This is the Pre-Published Version.

Wavelet Packet Based Approach for Image Retrieval in Compressed Domains

K.O. Cheng, N.F. Law and W.C. Siu
Centre for Signal Processing, Dept of Electronic and Information Engineering,
the Hong Kong Polytechnic University, Hong Kong.
E-mail: {encko, ennflaw, enwcsiu}@polyu.edu.hk

***Abstract*— Images are often compressed to reduce storage. In order to avoid full decoding in image retrieval, features extracted directly from transformed data have been investigated. The popularity of JPEG and JPEG2000 motivates us to investigate a fast conversion scheme from block-based discrete cosine transform (BDCT) to wavelet domain so that similar features can be extracted from the two different domains. In the proposed scheme, BDCT is first approximated to wavelet packet (WP) transform and then the synthesis WP filter bank is applied to construct wavelet subbands. Studies found that the proposed conversion scheme has a low computational complexity as compared to the baseline approach. Two algorithms are then developed for extracting features in the wavelet subbands for retrieval. Our experimental results show that the proposed algorithms provide a good retrieval performance while the complexity is kept to be low.**

I. INTRODUCTION

Image retrieval is a challenging but important problem in image processing. Much effort has been spent on content-based image retrieval (CBIR) which relies on describing visual contents using color, texture and shape. Unlike human annotation, CBIR is able to produce an objective and automatic image description. Features can be extracted for image representation using methods such as wavelet transform (WT), Gabor filters and multiresolution simultaneous autoregressive models. These extraction methods work on spatial domain which is unavailable without full image decoding. The computational burden of image decoding is crucial for retrieval in media like Internet. In order to avoid unnecessary processing, methods proposed in [1, 2] extract features from image file headers and transformed image data. Despite that, these methods have been designed for one specific transform domain. In real situation, images can be compressed using different formats. As a result, cross format extraction methods are much desired.

Recent works on compressed domain retrieval mainly focus on JPEG [3] and JPEG2000 [4] formats. JPEG is currently one of the most popular compression standards. JPEG2000 is a successor of JPEG which replaces the block-based discrete cosine transform (BDCT) by WT. Retrieval on JPEG and JPEG2000 images is usually achieved by extracting wavelet-based features for both image types. For JPEG2000 images, wavelet-based features can be obtained directly without full decoding. However, for JPEG images, fast conversion methods from BDCT to wavelet domains are required. Existing conversion methods include multiresolution reordering [1, 5], a filter bank method which fuses inverse BDCT and WT together [6] and a characteristic image based approach [7]. The performance of the multiresolution reordering is limited by its one-to-one assignment strategy. The fused filter bank is efficient for coarse scales but its complexity increases at fine scales. The characteristic images based approach cannot extract fine scale features. In this paper, we propose an efficient conversion method based on wavelet packet (WP) transform. Our proposed method explores the similarity between BDCT and WP decomposition as well as the relationship between WP and WT. In Section II, we first review WP transform. Section III then presents the proposed fast approach for BDCT to WT conversion. Our two proposed algorithms are described in Section IV while the experimental results are discussed in Section V. Finally, Section VI concludes this paper.

II. REVIEW ON WAVELET PACKETS

The wavelet packet (WP) decomposition [8] for a 2D image $x[n_1,n_2]$ can be realized by recursively applying a separable filter bank as,

$$\begin{aligned}
&w_{j+1}^{2p_1,2p_2}[n_1,n_2] \\
&\quad = \sum_{m_1}\sum_{m_2} w_j^{p_1,p_2}[m_1,m_2]\tilde{h}_0[m_1-2n_1]\tilde{h}_0[m_2-2n_2] \\
&w_{j+1}^{2p_1+1,2p_2}[n_1,n_2] \\
&\quad = \sum_{m_1}\sum_{m_2} w_j^{p_1,p_2}[m_1,m_2]\tilde{h}_1[m_1-2n_1]\tilde{h}_0[m_2-2n_2] \\
&w_{j+1}^{2p_1,2p_2+1}[n_1,n_2] \\
&\quad = \sum_{m_1}\sum_{m_2} w_j^{p_1,p_2}[m_1,m_2]\tilde{h}_0[m_1-2n_1]\tilde{h}_1[m_2-2n_2] \\
&w_{j+1}^{2p_1+1,2p_2+1}[n_1,n_2] \\
&\quad = \sum_{m_1}\sum_{m_2} w_j^{p_1,p_2}[m_1,m_2]\tilde{h}_1[m_1-2n_1]\tilde{h}_1[m_2-2n_2] \quad (1)
\end{aligned}$$

