Adaptive Wavelet Transforms with Spatially Varying Filters for Scalable Image Coding

Wenjun Zeng, Jin Li and Shawmin Lei

Sharp Laboratories of America, Inc.
5750 NW Pacific Rim Blvd.
Camas, WA 98607

Abstract

An adaptive wavelet transform algorithm for scalable image coding is proposed in this paper. A quadtree segmentation scheme based on an entropy criterion is proposed to segment the image into several regions to allow for adaptive filtering using different types of prototype wavelet filters. A joint bit allocation and coding method is presented to efficiently code the segmented image in an embedding fashion. Experimental results show that the proposed scheme generally provides higher peak signal-to-noise ratios (PSNRs) at higher bit rates, and significantly better visual quality at lower bit rates, compared to fixed-filter wavelet transform based coding schemes.

1 Introduction

Discrete wavelet transform is a popular transform which has the nice features of space-frequency localizations and multiresolution representation. The characteristics of images differ from image to image, and sometimes even from one region of the image to another one. Therefore an adaptive transform fine-tuned to image characteristics is more effective than a fixed transform. Arbitrary subband decomposition trees have been introduced in [1] as wavelet packets (WP's) to deal with classes of signals with diverse or unknown time-frequency characteristics. To further characterize the space-varying characteristics, the spatially adaptive WP's were introduced in [2, 3] by performing a spatial segmentation and adapting the WP frequency decompostion to each spatial segment.

However, just as a fixed nonadaptive WP's may not characterize well the local space-varying characteristics of an image, a spatially adaptive WP with a *single* prototype wavelet filter is often not good enough to fully characterize the local space-varying characteristics of an image. Wavelet transform and WP's provide an elegant framework in which both "anomalies" and "trends" can be analyzed [4]. However, using a single wavelet prototype filter, the optimality of the trade-offs between the quality of "trends" and "anomalies" usually can not be achieved. For example, to achieve a better compression for a fixed distortion, a filter with long length tends to be used. However, at very low bit rates, the coder tends to allocate too many bits to the "trends", and have few bits left over to represent "anomalies". As a result, severe ringing artifact is usually observed around sharp edges.

In this work, we address the issue of spatially varying wavelet transform for image compression. We focus on joint spatial segmentation and selection of different prototype wavelet filters for different segments. Note that this is different from the spatially adaptive WP's in [2, 3, 5] where the prototype wavelet filter is fixed but the WP decomposition varies from region to region. In general, the adaptive WP's help increase the coding efficiency of the texture area. However, with a single preselected prototype wavelet filter, it usually does not help to improve the quality of the "anomalies" – the sharp edges. Therefore the ringing artifact will not be reduced. Using filters tailored to the local characteristics of image regions, the subjective quality and often the objective quality of both the "trends" and the "anomalies" can be improved. In this paper, a quadtree segmentation scheme based on an entropy criterion to segment an image into different regions for different wavelet filterings is proposed. A joint embedded coding method for images with multiple (potentially arbitrarily shaped) segments is developed to facilitate bit allocation among different segments and to allow scalability functionality. Experimental results show that the proposed scheme generally provides higher PSNRs at high bit rates, and noticeably better visual quality at lower bit rates, compared to non-adaptive wavelet transform based coding schemes.

This paper is organized as follows. Section 2 proposes a quadtree-based segmentation scheme for joint spatial segmentation and selection of different wavelet filters for different segments. A joint embedded bit allocation and coding method is described in Section 3 to code the segmented images. Experimental results are presented in Section 4. Section 5 draws the conclusion.

2 Joint Spatial Segmentation and Adaptive Wavelet Transforms

Our objective is to segment the image into different regions, and then perform wavelet transform using different types of prototype filter for different regions. Given a small set of candidate wavelet filters, our approach is to jointly optimize the spatial segmentation and the matching of prototype filters to different segments. After adaptive transformation, a joint embedded bit allocation and coding method is proposed to actually code the transform coefficients. The overall diagram of the proposed coding system is shown in Fig. 1.

