images and videos are commonly compressed by JPEG. However, the artifacts of JPEG degrade the visual quality of images and videos. JPEG2000 provides a solution to compress images at high compression ratios with acceptable quality [71]. Thus, medical systems may compress images and video sequences into either JPEG or JPEG2000 domains. For diagnostic purposes, retrieval on medical images and videos become a valuable application for case references. To search for relevant images, our proposed image and video retrieval algorithms can be combined so as to facilitate searching in different domains. This retrieval system would benefit practical diagnosis as similar medical images can be search efficiently.

References

[1] A. Gupta and R. Jain, "Visual Information Retrieval", *Comm. ACM.*, Vol. 40, No. 5, pp. 71-79, 1997.

[2] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *SPIE Storage and Retrieval Image and Video Database II*, 1995.

[3] Agostine Abbate, Casimer M. DeCusatis and Pankaj K. Das, *Wavelets and Subbands: Fundamentals and Applications*, ISBN 081764136X, Birkhauser, 2002

[4] Asha Vellaikal and C. C. Jay Kuo, "Joint spatial-spectral indexing form image retrieval," *International Conference on Image Processing*, Vol. 3, pp. 867-870, 16-19 Sept. 1993

[5] B. V. Funt and G. D. Finlayson, "Color constrant color indexing," *Pattern Anal. Machine Intell.*, Vol. 17, No. 5, pp. 522-529, 1995

[6] Bryan E. Usevitch, "A Tutorial on Modern Lossy Wavelet Image Compression: Foundations of JPEG 2000," *IEEE Signal Processing Magazine*, Vol. 18, Issue 5, pp. 22-35, Sept. 2001

[7] C. Carson and V.E. Ogle, "Storage and Retrieval of Feature Data for a Very Large Online Image Collection," *IEEE Data Eng. Bull.*, Vol. 19, No. 4, Dec. 1996

[8] C. C. Chang; J.C. Chuang; Y.S. Hu, "Retrieving digital images from a JPEG compressed image database," *Image and Vision Computing*, Vol. 22, pp.471-484, 2004

[9] C. W. Ngo; T. C. Pong; R. T. Chin, "Exploiting image indexing techniques in DCT domain," *Pattern Recognition*, Vol. 34, pp.1841-1851, 2001

[10] D. Feng, W. C. Siu and H.J. Zhang, *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, ISBN 3540002448, Springer, Chapter 1, 2003

[11] D. Nishantha, Y. Hayashida and T. Hayashi, “Application Level Rate Adaptive Motion-JPEG Transmission for Medical Collaboration Systems,” *24th International Conference on Distributed Computing Systems Workshops*, pp. 64 – 69, 23-24 March 2004

[12] D. G. Sim, H. K. Kim and R. H. Park, “Fast texture description and retrieval of DCT-based compressed images,” *Electronics Letters*, Vol. 37, Issue 1, pp. 18-19, 4 Jan. 2001

[13] D. S. Wu and L. N. Wu, “Image Retrieval Based on Subband Energy Histograms of Reordered DCT Coefficients,” *International Conference on Signal Processing*, Vol.1 pp. 596-599, 26-30 Aug. 2002

[14] David S. Taubman and Michael W. Marcellin, *JPEG2000 : image compression fundamentals, standards and practice*, ISBN 079237519X, Kluwer Academic Publishers, 2002

[15] Digital Camera, [Online] Available:
http://www.jpeg.org/apps/cinema.html?langsel=en

[16] Emmanuel C. Ifeachor and Barrie W. Jervis, *Digital Signal Processing: A Practical Approach, Addison-Wesley*, ISBN 020154413X, 1993

[17] G. Feng; J. Jiang, “JPEG compressed image retrieval via statistical features,” *Pattern Recognition*, Vol. 36, pp.977-985, 2002

[18] G. Pass and R. Zabith, “Comparing images using joint histograms,” *Multimedia Systems*, Vol. 7, pp.234-240, 1999

[19] G. Van de Wouwer, P. Scheunders and D. Van Dyck, “Statistical Texture Characterization from Discrete Wavelet Representations,” *IEEE Transactions on Image Processing*, Vol. 8, Issue 4, pp. 592-598, April 1999

[20] Gregory K. Wallace, “The JPEG Still Picture Compression Standard,” *IEEE Transactions on Consumer Electronics*, Vol. 38, pp.18-36, Feb. 1992

[21] H. Tamura, and N. Yokoya, “Image database systems: A survey,” *Pattern Recognition,* Vol.17, No.1, pp.29-43, 1984

[22] H. A. Moghaddam, T. T. Khajoie and A. H. Rouhi, “A New Algorithm for Image Indexing and Retrieval Using Wavelet Correlogram,” *International Conference on Image Processing*, Vol. 3, pp. 497-500, 14-17 Sept. 2003

