SPATIAL-SPECTRAL FEATURE ANALYSIS IN JPEG AND JPEG-2000

K.M. Au, N.F. Law and W.C. Siu

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Hong Kong.

ABSTRACT

Images are often compressed using algorithms such as JPEG and JPEG-2000. Retrieving these kinds of compressed images typically require processing them onto the uncompressed spatial form in which features are extracted for further analysis. This approach incurs many pre-processing operations, especially for large image archives. The objective of this paper is to study the common features in different compressed domains so that indexing can be done directly from their respective domains. By employing a subband filter model, filters in the block-based discrete cosine transform (BDCT) and wavelet transform (WT) can be directly compared. We have found that both filters share similar characteristics. In fact, the lowpass filters in the two transforms are the same for Haar wavelet. Therefore, due to the filter similarity, similar features can be extracted from the BDCT and WT subband outputs. We have also studied the effect of compression on their similarity. Despite the high compression, large similarity can still be found. Both our theoretical and experimental studies confirm that common features can be extracted directly from the compressed domains irrespective of the value of the compression ratio and the use of BDCT or WT.

1. INTRODUCTION

With the tremendous increase in the amount of visual data, an efficient image retrieval system is highly desirable. The retrieval system should extract meaningful information from the data so that images can be retrieved efficiently based on their contents. In the past few years, many retrieval systems for uncompressed spatial images have been proposed, e.g., QBIC, Virage [1-2]. Recently, image compression standards such as JPEG have been used to reduce the storage requirements. Retrieving these kinds of compressed images then require some pre-processing operations to convert them back to the uncompressed form for feature analysis. To avoid some of the decompression operations, feature extraction was proposed to be done directly from the transformed domains.

Most of the compression systems are based on either the BDCT or the WT. BDCT has been selected as part of the international standards such as JPEG [3]. Indexing techniques in BDCT domains involve extracting features by analyzing the DCT coefficients [4]. Recently, wavelets have emerged as a new tool in compression. In fact, the new compression standard, JPEG-2000 [5], had adopted wavelets in its compression algorithm. Various WT-based indexing techniques have been proposed in which significant wavelet magnitudes are used as features [6].

Due to the various compression standards, images would be available in heterogeneous nature, i.e., some might be stored using BDCT technique while some using WT-based technique. It is very much desirable if we can explore common features in these domains so that indexing can be done directly from their original domain. The objective of this paper is to study the common features between the BDCT and the WT domains. A common subband filter model will first be derived. Using this model, common features in BDCT and WT will be investigated.

2. TRANSFORM TECHNIQUE

2.1. JPEG compression scheme

JPEG compression scheme employs BDCT [3]. The whole image in spatial domain is divided into a number of 8×8 pixels blocks. Each block then undergoes a 2D discrete cosine transform (DCT) to convert to frequency domain. Mathematically, let $x(k,l)$ be the original image, where k and l are from 0 to $N-1$ and N is the size of the image. Then, the DCT coefficients $X_{m,n}(u,v)$ at (u,v) is defined as,

$$X_{m,n}(u,v) = \frac{1}{4}K(u)K(v)\sum_{i=0}^{7}\sum_{j=0}^{7}x(8m+i,8n+j)\cos\frac{(2i+1)u\pi}{16}\cos\frac{(2j+1)v\pi}{16} \quad (1)$$

where $K(u)$ equals to $1/\sqrt{2}$ for u=0 and 1 otherwise. Those DCT coefficients with same u in each block can be concatenated together so as to provide spatial-frequency information. For example, the DC coefficient (i.e., u=0) in each block can be concentrated together to form a blurred version of the original signal as shown in Figure 1a. We named this blurred image as the DC image.

2.2. JPEG-2000 compression scheme

JPEG-2000 compression scheme uses WT to map images from spatial domain into spatial-frequency domain. Since wavelets are well localized in both spatial and spectral domains; they can produce a multiple resolution view of an image. As can be seen in Figure 1c, the low-low band gives coarse information while the other subbands give details of the original image. We named this coarse image from the low-low band as the lowpass image (Figure 1b).

