

# A Fast and Efficient Topological Coding Algorithm for Compound Images

**Xin Li**

*Lane Department of Computer Science and Electrical Engineering, West Virginia University,  
Morgantown, WV 26506, USA  
Email: xinl@csee.wvu.edu*

*Received 11 September 2002 and in revised form 8 June 2003*

We present a fast and efficient coding algorithm for compound images. Unlike popular mixture raster content (MRC) based approaches, we propose to attack compound image coding problem from the perspective of modeling location uncertainty of image singularities. We suggest that a computationally simple two-class segmentation strategy is sufficient for the coding of compound images. We argue that jointly exploiting topological properties of image source in classification and coding stages is beneficial to the robustness of compound image coding systems. Experimental results have justified effectiveness and robustness of the proposed topological coding algorithm.

**Keywords and phrases:** compound image coding, level set, location uncertainty, topological property, strongly connected component, rate-distortion optimization.

## 1. INTRODUCTION

Compound image coding arises from various applications related to the storage and the distribution of document images. Document images usually contain the mixture of textual, graphical, and pictorial contents. Mixture raster content (MRC) model, a layered representation, has been widely used in the literature of compound image coding [1, 2, 3]. In spite of the popularity of MRC representation, the computational complexity of generating layers by image segmentation is prohibitive. For example, in MRC-based DjVu algorithm [2], segmentation stage often takes significantly longer time than the following coding stage.

In this paper, we attack compound image coding from a different perspective. We argue that computationally expensive document segmentation [4, 5, 6] is not an indispensable component to compound image coding system. Instead, we propose a simple yet effective two-class segmentation strategy to accommodate the compound nature of document images. The key observation is that image coding does not need to fully separate images from texts, graphics, and pictures as document segmentation does. For the task of compression, we advocate that it is sufficient to separate the compound image into two subsources: texts/graphics for which location uncertainty of image singularities should be directly modeled in the spatial domain, and pictures for which wavelet representations have shown to be appropriate [7, 8, 9, 10]. It is easy to see that such two-class model can be viewed as a special case of MRC representation (i.e., lift texts from mask

layer to foreground layer). The advantage with our two-class model is reduced complexity.

We will show that topological properties of two subsources provide a useful cue for fast segmentation. A linear-time algorithm based on finding strongly connected components [11] is proposed for the identification of textual/graphic regions. We then study how to exploit topological properties of image source in the coding stage where the support of each subsource can have arbitrary shape. The benefit of topological coding can be well understood from the perspective of modeling location uncertainty of image singularities. We argue that the fundamental limitation with data-filling approach [12] in MRC-based coding lies in its ignorance of topological information contained in the mask. The other advantage with joint exploitation of topological properties in segmentation and coding is improved robustness, that is, small errors in the segmentation result do not have significant impact on the overall coding performance. We also briefly study the rate-distortion (RD) optimization problem within the proposed two-class coding framework. Extensive experiment results are used to justify the effectiveness and robustness of the proposed compound image coding algorithm.

The rest of this paper is organized as follows. Section 2 introduces two-class model for compound image source and presents a fast topological segmentation algorithm. Section 3 describes topological coding algorithms in the spatial and the wavelet domain for texts/graphics and pictures, respectively. Section 4 studies RD optimization within the

framework of two-class coding. We report our simulation results in Section 5.

## 2. TWO-CLASS MODEL FOR COMPOUND IMAGE SOURCE

There are many different ways to model a compound image source. For example, MRC representation structures a compound image into three layers: mask (texts), foreground (graphics), and background (pictures). Extensive studies have been done on the problem of document segmentation [4, 5, 6], that is, separate texts, graphics, and pictures apart. We note that although document segmentation is a trivial task for human eyes, it has been one of the long open problems in computer vision research. Especially from the computational point of view, there is little evidence to believe that cumbersome document segmentation is an indispensable component to compound image coding systems.