In general, for compression applications, the spatial segmentation is rate-sensitive. Therefore, the characteristics of the candidate prototype wavelet filters, and the rate-distortion (R-D) performance of the coder, play important roles in the development of the optimal spatial segmentation. Several works have been developed to optimize the R-D performance of image coders, using the Lagrangian optimization method [2, 3, 5, 6]. The computational cost, however, is usually very high, thus hindering the practical implementation. Li et. al. [7] proposed a fast WP decomposition scheme which efficiently determines the best WP transform in the R-D sense. With some empirical R-D model, it was shown [7] that the constant R-D slope criterion for optimal bit allocation is equivalent to the constant distortion criterion. In fact, as is well known, this constant distortion criterion is always true for high rate coding.

For simplicity, we use the quadtree structure for spatial segmentation. The tree is first grown to a predetermined depth. Then each leaf segment is wavelet decomposed to a preset depth using all the prototype filters from the candidate set, and uniformly quantized with a *fixed* quantizer step size which depends on the target bit rate or average distortion of the whole image. Note that this fixed quantizer step size strategy complies with the constant distortion criterion mentioned above for optimal bit allocation among different segments. After that, the first order entropy of the quantized coefficients is calculated for each potential wavelet prototype filter. The filter type which results in the minimal entropy is picked and assigned to this leaf segment. After all leaf segments have been assigned a filter type, the tree is merged from bottom up in a recursive way such that if all the children of a parent node are assigned the same type of filter, the children will be merged and the same type of filter will be assigned to the parent node. Note that, typically, segments assigned the same type of filter tend to cluster together.

It should be noted that we do not perform the selection of the best type of filter based on wavelet transform of the intermediate node. This is because if the children of the parent node differ from each other in the selection of the filter type, then the entropy of the parent node with any single type of filter is expected to be higher than the average entropy of the children with their best associated filters. Note that the scenario here is different from those in [3, 5] in that the depth of the wavelet decomposition in our work is fixed regardless of the depth of the node in the spatial quadtree. In [3, 5], the spatial quadtree and the WP tree are combined into a single tree, and the restriction that the combined tree should be *balanced* is imposed. With this restriction, a parent node in the spatial quadtree has one more layer of WP decomposition than their children nodes. This sacrifice of an iteration of WP decomposition for a spatial segmentation will most probably result in loss of coding efficiency. In our work, the spatial quadtree *structure* does not play a significant role in the R-D optimization process. It only facilitates the coding of the segmentation map by merging leaf segments with the same type of filter into a larger segment. This merging helps to reduce the number of boundaries in the segmentation map, thus reducing the potential boundary artifact which may occur at low bit rates.

In principal, the segmentation result will depend on the final target bit rate. Using different segmentation maps for different intermediate rates makes it extremely difficult, if not impossible, to generate a (preferably continuously) scalable bit stream. From a perceptual point of view, the map generated at higher target bit rate is preferred because it more accurately reflects the characteristics of each region. Note that at lower bit rates, the decoded image may have been *visually* distorted, thus the entropy may not reflect the visual quality correctly. Using a segmentation map generated at low target bit rate may lead to a reconstructed image with severe visual artifacts at low bit rates, though the quantitative measure such as PSNR may be higher. With the considerations of perceptual quality and desirable scalability functionality, we will use a *single* segmentation map generated at a *higher* bit rate for adaptive wavelet transform.

3 Joint Embedded Bit Allocation and Coding

Once the image has been segmented into different regions, each region is wavelet transformed using the associated best prototype filter. Note that in our system, the wavelet decomposition level for each segment is set to the same value. Arbitrary shape wavelet transform (ASWT) [8] is used, considering the irregularity (non-rectangular shape) of each region. Then the resulting coefficients from different regions are coded *jointly* using the R-D optimized embedded coding scheme (RDE) [9]. The RDE uses context adaptive arithmetic code and R-D optimized embedding where the available coding bits are first assigned to coefficients with the steepest rate-distortion slope. To synchronize the decoder with the encoder, the actual optimization is performed with the *predicted* R-D slope which can be calculated on the decode side. For more details about the RDE, the reader is referred to [9].