(a) (b) (c)

Figure 1: (a) DC image from BDCT (b) lowpass image from WT and (c) wavelet subbands in three decomposition levels.

3. SUBBAND FILTER MODEL

In both BDCT and WT domains, spatial and spectral information regarding the original image are preserved. The DC image from DCT and lowpass image from WT can be extracted from their respective compressed domains for further analysis due to their similarity [7]. It is shown that the DC and lowpass filter that are used to obtain the DC and lowpass image are very similar [8]. In the following, mathematical formulations are derived to describe a subband filter model for the filters in BDCT and WT cases. Since both transforms belong to the class of a separable 2D transform, our formulation is done in 1D for simplicity.

3.1. Filters in BDCT

The DCT coefficients with same u (Eqn. 1) in each block can be concatenated together so as to provide spatial-frequency information. In matrix form,

$$\begin{bmatrix} X_0(u) \\ X_1(u) \\ \vdots \\ X_{\frac{N}{8}-1}(u) \end{bmatrix}_{\frac{N}{8}x1} = \begin{bmatrix} A_{u,1x8} & \mathbf{0}_{1x8} & \cdots & \mathbf{0}_{1x8} \\ \mathbf{0}_{1x8} & A_{u,1x8} & \cdots & \mathbf{0}_{1x8} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{1x8} & \mathbf{0}_{1x8} & \cdots & A_{u,1x8} \end{bmatrix}_{\frac{N}{8}xN} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}_{Nx1} \quad (2)$$

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

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

Analyzing eqn.2 shows that the concatenated DCT coefficients can be obtained by employing a filtering operation followed by a sub-sampling operation, i.e.,

$$\tilde{X}_{u,DCT}(z) = \tilde{F}_{u,DCT}(z)\tilde{X}(z) \quad \downarrow 8 \quad (4)$$

where $\tilde{F}_{u,DCT}(z)$ is formed using Eqn..3 as

$$\tilde{F}_{u,DCT}(z) = \frac{K(u)}{2}\sum_{i=0}^{7}\cos\frac{(2i+1)u\pi}{16}z^i \quad (5)$$

The spatial-frequency characteristics of the concatenated DCT coefficients then depend on the filter $\tilde{F}_{u,DCT}(z)$.

3.2. Filters in WT

In each decomposition level, the signal is lowpass and bandpass filtered and then followed by a sub-sampling by two operation. For a three-level decomposition, the lowpass and bandpass signals can be written respectively as,

$$\tilde{X}_{0,WT}(z) = \tilde{F}_{o,WT}(z)\tilde{X}(z) \quad \downarrow 8$$
$$\tilde{X}_{i,WT}(z) = \tilde{F}_{i,WT}(z)\tilde{X}(z) \quad \downarrow 2^{4-i}, \ i=1,2,3 \quad (6)$$

where
$$\tilde{F}_{i,WT}(z) = \begin{cases} \tilde{H}(z)\tilde{H}(z^2)\tilde{H}(z^4) & i=0 \\ \tilde{G}(z) & i=3 \\ \tilde{G}\left(z^{2^{3-i}}\right)\prod_{j=1}^{3-i}\tilde{H}(z^j) & i=1,2 \end{cases} \quad (7)$$

$$\tilde{H}(z) = \sum_{n=0}^{N-1} h_n z^n \quad \tilde{G}(z) = \sum_{n=0}^{N-1} g_n z^n \quad (8)$$

h_i and g_i denotes the lowpass and bandpass filters in WT respectively.