where $w_j^{p_1,p_2}[n_1,n_2]$ is the (p_1,p_2)th wavelet packet subband at level j, $\tilde{h}_0[-n]$ and $\tilde{h}_1[-n]$ are 1D lowpass and highpass filters respectively. Note that the decomposition is initialized as $w_0^{0,0}[n_1,n_2]=x[n_1,n_2]$. The decomposition is equivalent to filtering followed by downsampling by a factor of 2 in each dimension. Perfect reconstruction requires existence of filters $h_0[n]$ and $h_1[n]$ such that

$$w_j^{p_1,p_2}[n_1,n_2] = \sum_{\delta_1\in\{0,1\}}\sum_{\delta_2\in\{0,1\}}\sum_{m_1}\sum_{m_2} w_{j+1}^{2p_1+\delta_1,2p_2+\delta_2}[m_1,m_2]h_{\delta_1}[n_1-2m_1]h_{\delta_2}[n_2-2m_2] \quad (2)$$

Equation (2) also states that reconstruction of a subband at level j is achieved by first upsampling each of the corresponding four subbands at level j+1. Then the results are filtered and added together. In principle, subbands can be decided arbitrarily for further decomposition in next level. The full WP decomposition refers to the one with all subbands decomposed in each level. If only $w_j^{0,0}[n_1,n_2]$ is involved, WPs reduce to WTs. Thus WPs extend wavelets for a more general signal representation and analysis.

III. PROPOSED FAST CONVERSION SCHEME

The full wavelet packets (WPs) at level l have similar space-frequency localization as the basis functions of BDCT with a block size of $2^l \times 2^l$. In particular, both transforms divide the frequency spectrum in a nearly uniform manner and have the same sampling rate of 2^l in each dimension. Because of their similarity in frequency partition, the BDCT coefficients at the same frequency can be grouped to form subbands [9]. These DCT subbands are then used to approximate the WP subbands based on the idea of frequency ordering [8]. Without loss of generality, 1D signal is considered. Let $p = p_l p_{l-1} \ldots p_1$ and $k = k_l k_{l-1} \ldots k_1$ be the binary representations of the index of a WP subband and the frequency index of a DCT subband respectively. The *k*th DCT subband is matched to the *p*th WP subband using the following formula,

$$p_i = \begin{cases} (k_i + k_{i+1}) \bmod 2, & 1 \le i < l \\ k_i, & i = l \end{cases} \quad (3)$$

The subband matching in 2D is achieved by applying the 1D mapping to each dimension. Since we can obtain the full WP decomposition by further decomposing the bandpass subbands in WT, the wavelet subbands are reconstructed from the DCT subbands by performing inverse operations. Fig. 1 shows the proposed conversion method for BDCT with a block size of 8 to WT.

As illustrated in Fig. 1, the proposed algorithm can construct a wavelet coefficient from the BDCT coefficients which are located around it in spatial and frequency domains. First the DCT subband $X_k[n]$ is considered to be equal to the WP subband $\hat{w}_p[n]$ according to the frequency order in equation (3). By applying a synthesis filter bank of WP decomposition, the WT subband can be obtained. For example, the 4th-7th BDCT components are combined to estimate the bandpass subband at the first level by using a WP synthesis filter bank with two levels. The bandpass subband at the second level is calculated from the DCT subbands 2 and 3 using the synthesis filter bank of one level. The bandpass and lowpass subbands at the third level are the same as the first and the zeroth DCT subbands respectively. In contrast to the multiresolution reordering approach that obtains wavelet subbands by rearranging BDCT coefficients into the octave band structure, our method takes into account the fact that a wavelet basis could overlap with a set of adjacent BDCT bases. Thus, the proposed algorithm can provide better approximation to wavelet subbands than multiresolution ordering.