After the transformation, a rectangular *compound* wavelet coefficient image which has exactly the same size as the original image is generated by gathering the wavelet coefficients of different regions together. Essentially, the compound wavelet image appears to have exactly the same structure (scale, subbands, etc.) as a wavelet image generated using a single rectangular wavelet transform. This is made possible by the way the coefficients are organized in the arbitrary shape wavelet transform. In this work, we solve the bit allocation problem simply yet effectively by jointly coding the compound wavelet coefficient image in a way as if it were generated by performing a single wavelet transform on the *entire* original image. The RDE is used in our work, although in principle, any traditional coding method for wavelet transform coding can be used. This joint embedded coding approach has the nice feature that data coming from different regions of an image are treated as if they were coming from the same source, so bit allocation among different regions is realized implicitly. With the joint embedded bit allocation and coding, the resulting bit stream is generated in the order of importance, and the bits from different segments are multiplexed together in that order. In fact, with some small details being taken care of at the encoder, the system can also output separate bit stream for each segment, hence allowing object-based bit stream scalability.

4 Experimental Results and Discussions

In our experiments, the candidate set of prototype wavelet filters consists of three filters: the haar filter, the 9-3 tap, and the 9-7 tap Daubechies biorthogonal filters [10]. Although the 9-7 and 9-3 tap filters are biorthogonal, they are close to orthogonal so that the error introduced in the wavelet domain is close to the error in the space domain. The haar filter is the only symmetric, orthogonal filter, and is a good choice for text, graphics regions. The 9-7 tap filter is the most popular filter used for coding of natural images in the literature, and has generally produced very good R-D performance. The 9-3 tap filter, with shorter length than 9-7 tap filter, is expected to have less ringing artifact around sharp edges at low bit rates. A five level wavelet decomposition with symmetric boundary extension is used in our experiments.

For all the black and white test images, the segmentation map is generated at a target bit rate of 4 bits per pixel (bpp). This map is then used for the adaptive wavelet transform, and accounts for all presented coding results at other bit rates. The depth of the initial spatial quadtree is adapted to the size of the test image in a way such that the shorter size of the leaf block is between 32 and 64. The overhead for coding the segmentation map is typically negligible. For example, for the 512×768 compound image "cmpnd1" which consists of both text and natural image, the overhead is 14 bytes, corresponding to 0.1 % overhead at 0.25 bpp. To reduce the potential segment boundary effect at lower bit rates, a slightly modified version of the simple deblocking algorithm presented in [11] is applied to the region boundaries. Fig. 2 shows an example of the segmentation maps.

The R-D performance of the proposed scheme is summarized in Table 1. Note that the same RDE coding scheme is used for all different schemes. It is seen that, among non-adaptive schemes, none of the filters will *always* give the best performance. In fact, similar behaviors have been previously observed in [12]. For relatively higher bit rates, PSNR improvements (up to 2 dB) of the adaptive scheme over the best non-adaptive transform are observed for "cmpnd1", and "target" which consists of some graphics and charts. At lower bit rates, PSNR is not necessarily improved (with the exception of "cmpnd1"). This is reasonable, because the segmentation map is derived at a high bit rate and used for all rates. For "container", the PSNR of the proposed scheme is a little bit lower than the best non-adaptive transform even at the bit rate of 2 bpp. This is probably because the segmentation is based on the first order entropy, rather than the actual coding bit rate by RDE. Nevertheless, for all three test images, the visual quality is significantly improved by the proposed scheme, especially at medium and lower bit rate. Fig. 3 shows the decoded images of "container" using the adaptive scheme and 9-7 tap filter. It is seen that the adaptive transform scheme significantly reduces the ringing artifacts around sharp edges such as the pole and the edge near the right boundary. For "cmpnd1", the adaptive scheme eliminates almost all the ringing artifacts in the text region which otherwise will be presented in the decoded images using fixed transforms such as 9-7 tap or 9-3 tap filters; it also avoids the blocky effects presented in the natural image region if a fixed transform with haar filter is used.

The proposed adaptive scheme has essentially the same complexity at the decoder as a non-adaptive scheme. At the encoder, the extra complexity mainly lies in the wavelet transformations (3 times in our implementation) computed for each filter in the candidate set in the segmentation stage. The complexity incurred for segmentation, however, is much less than the iterative procedure adopted in [2, 3, 5, 6]. The proposed scheme is particularly useful for images exhibiting space-varying characteristics such as compound images and images having both smooth regions and very sharp edges. The selection of the set of candidate filters in our implementation is somewhat heuristic. The issue of how to select the best set of candidate filters remains open and warrants further investigation. One potential approach which is currently under investigation is to try to characterize (possibly through training) image regions using some models. Then, given a set of parameters characterizing the image source, a set of wavelet filters which result in best coding gain can be identified. In this way, by looking at the model which best characterizes the underlying image region, it may be possible to determine which wavelet filter is a good candidate for that particular region. This approach thus may potentially avoid the need to do transformation using all candidate filters in order to find the best filter for a specific region.