3.3. Subband Structure for BDCT and WT

Eqn. 4 shows that the DCT output, $\tilde{X}_u(z)$, can be obtained by passing input $\tilde{X}(z)$ through a filter $\tilde{F}_{u,DCT}(z)$ followed by a down-sampling by 8 operation. Eqn. 6 shows that the WT output, $\tilde{X}_i(z)$, can be obtained by passing the input through a filter $\tilde{F}_{i,WT}(z)$ followed by a down-sampling operation. Therefore, for both BDCT and WT, the transformed outputs are always obtained by filtering followed by a down-sampling operation. This common subband model implies that spatial-spectral information is preserved by the two transforms. Our objective is to analyze and compare this information so as to characterize their similarity. Firstly, we will compare the filters, $\tilde{F}_{u,DCT}(z)$ and $\tilde{F}_{i,WT}(z)$ to find out their similarity. Then we will compare the transformed outputs, $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$.

4. FILTER ANALYSIS

The subband formulation in Eqn. 4 and Eqn. 6 provides a foundation for the theoretical analysis of the spatial-frequency information in the BDCT and WT. By comparing $\tilde{F}_{u,DCT}(z)$ and $\tilde{F}_{i,WT}(z)$, one can find out their spectral characteristics and then determine their spectral similarity. Two measures are used, namely the passband region and filter similarity.

4.1. Passband region

The magnitude spectra of $\tilde{F}_{u,DCT}(z)$ and $\tilde{F}_{i,WT}(z)$ are compared. It is found that $\tilde{F}_{u,DCT}(z)$ partitions the frequency spectrum uniformly over the frequency range [-π,π]. For u from 0 to 7, the passband region of $\tilde{F}_{u,DCT}(z)$ are (0-0.11π), (0.09π-0.26π), (0.18π-0.39π), (0.30π-0.51π), (0.42π-0.63π), (0.54π-0.76π), (0.68π-0.90π) and (0.85π-π) respectively. In contrast, the width of the passband region of $\tilde{F}_{i,WT}(z)$ increases with i. This implies that a particular wavelet filter might share similar characteristics with a number of DCT filters. To complicate the issue,

$\tilde{F}_{i,WT}(z)$ depends on the choice of the wavelet kernel as shown in Table 1. Haar, Daubechies 4 (DB4), Biorthogonal 93 (B93) and Biorthogoanl (B97) kernels give slightly different 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 1: Passband region of WT filters using different kernels

4.2. Similarity

To quantify the filter similarity, one can use the similarity measure which is defined in [3],

$$Similarity = \max \sum_{k=0}^{N-1} |A(k)||B(k-j)| \left[\left(\sum_{k=0}^{N-1} |A(k)|^2 \right) \left(\sum_{k=0}^{N-1} |B(k)|^2 \right) \right]^{-1/2} \quad (9)$$

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.

4.2.1. Lowpass Filters

For *i=0* and *u=0*, we found that S equals to 1 for the Haar kernel, i.e., an exact match. Theoretically, from Eqn..5 and Eqn..7,

$$\tilde{F}_{0,WT}(z) = \frac{1}{2\sqrt{2}} \sum_{i=0}^{7} z^i = \tilde{F}_{0,DCT}(z) \quad (10)$$

Therefore, the DC image from DCT and the lowpass image 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 0.9831, 0.9522 and 0.9659 respectively for DB4, B93 and B97 kernels. They are all close to 1; which means that filters are very similar. Also as shown in Table 1, the passband regions are similar. In summary, the DC image from BDCT and the lowpass image from WT retain similar spatial-spectral content.

4.2.2. Bandpass Filters

Using the subband filter model, we can see that $\tilde{F}_{1,WT}(z)$ can be directly compared to $\tilde{F}_{1,DCT}(z)$ since both involved a down-sampling by 8 operation. The passband region of $\tilde{F}_{1,DCT}(z)$ is (0.09π-0.26π) that is comparable to that of $\tilde{F}_{1,WT}(z)$ for all 4 kernels. The average similarity measure over the four kernels is 0.917.

Due to the different width of the passband regions, a few DCT filters should be combined to match the WT filters. For example, the resultant filter formed from $\tilde{F}_{2,DCT}(z)$ and $\tilde{F}_{4,DCT}(z)$ shows a good match to $\tilde{F}_{2,WT}(z)$ in terms of both filter similarity and passband region. In particular, the resultant passband region from the DCT is (0.18π-0.64π) which is comparable to that of WT as shown in Table 1. Also, the average similarity measure over the four kernels to this resultant DCT filter is 0.9485. Therefore, similar spatial-spectral contents are preserved.