It has been assumed that the filter used in the WP synthesis filter bank is the same as that used in WT. Nonetheless, it is possible to select another filter which may improve approximation accuracy or reduce complexity. In the subsequent discussion, filters in the WP synthesis filter bank are called primary filters if they are same as that in WT. Otherwise, they are called secondary filters. In the approximation, the primary filter incurs error because of its difference with the BDCT kernels. For the secondary filter, there is an additional error caused by its difference with the wavelet kernel. As Haar filter is close to the BDCT kernels and has low complexity, we select it as the secondary filter.

In order to study the approximation accuracy of these two synthesis filters in BDCT to WT conversion, root mean square error (RMSE) was calculated between the converted coefficients and the actual coefficients obtained by applying WT. Table I summarizes the results. The baseline approach consists of an inverse BDCT and a forward WT. Thus it incurs no errors in the conversion. The proposed WP approach achieves a smaller RMSE than the multiresolution reordering, especially for images with lots of textures like Barbara. Besides, the secondary filter gives a better approximation than the primary filter. This is because the BDCT and the WP with Haar kernel have the same delay in the transformed coefficients, but that with the CDF9/7 have different delay due to its odd-symmetry nature. Hence, the reconstruction from using the Haar kernel is much better than that using the CDF9/7 kernel.

TABLE I

RMSE OF MULTIRESOLUTION REORDERING (REORDER) AND THE PROPOSED WP APPROACH WITH PRIMARY (WPP) AND SECONDARY FILTER (WPS).

Image	Baseline	Reorder	WPP	WPS
Lena	0	26.96	26.57	26.18
Barbara	0	34.86	32.71	30.75

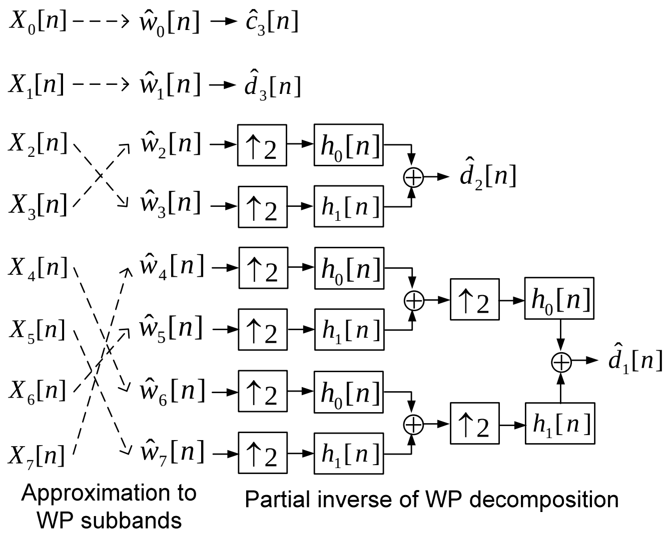


Fig. 1 The proposed fast conversion scheme for 1D signals. $X_k[n]$ is the kth DCT subband and $\hat{w}_p[n]$ is the pth WP subband.

IV. EFFICIENT ALGORITHM FOR JOINT RETRIEVAL

In retrieval of JPEG and JPEG2000 compressed images, wavelet-based features are often extracted as common features because of their promising retrieval performance. These features can be directly computed from JPEG2000 images after entropy decoding. For JPEG images, however, conversion from BDCT to wavelet domain is required. We thus apply the proposed fast scheme to JPEG images to obtain wavelet-based features. Two algorithms are considered. Algorithm I refers to the use of the WP scheme described in Section 3 to all scales while Algorithm II applies the WP to the lowpass subband and bandpass subbands at the 1st and the 2nd levels. For the bandpass subband at the 3rd level, the fused filter bank that integrated the process of inverse BDCT and WT is applied to have a higher approximation accuracy [6].

A. Computational Complexity

To evaluate the complexity of the two proposed algorithms, the number of arithmetic operations for an image of size $N \times N$ has been estimated and provided in Table II. Our approach is compared with two other methods, namely the baseline approach and the multiresolution reordering approach. As CDF 9/7 is one of the default filters in JPEG2000 standard, it is the primary filter. Lifting implementation [11] is considered for WT and WP filter banks. The BDCT in the baseline approach uses a fast scaled DCT on 8×8 points [12].