We propose a comprised two-class segmentation strategy, that is, to view a document image as the mixture of two sub-sources: texts/graphics and pictures. Such two-class model can be viewed as a special case of three-layer MRC representation; but we argue that our model dramatically alleviates the computational burden on segmentation. The basic motivation behind our fast segmentation strategy is that the two sub-sources have different topological properties. That is, if we consider the level-set representation [13] for each subsource, textual/graphical regions typically have a support with regular shape and large size, while the level set in pictorial areas has irregular shape and small size (due to noise interference). Such distinction of the characteristics of level-set shape and size leads to a fast segmentation algorithm in the topological space.

### *Fast segmentation algorithm in the topological space*

- (1) Initialization:  $C(i, j) = 0$  (class 0), for all  $i, j$ .
- (2) Loop over level-set value  $k = 0-255$ .
  - (i) Generate level set  $\Omega_k = \{(i, j) \mid X(i, j) = k\}$  and its indicator function

$$I(i, j) = \begin{cases} 1, & (i, j) \in \Omega_k, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

- (ii) Identify each strongly connected component and calculate its topological parameters (size  $A$  and contour smoothness  $\alpha$ ).
  - (iii) If  $A < th_1$  or  $\alpha < th_2$ , set  $C(i, j) = 1$  (class 1).

In the above algorithm, strong connectivity refers to the connection through the eight nearest neighbors. The size of a set is defined by the number of pixels in the set and the contour smoothness is measured by the average of absolute differential tangent vector along the contour. It is well known that there exists a linear-time algorithm for finding strongly connected component in an undirected graph [11].

We note that the segmentation results generated by the above algorithm are mostly satisfactory but seldom perfect. A tantalizing question arises: how should we handle an imper-

fect segmentation result? Such issue is fundamentally important to the optimization of compound image coding systems but has been largely overlooked by the existing MRC-based approaches. We suggest that the robustness of compound image coding systems can be improved by jointly exploiting the topological properties of image subsource (connectivity and shape constraints) in both segmentation and coding stages. The above claim can be intuitively justified by thinking of compound image coding as a problem of resolving location uncertainty of image singularities. Segmentation errors are typically associated with ambiguity regions, that is, the set of pixels whose characteristics lie between texts/graphics and pictures (Figure 1). However, if the coding algorithms designed for each subsource, indeed, exploit the topological properties, the overall coding performance will be insensitive to the choice of coding algorithm, which compensates the wrong decision made by two-class segmentation.

## 3. TOPOLOGICAL CODING OF SUBSOURCES

In this section, we study the coding of two sub-sources with the segmentation result (binary classification map) available. We first introduce some notations. The compound image is decomposed into two sub-sources:  $\Omega = \Omega_{tg} \cup \Omega_{pi}$ , where  $\Omega_{tg}$  and  $\Omega_{pi}$  denote the support region of texts/graphics and pictures, respectively. For  $\Omega_{tg}$ , which consists of a small number of level sets, spatial domain is the appropriate space for modeling location uncertainty of image singularities (note that wavelet transform is not level-set preserving). For  $\Omega_{pi}$ , it is well known that wavelet space is suitable due to good energy compaction property of wavelet transform in both spatial and frequency domains. The challenge here is that both  $\Omega_{tg}$  and  $\Omega_{pi}$  have the support of arbitrary shape, which calls for coding algorithms capable of exploiting topological properties. It should be noted that a straightforward approach to handle arbitrary-shape support is by data filling [12] as used in most MRC-based coding systems. However, from the viewpoint of resolving location uncertainty of image singularities, data filling is unlikely to be optimal because it ignores useful topological information contained in the classification map. Instead, we propose to study topological coding for textual/graphical and pictorial sub-sources, respectively.

### 3.1. Textual/graphical subsource

The coding of textual/graphical images has been studied in the literature as palette-based image coding problem [14, 15, 16]. The main motivation behind palette-based coding is based on the following observations with texts/graphics:

- (1) there are typically far fewer colors than the number of pixels;
- (2) pixels of the same color tend to be contiguous.

The first observation implies that the subsource entropy is primarily determined by the location of image singularities (color transition). The second observation leads to the potential of exploiting topological properties during the actual coding process.