Besides, $\tilde{F}_{3,WT}(z)$ 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 the sum of the 4 DCT filters is (0.46π-π) which is comparable to that of the WT filter (Table 1). The average similarity over the four kernels is 0.9873.

5. EXPERIMENTAL RESULTS

Section 4 shows that large similarity has been found between the WT and the BDCT filters. However, different images have different spectral contents. For example, the spectral content of natural images concentrates in the low frequency part. Thus difference in filters could be emphasized or de-emphasized when applying to images. In this section, a similarity study will be carried out on various images. In our image database, there are four different kinds of images: 29 natural scenes, 11 structured texture patterns, 10 random texture patterns, and 17 man-made (artificial) object images.

5.1. No Compression

We consider the following four kernels: Haar, DB4, B93 and B97. Similarity is calculated on the four image kinds to find out how similar the outputs, $\tilde{X}_{u,DCT}(z)$ and $\tilde{X}_{i,WT}(z)$, are. Table 2 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 2: Similarities in spatial domains between the BDCT and WT images for four different kernels.

Most of the output images show a large similarity between the BDCT and WT. The only exception is for B93 and B97 in $\tilde{X}_{1,DCT}(z)$ and $\tilde{X}_{1,WT}(z)$. It is because the passband regions are different mostly near the low-frequency end. This thus emphasizes the differences in the spatial-spectral information extracted. However, for all other outputs, the similarity is close to 1. Therefore, common features should be able to find from the BDCT and WT subband outputs for direct comparison.

5.2. Effect of compression

Besides the use of different transforms in JPEG and JPEG-2000, they also differ in their compression algorithm. On one hand, features that are extracted directly from their compressed domains could be different under different compression ratio.

On the other hand, features extracted from JPEG and JPEG-2000 can be different even under the same compression ratio. In this sub-section, JPEG and JPEG-2000 compression schemes will be applied to compress images to various degrees. Then the subband images are extracted from their respective compressed domains for similarity measure.

We considered six compression ratios, 6:1, 12:1, 18:1, 24:1, 32:1 and 44:1. The average similarity values of the four image kinds over the four wavelet kernels verse compression ratio is plotted in Figure 2. We can see that the similarity between $\tilde{X}_{0,DCT}(z)$ and $\tilde{X}_{0,WT}(z)$ remains large even under a high compression ratio. This means that very similar spectral information can be extracted from BDCT and WT at this lowpass band. Note that most of the image information is concentrated at this frequency band. $\tilde{X}_{0,DCT}(z)$ and $\tilde{X}_{0,WT}(z)$ are thus ideal for indexing because retrieval performance would not be greatly affected by the compression.

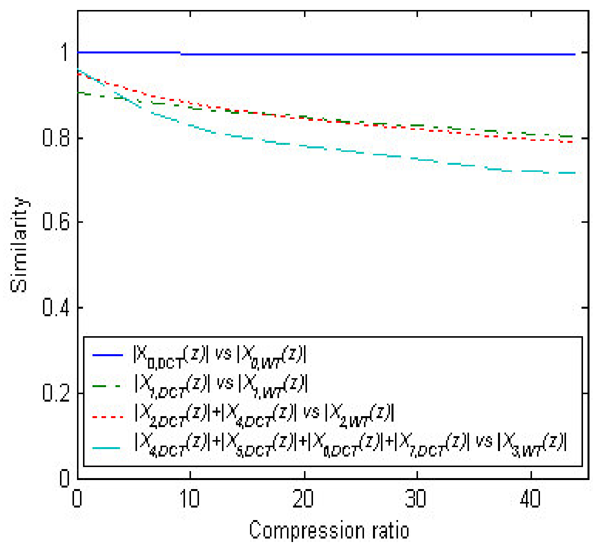

Figure 2: Similarities in spatial domains between the DCT and WT subband image at different compression ratio.