5 Conclusions

An adaptive wavelet transform algorithm with spatially varying filters for scalable image coding is proposed in this paper. A quadtree segmentation scheme based on an entropy criterion is proposed to segment the image into several regions where different types of prototype wavelet filters will be used for different regions. A joint embedded bit allocation and coding method is then used to code the images in a scalable fashion. It is shown that with the spatially adaptive algorithm, PSNR performance has been improved (by up to about 2 dB) over the best nonadaptive scheme, and the resulting reconstructed images generally have significantly better visual quality at lower bit rates. This scheme is particularly useful for images exhibiting space-varying characteristics such as compound images and images having both smooth regions and very sharp edges.

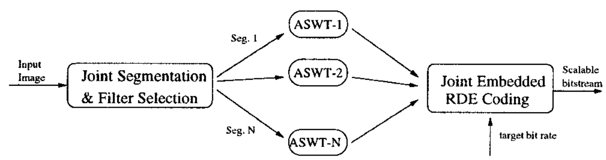

Figure 1: Spatially adaptive wavelet transform coding architecture

References

[1] R. Coifman and V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38:713–718, March 1992.

[2] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli. Tilings of time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms. *IEEE Trans. Signal Proc.*, 41(12):3341–3360, Dec. 1993.

[3] K. Ramchandran, Z. Xiong, K. Asai, and M. Vetterli. Adaptive transforms for image coding using spatially varying wavelet packets . *IEEE Tran. Image Proc.*, 5(7):1197–1204, July 1996.

[4] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Tran. Signal Proc.*, 41(12):3445–3462, Dec. 1993.

[5] J. Smith and S. Chang. Frequency and spatially adaptive wavelet packets. In *Proc. Inter. Conf. Acous., Speech and Sig. Proc.*, May 1995.

[6] E. Reusens et. al. Dynamic approach to visual data compression . *IEEE Tran. Cir. and Sys. for Video Technology*, 7(1):197–211, Feb. 1997.

[7] J. Li, C. Kuo, and P. Cheng. Embedded wavelet packet image coder with fast rate-distortion optimized decomposition. In *Proc. SPIE: Visual Communication and Image Processing*, volume 3024, pages 1077–1088, 1997.

[8] J. Li and S. Lei. Arbitrary shape wavelet transform with phase alignment. *ISO/IEC JTC1/SC29/WG11 MPEG97/M2381*, July 1997; also to appear in *Proc. ICIP'98.*

[9] J. Li and S. Lei. Rate-distortion optimized embedding. In *Proc. Picture Coding Symposium*, pages 201–206, 1997.

[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Trans. Image Proc.*, 1:205–230, April 1992.

[11] S. Kim, J. Yi, J. Ra, J. Kim, and Y. Lee. Results of Core Experiment N1. *ISO/IEC JTC1/SC29/WG11 MPEG97/M2099*, April 1997.

[12] A.R. Calderbank, I. Daubechies, W. Sweldens, and B. Yeo. Lossless image compression using integer to integer wavelet transforms. In *Proc. IEEE Int. Conf. Image Processing*, 1997.

PSNR (dB)	filter	bit rate (bpp)					
		0.125	0.25	0.5	0.75	1	2
cmpnd1 (512 × 768)	haar	22.37	28.35	38.90	45.56	50.82	90.46
	9-3 tap	21.77	26.50	31.98	34.92	36.31	37.28
	9-7 tap	22.46	26.69	32.04	35.07	36.58	38.14
	adaptive	23.99	30.26	40.11	46.73	51.80	89.76
target (512 × 512)	haar	19.31	23.15	30.55	36.54	41.45	55.12
	9-3 tap	21.56	26.16	32.16	37.23	41.12	48.26
	9-7 tap	22.11	26.54	33.54	37.89	42.26	49.15
	adaptive	22.04	26.80	34.14	39.38	44.78	63.30
container (352 × 288)	haar	25.21	28.11	31.73	34.23	36.31	42.33
	9-3 tap	25.49	28.38	32.05	34.74	36.86	43.44
	9-7 tap	25.84	28.56	32.16	34.88	37.11	43.59
	adaptive	25.46	28.41	32.13	34.88	37.00	43.49