The baseline approach consists of an inverse BDCT and a forward WT. Its complexity is the highest. The multiresolution reordering approach involves rearranging DCT coefficients only, thus there is no arithmetic operation involved. The Algorithm I with the secondary filter (Haar) uses less arithmetic operations than that with the primary filter (CDF 9/7). It needs no multiplications and only about 19.1% additions of the baseline approach. Although there is an increase in shift operation, the influence on the overall complexity is negligible because additions and multiplications are more complicated than binary shifts. Compared with the multiresolution reordering, the Algorithm I with Haar filter requires slightly more operations to provide a better approximation accuracy. For Algorithm II, the WP is applied to the lowpass and the bandpass subbands at the 1st and the 2nd levels. For the bandpass subband at the 3rd level, extra computation is need for the conversion from the BDCT to WT. Hence, Algorithm II has a slightly higher complexity than Algorithm I. As in Table II, our two proposed algorithms are less complex than the baseline approach. The computational complexity is greatly reduced.

TABLE II
NUMBER OF ARITHMETIC OPERATIONS FOR AN IMAGE OF SIZE $N\times N$

Operation	Addition	Multiplication	Shift
Baseline approach	$17.7N^2$	$8.72N^2$	$0.0938N^2$
Multiresolution reordering	0	0	0
Algorithm I:			
primary: CDF 9/7	$13.5N^2$	$10.1N^2$	0
secondary: Haar	$3.38N^2$	0	$1.69N^2$
Algorithm II	$4.91\ N^2$	$1.80N^2$	$1.69N^2$

B. Features Extraction

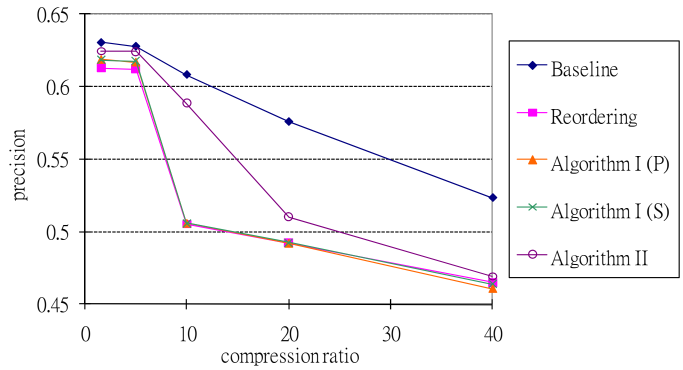
In feature extraction, sum of squared coefficients is calculated as energy from all subbands in WP. In addition, seven normalized central moments are computed from significance maps of all subbands excluding the lowpass subband because the lowpass subband lacks edge information. The overall similarity between two images is measured as the weighted sum of the distances in the two feature spaces of the energy and the moments as described in WaveGuide [10]. Given a query image, images in the database of the retrieval system are ranked according to their similarity with the query image.

V. RETRIEVAL EXPERIMENTS

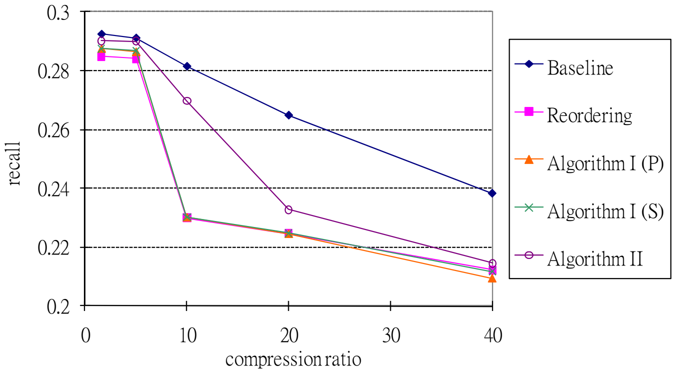
Experiments were conducted to evaluate the proposed fast algorithm in retrieval of images compressed in JPEG and JPEG2000 formats. There are 3600 images divided into nine classes [5] with each image contains two copies: one in JPEG and the other in JPEG2000 formats. The CDF 9/7 filter was selected for JPEG2000 compression. In the experiments, each image was used as a query image once. All images excluding the copies of the query image formed the database set. Thus, there are 358 images relevant to the query image in the database. Precision and recall [13] are calculated at different numbers of retrieved images varied from 1 to 358. The experiments were repeated at five compression ratios of 1.6, 5, 10, 20 and 40.