[23] Hong Yan, “Image analysis for digital media applications,” *IEEE Computer Graphics and Applications*, Vol. 21, Issue 1, pp.18–26, Jan.-Feb. 2001

[24] Ingrid Daubechies, *Ten Lectures on Wavelets*, ISBN 0898712742, Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 1992

[25] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu and R. Zabih, “Image Indexing Using Color Correlograms”, *IEEE Computer society Conference on Computer Vision and Pattern Recognition*, pp. 762-768, 17-19 June 1997

[26] J. Jiang, A. Armstrong, G.C. Feng, “Direct content access and extraction from JPEG compressed images”, *Pattern Recognition*, Vol. 35, Issue 11, pp.2511-2519, Nov. 2002

[27] J. Wei, “Color Object Indexing and retrieval in digital libraries,” *IEEE Transactions on Image Processing*, Vol. 11, Issue 8, pp. 912-922, Aug. 2002

[28] J. Zhu, “Image compression using wavelets and JPEG2000: a tutorial,” *Electionics and Communication Engineering Journal*, Vol. 14, Issue 3, pp. 112-121, June 2002

[29] J. L. Shih and L. H. Chen, “Color image retrieval based on primitives of colour moments,” *IEE Proceedings on Vision, Image and Signal Processing*, Vol. 149, Issue. 6, Dec. 2002

[30] J. Z. Wang, G. Wiederhold, O. Firschein and S. X. Wei, “Content-based image indexing and searching using Daubechies’ wavelets,” *International Journal on Digital Libraries*, Vol. 1, pp.311-328, 1997

[31] John R. Smith and S. F. Chang, “Transform Features for Texture Classification and Discrimination in Large Image Databases,” *Proc. Of ICIP*, pp. 407-411, 1994

[32] John R. Smith and S. F. Chang, “Single color extraction and image query,” *International Conference on Image Processing*, Vol. 3, pp. 528-531, 23-26 Oct. 1995

[33] John R. Smith and S. F. Chang, “VisualSEEK: A Fully Automated Content-Based Image Query System,” *ACM Multimedia*, 1996

[34] John R. Smith and S. F. Chang, “Tools and techniques for color image retrieval,” *Proc. SPIE*, Vol. 2670, pp. 1630-1639, Feb. 1996

[35] John R. Smith and S. F. Chang, “Local Color and Texture Extraction and Spatial Query,” *International Conference on Image Processing*, Vol. 3, pp.1011-1014, 16-19 Sept. 1996

[36] JPEG Homepage, [Online] Available: http://www.jpeg.org/jpeg/index.html

[37] JPEG2000 Homepage, [Online] Available:
http://www.jpeg.org/jpeg2000/index.html

[38] JBIG Homepage, [Online] Available: http://www.jpeg.org/jbig/index.html

[39] Ju Han and K.K. Ma, “Fuzzy color histogram and its use in color image retrieval,” *IEEE Transactions on Image Processing*, Vol. 11, Issue 8, pp. 944-952, Aug. 2002

[40] Kai-Chieh Liang and C. C. Jay Kuo, “Progressive Image Indexing and Retrieval Based on Embedded Wavelet Coding,” *International Conference on Image Processing*, Vol. 1, pp. 572-575, 26-29 Oct. 1997

[41] Kai-Chieh Liang and C. C. Jay Kuo, “Wavelet-Compressed Image Retreival Using Successive Approximation Quantization (SAQ) Features,” *SPIE Voice, Video, and Data Comm.*, Nov. 1997

[42] Kai-Chieh Liang and C. C. Jay Kuo, ”WaveGuide: a joint wavelet-based image representation and description system,” *IEEE Transactions on Image Processing*, Vol. 8, Issue 11, pp.1619-1629, Nov. 1999

[43] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Qian Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker, “Query by Image and Video

Content: the QBIC System," *IEEE Transactions on Computer*, Vol. 28, Issue 9, pp. 23-32, Sept. 1995

[44] M. Stricker and A. Dimai, "Spectral covariance and fuzzy regions for image indexing," *Mach. Vis. Applicat.*, Vol. 10, pp. 66-73, 1997

[45] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, Vol. 7, No. 1, pp. 11-32, 1991

[46] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory*, pp.179-187, 1962

[47] M. K. Mandal, T. Aboulnasr and S. Panchanathan, "Image Indexing Using Moments and Wavelets," *International Conference on Image Analysis and Processing*, pp.904-908, 27-29 Sept. 1999

[48] Mark Nixon and Alberto Aguado, *Feature Extraction and Image Processing*, ISBN 0750650788, Newnes, 2002

[49] Martin Vetterli and Jelena Kovacevic, *Wavelets and Subband Coding*, ISBN 0120970808, Prentice Hall, 1995