Table 1: Comparison of the R-D performance of the adaptive and nonadaptive wavelet transforms

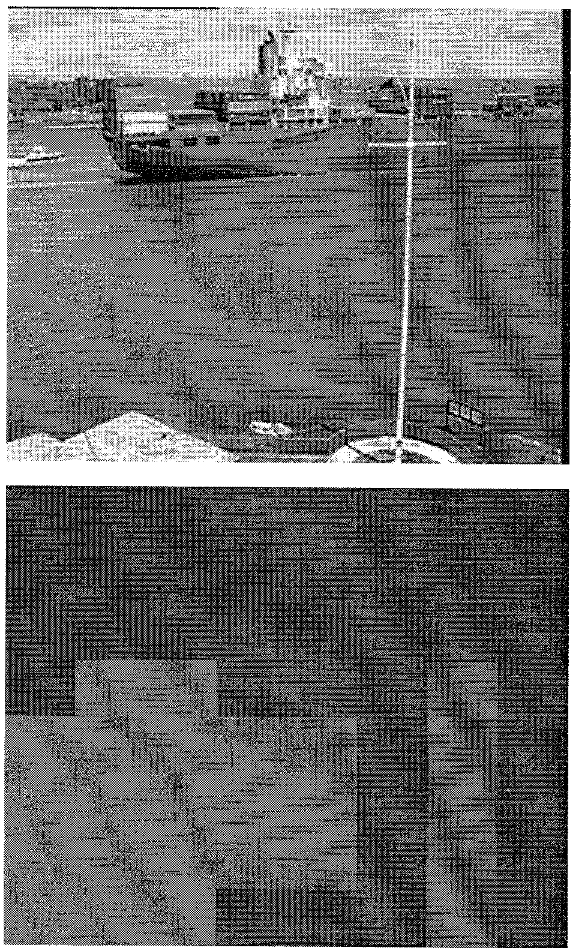

Figure 2: Top: original "container"; Bottom: segmentation map: gray for 9-7 tap, and black for 9-3 tap

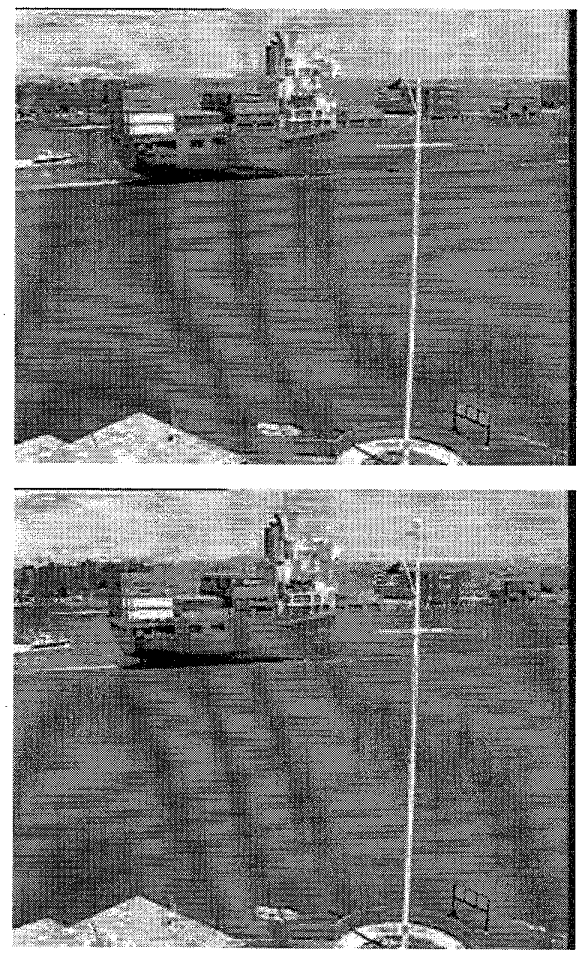

Figure 3: Decoded images at 0.25 bpp. Top: 9-7 tap; Bottom: adaptive filtering