As shown in Fig. 2, the performance of the Algorithm I with the primary filter (CDF 9/7) and the secondary filter (Haar) is nearly the same. Despite that the secondary filter, i.e., Haar filter, has two sources of error as mentioned in Section III, it gives the same delay as the DCT kernels. As a result, Haar filter matches better to the BDCT kernel than the CDF 9/7 which compensates for its dissimilarity to the wavelet kernels. Since Haar filter can also reduce the computational complexity substantially as shown in Table II, it is recommended for practical uses.

Comparing with the multiresolution reordering, the Algorithm I has a higher precision and recall at low compression ratios. At high compression ratios, their performance becomes similar. This is because the approximation to wavelet subbands is mainly improved at fine scales. In fact, the approximated subbands at the third level are the same as those constructed by the multiresolution reordering. When compression ratio increases, more and more high frequency coefficients of JPEG images are quantized to zero. As a result, only marginal improvement is obtained. Algorithm II uses a filter bank to convert the BDCT coefficients to wavelet coefficients at the 3rd level and thus has a better approximation accuracy than Algorithm I. It can be seen from Fig. 2 that the Algorithm II can improve the performance at all compression ratios. More importantly, both precision and recall of the Algorithm II are better than that of the multiresolution reordering and the Algorithm I.



(a)



(b)

Fig. 2 (a) Precision and (b) recall at different compression ratios.

VI. CONCLUSIONS

In this paper, we propose a new scheme which uses wavelet packet (WP) filter banks to construct wavelet subbands from block-based discrete cosine transform (BDCT). Our study finds that the computational complexity of the proposed conversion scheme is significantly smaller than the baseline approach which consists of an inverse BDCT and a forward WT. Two algorithms are then proposed to extract wavelet-based features from JPEG images so that images compressed in JPEG and JPEG2000 formats can be processed and retrieved in the same system efficiently. Retrieval experiments were conducted to evaluate their performances. It is found that both algorithms have a better recall and precision than the multiresolution reordering. Algorithm I has a lower computational complexity than Algorithm II, but Algorithm II provides a better approximation accuracy and thus achieves a higher retrieval performance than Algorithm I.

ACKNOWLEDGMENT

This work is supported by the Center for Signal Processing, Department of Electronic and Information Engineering, the Hong Kong Polytechnic University.

REFERENCES

[1] Y.-L. Huang and R.-F. Chang, "Texture features for DCT-coded image retrieval and classification", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3013-3016, Mar. 1999.

[2] M.K. Mandal and C. Liu, "Efficient image indexing techniques in the JPEG2000 domain", *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 182 – 190, 2004.

[3] W.B. Pennebaker and J.L. Mitchell, *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold, USA, 1993.

[4] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, USA, 2002.

[5] K.M. Au, N.F. Law and W.C. Siu, "Unified feature analysis in JPEG and JPEG 2000-compressed domains", *Pattern Recognition*, vol. 40, no. 7, pp. 2049-2062, 2007.

[6] K.O. Cheng, N.F. Law and W.C. Siu, "A fast approach for identifying similar features in retrieval of JPEG and JPEG2000 images", *Proceedings of Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, pp. 258-261, Oct. 2009.

[7] J. Liu and H. Gu, "Image retrieval in various domains", *Computers & Graphics*, vol. 5, pp. 807-81227, 2003.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, UK, 1999.

[9] Y.S. Hsu, S. Prum, J.H. Kagel and H.C. Andrews, "Pattern recognition experiments in the Mandala/cosine Domain", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 5, pp. 512-520, 1983.

[10] K.C. Liang and C.-C.J. Kuo, "WaveGuide: a joint wavelet-based image representation and description system", *IEEE Transactions on Image Processing*, vol. 8, no. 11, pp. 1619 – 1629, 1999.

[11] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets", *Applied and Computational Harmonic Analysis*, vol. 3, pp. 186-200, 1996.

[12] E. Feig and S. Winograd, "Fast algorithms for the discrete cosine transform", *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2174-2193, 1992.

[13] D. Feng, W.C. Siu and H.J. Zhang, *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, Springer-Verlag Berlin Heidelberg, 2003.