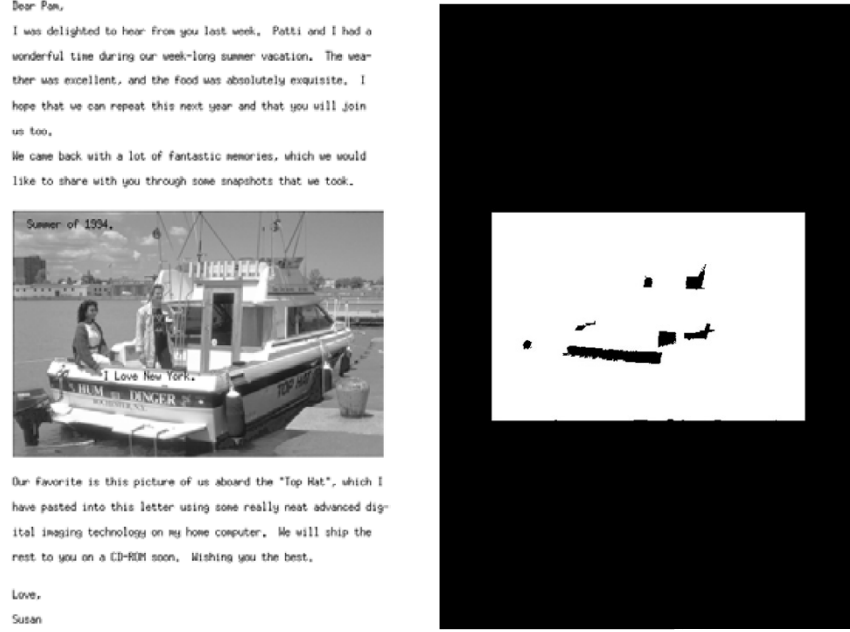


FIGURE 1: Left: original compound image; right: classification map.

We label all colors in the subspace by  $1, 2, \dots, N_c$ . Since  $N_c$  is usually a small number, the overhead of coding a palette of  $N_c$  colors is negligible. To code the index map  $\text{index}[X(i, j)]$ , we define the set of pixels having color  $k$  by

$$R_k = \{(i, j) \mid \text{index}[X(i, j)] = k, c(i, j) = 0\}, \quad (2)$$

and the union set of pixels whose color index is not less than  $k$  is thus given by

$$U_k = \bigcup_{l=k}^{N_c} R_l. \quad (3)$$

It is easy to see that  $R_k$  is related to  $U_k$  by

$$R_k = U_k - U_{k+1}, \quad (4)$$

where “the minus sign” denotes set subtraction operation and  $U_1 \supset U_2 \supset \dots \supset U_{N_c}$ . Equation (4) decomposes the original index map into  $N_c - 1$  binary maps (layers), which can be coded in  $N_c - 1$  passes. Each coding pass only needs to resolve the uncertainty of  $U_{k+1}$  from  $U_k$ , and therefore deals with a binary map with monotonically decreasing support. The existing context-based adaptive binary arithmetic coding (e.g., JBIG) can be easily modified to handle a binary map with arbitrary-shape support. For example, we can assign zero values to all the causal neighbors outside of the support. In fact, the binary classification map can also be incorporated into the above topological coding as an initial layer.

To exploit topological properties (observation 2), we note that both  $R_k$  (level sets) and  $U_k$  (union of level sets) are usually decomposed of strongly connected components with arbitrary shapes. Since the pixels with the same color tend to be

contiguous, the topological structure of  $U_k$ , which is already available at the decoder after the previous  $k - 1$  coding passes, carries useful information. That is, we can label each strongly connected set of  $U_k$  by 0 if all its pixels belong to  $R_k$ , by 1 if its pixels are all in  $U_{k+1}$ , and by 2 if it contains the mixture of  $R_k$  and  $U_{k+1}$ . It is easy to see that only the sets labeled by 2 need to be coded in the  $k$ th coding pass.

### 3.2. Pictorial subspace

It has been widely recognized that the success of wavelet coders for pictorial images is attributed to the effectiveness of modeling location uncertainty of image singularities in the wavelet space [17]. Our coding scheme consists of two stages (similar to LZC [8]): position coding (resolve location uncertainty of significant coefficients), and sign/magnitude coding for those coefficients which have been identified to be significant in the first stage. Most wavelet coders [7, 8, 9, 10] assume that images have regular support with a rectangular shape. However, the subspace of pictures is a partially masked image whose support  $\Omega_{\text{pi}}$  could have arbitrary shape. Recently, several works have appeared on the implementation of arbitrary-shape wavelet transform (ASWT) [18, 19, 20]. We employ the implementation based on lifting construction [20, 21].