[50] Michael Shneier and Mohamed Abdel-Mottaleb, "Exploiting the JPEG compression scheme for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, Issue 8, pp. 849-853, Aug. 1996

[51] Michael W. Marcellin, Michael J. Gormish, Ali Bilgin, Partin P. Boliek, "An overview of JPEG-2000," *Proc. Data Compression Conference*, pp. 523 – 541, 28-30 March 2000

[52] Mihaela D. Pal, Christopher M. Brislawn and Steven P. Brumby, "Feature Extraction from Hyperspectral Image Compressed Using the JPEG-2000 Standard," *IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 168-172, 7-9 April 2002

[53] Niles Ritter, *What is TIFF?*, [Online] Available:
http://home.earthlink.net/~ritter/tiff/

[54] R. A. Foulds, "Biomechanical and perceptual constraints on the bandwidth requirements of sign language," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 12, Issue 1, pp. 65-72, March 2004

[55] Ruey-Feng Chang; Wen-Jia Kuo; Hung-Chi Tsai, "Image retrieval on uncompressed and compressed domains," *International Conference on Image Processing*, Volume 2, pp. 546 –549, 10-13 Sept. 2000.

[56] S. Climer; S. K. Bhatia, "Image database indexing using JPEG coefficients," *Pattern Recognition*, Vol. 35, pp.2479-2488, 2002

[57] S. Deb and Y. Zhang, "An overview of content-based image retrieval techniques," *International Conference on Advanced Information Networking and Applications*, Vol. 1, pp. 59-64, 29-31 March 2004

[58] S. Liapis and G. Tziritas, "Color texture image retrieval using chromaticity histograms and wavelet frames," *IEEE Transactions on Multimedia*, Vol. 6, Issue 5, pp. 676-686, Oct. 2004

[59] S. K. Chang, and A. Hsu, "Image information systems: where do we go from here?" *IEEE Transactions on Knowledge and Data Engineering*, Vol.5, No.5, pp.431-442, Oct. 1992

[60] Siegfried Foessel, Motion JPEG2000 and Digital Camera, [Online] Available: http://www.jpeg.org/public/DCINEMA-JPEG2000.ppt

[61] Suryani Lim and Guojun Lu, "Spatial Statistics for Content Based Image Retrieval", *International Conference on Information Technology: Coding and Computing [Computers and Communications]*, pp.155-159, 28-30 April 2003

[62] T. Chang and C. C. Jay Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transform," *IEEE Transactions on Image Processing*, Vol. 2, Issue 4, pp. 429-441, Oct. 1993

[63] T. Kato, T. Kurita and K. Hirata, "A sketch retrieval method for full color image database-query by visual example," *11th IAPR International Conference on Computer Vision and Applications*, Vol. 1, pp. 530-533, 30 Aug. – 3Sept. 1992

[64] Theo Gevers and Arnold W. M. Smeulders, "PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval," *IEEE Transactions on Image Processing*, Vol. 9, Issue 1, pp. 102-119, Jan. 2000

[65] Voulgaris G., Jiang J., "Texture-based image retrieval in wavelets compressed domain", *Proceedings of International Conference on Image Processing*, Vol. 2, pp.125-128, Oct. 2001

[66] W. I. Grosky, R. Jain, and R. Mehrotra, *The Handbook of Multimedia Information Management*. Englewood Cliffs, ISBN 0132073250, NJ: Prentice-Hall, 1997

[67] William B. Pennebaker and Joan L. Mitchell, *JPEG: Still Image Data Compression Standard*, ISBN 0442012721, Van Nostrand Reinhold, Chapter 4, 1993

[68] William K. Pratt, *Digital Image Processing (2nd Edition)*, ISBN 0471857661, John Wiley & Sons, Chapter 2, 1991

[69] X. Liu and D. Wang, "Texture Classification Using Spectral Histograms," *IEEE Transactions on Image Processing*, Vol. 12, Issue 5, pp. 661-670, June 2003

[70] Y. Gong, H. J. zhang, and T. C. Chua, "An image database system with content capturing and fast image indexing abilities", *Proc. IEEE International conference on Multimedia Computing and Systems*, Boston, pp.121-130, May 1994

[71] Yani Zhang, B. Pham and M.P. Eckstein, "Evaluation of JPEG 2000 encoder options: human and model observer detection of variable signals in X-ray coronary

angiograms," *IEEE Transactions on Medical Imaging*, Vol. 23, Issue 5, pp. 613-632, May 2004

[72] Yves Nievergelt, *Wavelets made easy*, ISBN 0817640614, Boston: Birkhäuser, 1999

[73] Z. Xiong and Thomas S. Huang, "Subband-based, Memory-efficient JPEG2000 Images Indexing in Compressed-domain," *IEEE Southwest Symposium on Image Analysis and Interpretation*, pp.290-294, 7-9 April 2002