The similarity values of the other three subbands under different compression ratio are lower than that of the lowpass subband. The similarity values of the $\tilde{X}_{1,DCT}(z)$ and $\tilde{X}_{1,WT}(z)$ is obviously the smallest at no or low compression. This matches our findings in section 5.1 where the passband regions of these two filters do not show a good correspondence. As for the other two higher bandpass subbands, i.e., ($\tilde{X}_{2,DCT}(z)$ + $\tilde{X}_{4,DCT}(z)$, $\tilde{X}_{2,WT}(z)$) and ($\tilde{X}_{4,DCT}(z) + \tilde{X}_{5,DCT}(z) + \tilde{X}_{6,DCT}(z) + \tilde{X}_{7,DCT}(z)$, $\tilde{X}_{3,WT}(z)$), the similarity values are higher under low compression. However, the similarity values decrease rapidly with the increase in compression ratio. This is due to the difference in JPEG and JPEG-2000 algorithm. It is well known that JPEG-2000 retains more high frequency information than JPEG under the same compression ratio. As a result, the difference in the high frequency subband between the BDCT and WT will be amplified in large compression cases.

Our findings show that the lowpass subband shows the highest similarity among all the subbands. In image indexing, more emphasis can be put on the lowpass subband. The bandpass subbands can be used as a refinement during image indexing. Thus, the difference between higher subbands would not affect the indexing results significantly.

6. CONCLUSION

Compression schemes such as JPEG and JPEG-2000 employ block-based discrete cosine transform (BDCT) and wavelet transform (WT) respectively. In this paper, we first formulated a common subband model for BDCT and WT. Using this model, subband filters for both BDCT and WT are found to have similar passband region and large filter similarity. In fact, high similarity values are obtained between the pairs: lowpass: ($\tilde{F}_{0,WT}(z), \tilde{F}_{0,DCT}(z)$), bandpass 1: ($\tilde{F}_{1,WT}(z)$, $\tilde{F}_{1,DCT}(z)$), bandpass 2: ($\tilde{F}_{2,DCT}(z) + \tilde{F}_{4,DCT}(z)$, $\tilde{F}_{2,WT}(z)$) and bandpass 3: ($\tilde{F}_{4,DCT}(z) + \tilde{F}_{5,DCT}(z) + \tilde{F}_{6,DCT}(z) + \tilde{F}_{7,DCT}(z)$, $\tilde{F}_{3,WT}(z)$). Since image characteristics could emphasize or de-emphasize the effect of filter difference, we have also studied the similarity between the subband outputs. Our experimental results show that a high similarity still exists between the outputs from the BDCT and WT. Our result on the effect of compression shows that despite the high compression ratios, common features can be extracted from these two compressed domains. Therefore, both the theoretical and experiment results confirm that indexing can be done directly from their respective transformed domains no matter the image is compressed using JPEG or JPEG-2000.

7. REFERENCES

[1] 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", *IEEE Transactions on Computer*, Vol. 28, pp. 23-32, 1995.

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

[3] Gregory K. Wallace, "The JPEG Still Picture Compression Standard", *Comm. ACM*, Vol. 34, pp. 30-44, 1991.

[4] Shih-Fu Change, "Compressed Domain Techniques for Image/Video Indexing and Manipulation", *IEEE International Conference on Image Processing*, ICIP'95, p. 314-317, 1995.

[5] Usevitch, B.E., "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.

[6] K.C. Liang and C.C. J. Kuo, "Retrieval and Progressive Transmission of Wavelet Compressed Images", *Proc. of ISCAS*, pp. 1464-1467, 1997.

[7] R.F. Chang, W.J. Kuo and H.C. Tsai, "Image Retrieval on Uncompressed and Compressed Domains", *International conference on Image Processing*, Vol. 2, 546-549, 2000.

[8] K.M. Au, N.F. Law and W.C Siu, "Unified Feature Analysis in Different Compressed Domains", *International Conference on Information, Communications and Signal Processing and Pacific-Rim Conference on Multimedia*, Dec. 2003.