Within the context of ASWT coding, it is natural to ask, how can we effectively exploit the topological information contained in the mask (binary classification map) to help resolve the location uncertainty of image singularities (significant coefficients)? We suggest the following two techniques. First, the positions of masked (do-not-care) coefficients are exactly known if we choose to preserve the correspondence of an image pixel to its mask value (class information) during

wavelet transform. In other words, there exists a one-to-one mapping between the mask in the spatial domain and its counterpart in the wavelet domain. Therefore, we can simply skip the masked coefficients when coding the high-band coefficients. Secondly, due to the good localization property of wavelet transform in both spatial and frequency domain, we could further exploit the topological property during the process of coding the position of significant coefficients. For example, for an isolated high-band coefficient (i.e., its prediction neighbors are all masked coefficients), we know *deterministically* that it would remain significant after ASWT because no prediction is available. Empirical study shows that such observation leads to noticeable bit savings in the stage of position coding.

#### 4. RATE-DISTORTION OPTIMIZATION FOR COMPOUND IMAGE SOURCE

Previous works such as optimizing block-thresholding segmentation [1] and RD optimized segmentation [3] emphasize on the study of RD optimization techniques during image segmentation for MRC-based coding. However, rate and distortion in the segmentation stage can only be an estimate because the actual RD characteristics depend on the segmentation result (like a chicken and egg problem). The other advantage offered by our two-class modeling paradigm is that it facilitates the RD optimization for compound image source.

We formulate RD optimization for a two-class source by minimizing the distortion  $D = D_0 + D_1$  under the constraint  $R_0 + R_1 \leq R$ , where  $R_0$  and  $R_1$  refer to the bit rate allocated to the two subsources, respectively. A commonly used technique for such constrained optimization problem is to use Lagrange multiplier. The Lagrange multiplier-based optimization technique [22] is to transform the original constrained problem into an unconstrained problem (minimize  $D + \lambda R$ , where  $\lambda$  is the Lagrange multiplier).

For two-class source model, we propose to decompose the original problem

$$\min (D_0 + D_1) + \lambda(R_0 + R_1) \quad (5)$$

into the following two independent problems:

$$\min D_0 + \lambda R_0, \quad \min D_1 + \lambda R_1. \quad (6)$$

For a single-class source, the optimal rate allocation is often achieved by iterative search along the operational RD curve [10]. Here, we solve the optimal rate allocation for two-class source in a similar fashion. Suppose for subsourse 0 (texts/graphics),  $N_c$  points along the operational RD curve  $(R_0^i, D_0^i)$ ,  $i = 0, 1, \dots, N_c - 1$ , have been found, each of which corresponds to one coding pass; for subsourse 1 (pictures), we can apply an embedding coding strategy similar to the existing wavelet coding schemes and obtain a collection of points  $(R_1^j, D_1^j)$ ,  $j = 0, 1, \dots$ , that are densely sampled along the operational RD curve. The following iterative RD optimization techniques are proposed for the two-class source.

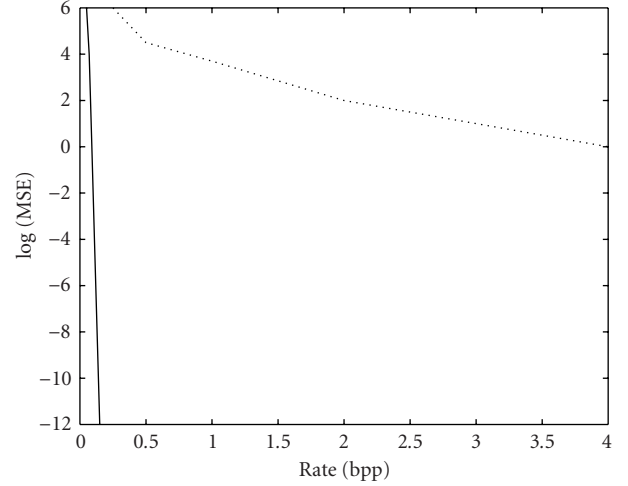


FIGURE 2: Operational rate-distortion curve comparison between subsourse 0 (solid) and subsourse 1 (dotted).

- (1) Initialization:  $i_{\text{opt}} = 0$ ,  $j_{\text{opt}} = 0$ , obtain  $(R_0^0, D_0^0)$  and  $(R_1^0, D_1^0)$ .
- (2) Iteration:
  - (i) for  $i = 1, 2, \dots$ , set  $\delta R = R_0^i - R_0^{i_{\text{opt}}}$  and  $\delta D = D_0^i - D_0^{i_{\text{opt}}}$ ; if  $\delta D / \delta R > \lambda$ , update  $i_{\text{opt}} = i$ ; otherwise continue;
  - (ii) for  $j = 1, 2, \dots$ , set  $\delta R = R_1^j - R_1^{j_{\text{opt}}}$  and  $\delta D = D_1^j - D_1^{j_{\text{opt}}}$ ; if  $\delta D / \delta R > \lambda$ , update  $j_{\text{opt}} = j$ ; otherwise stop.

Due to the distinct characteristics of two subsources, their operational RD curves differ dramatically. Figure 2 shows an example of the operational RD curves for a portion of cmpnd2 image. It can be seen that the slope of subsourse 0 is dramatically larger than that of subsourse 1. Therefore, the subsourse 0 has higher priority than the subsourse 1 when the Lagrange multiplier is large (at very low bit rates). This matches our intuition because the distortion in texts/graphics is often more visible than that in pictures.

#### 5. SIMULATION RESULTS

In this section, we report our experiment results with two compound images in the JPEG2000 test set: cmpnd1 ( $512 \times 768$ ) and cmpnd2 ( $5120 \times 6624$ ). The cmpnd2 image is composed of 8 concatenated small subimages. Since the size of cmpnd2 is huge, we choose to cut out one subimage (sized  $1568 \times 1568$ ) from cmpnd2 and use it as the test image. It should be noted that both cmpnd1 and cmpnd2 are computer-generated images containing no noise. Coding of noisy compound images (e.g., scanned documents) is beyond the scope of this paper.

We have implemented a new topological image coder based on two-class modeling of compound image source. The topological coder in the spatial domain employs a sixth-order context model at each coding pass. The implementation of adaptive binary arithmetic coder (QM coder) is taken



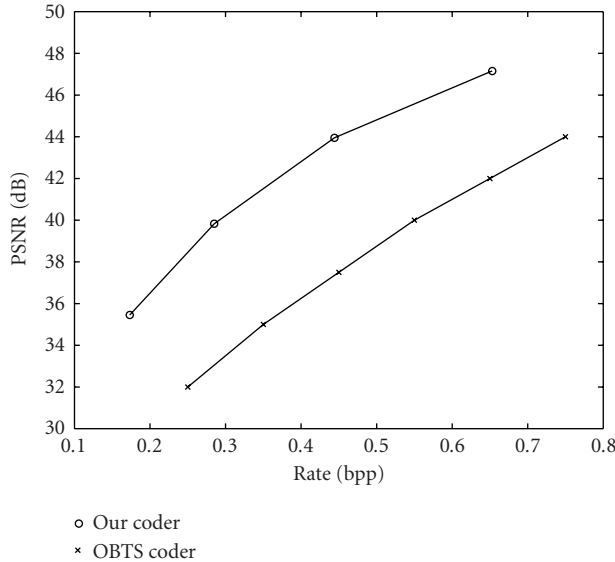


FIGURE 3: Rate-distortion performance comparison for cmpnd1 between our two-class coder and OBTS coder [1].

from the existing JBIG standard. The topological coding in the wavelet space is based on an implementation of masked Daubechies' 9-7 transform [23]. A simplified two-stage coding algorithm similar to LZC is used to code the unmasked wavelet coefficients. It should be noted that both JBIG and simplified LZC do not represent state-of-the-art coders. More sophisticated coders such as JBIG2 and JPEG2000 could lead to even better coding performance. The coding results reported here are mainly for the purpose of justifying the efficiency of the proposed two-class source modeling and topological coding techniques. Decoder executable and encoded bit streams in our experiments can be downloaded from <http://www.ee.princeton.edu/~lixin/cmpnd.htm>.

We first compare our two-class image coder and the OBTS coder for cmpnd1 image. It appears that the image quality offered by our coder at 0.285 bpp is visually lossless compared to the original. As an example, the bits spent on textual/graphical and pictorial subsources are 20 480 and 91 144, respectively at the rate of 0.285 bpp. The coding results of OBTS are cited from [1, Figure 9]. The RD performance comparison is shown in Figure 3. Large PSNR improvements (greater than 6 dB) can be observed. We note that such significant coding gain should be interpreted properly. Wavelet coding techniques (LZC, JPEG2000) typically could achieve at least 3 dB gain over DCT-based coding techniques (e.g., JPEG employed in OBTS coder). Therefore, partial credits of 6 dB gain go to wavelet coding techniques. Nonetheless, topological coding algorithms described in Section 3 do achieve impressive coding performance.

Figure 1 shows the segmentation result for cmpnd1 image. The segmentation result of texts/graphics from pictures is mostly satisfactory despite a few wrongly classified regions

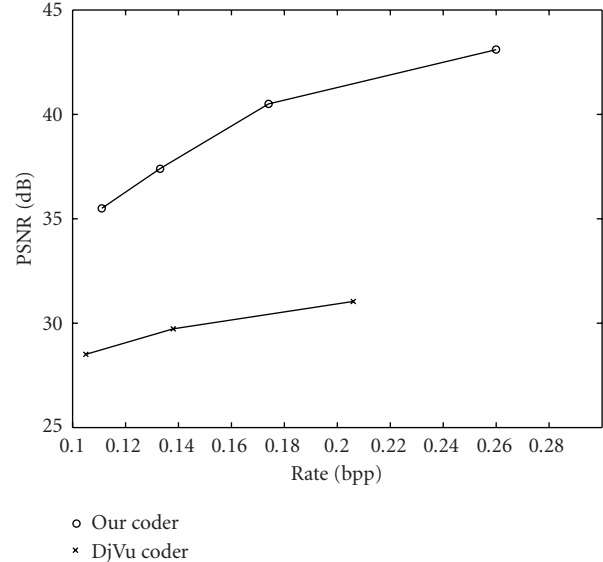


FIGURE 4: Rate-distortion performance comparison for cmpnd2 between our two-class coder and DjVu coder.

scattered in the pictorial content. Indeed, those segmentation errors are due to the fact that some areas in the pictorial content are locally constant, causing the ambiguity. To testify the robustness of our coder, we have generated an optimal mask manually for cmpnd1 image to see how it could further enhance the coding performance. It appears that the PSNR loss brought by segmentation errors is quite modest (less than 0.3 dB). We conclude that the ambiguity regions in the pictorial content can be efficiently handled by topological coding in either spatial or wavelet domain.

We also compare our two-class coder and the well-known DjVu coder for cmpnd2 image. The DjVu coder implementation is already available as a commercial software (DjVuShop 2.0). We have chosen the default parameter settings for color document selection but enforce the resolution of all layers to be 300 dpi (lower resolution for text color and background only renders worse PSNR results). Figure 4 shows the RD performance comparison between our coder and DjVu coder. Again, the PSNR improvements are in the range of 6–10 dB. Figure 5 compares the various portions of cmpnd2 image decoded by our coder at 0.133 bpp and by DjVu at 0.138 bpp (texts/graphics: 199664 bits, pictures: 126312 bits). Subjective quality improvements are also striking. Such dramatic improvements are partially due to the fact that DjVu coder mainly targets at web-browsing applications where compression ratio is extremely high. The RD performance of DjVu coder is far from being optimized at the bit rate of above 0.1 bpp. However, we believe that the gap will not be fully closed even with carefully tuning the coding parameters of DjVu coder. As we can see from Figure 6, document segmentation results generated by DjVu algorithm are relatively poor and coding efficiency loss is inevitable.

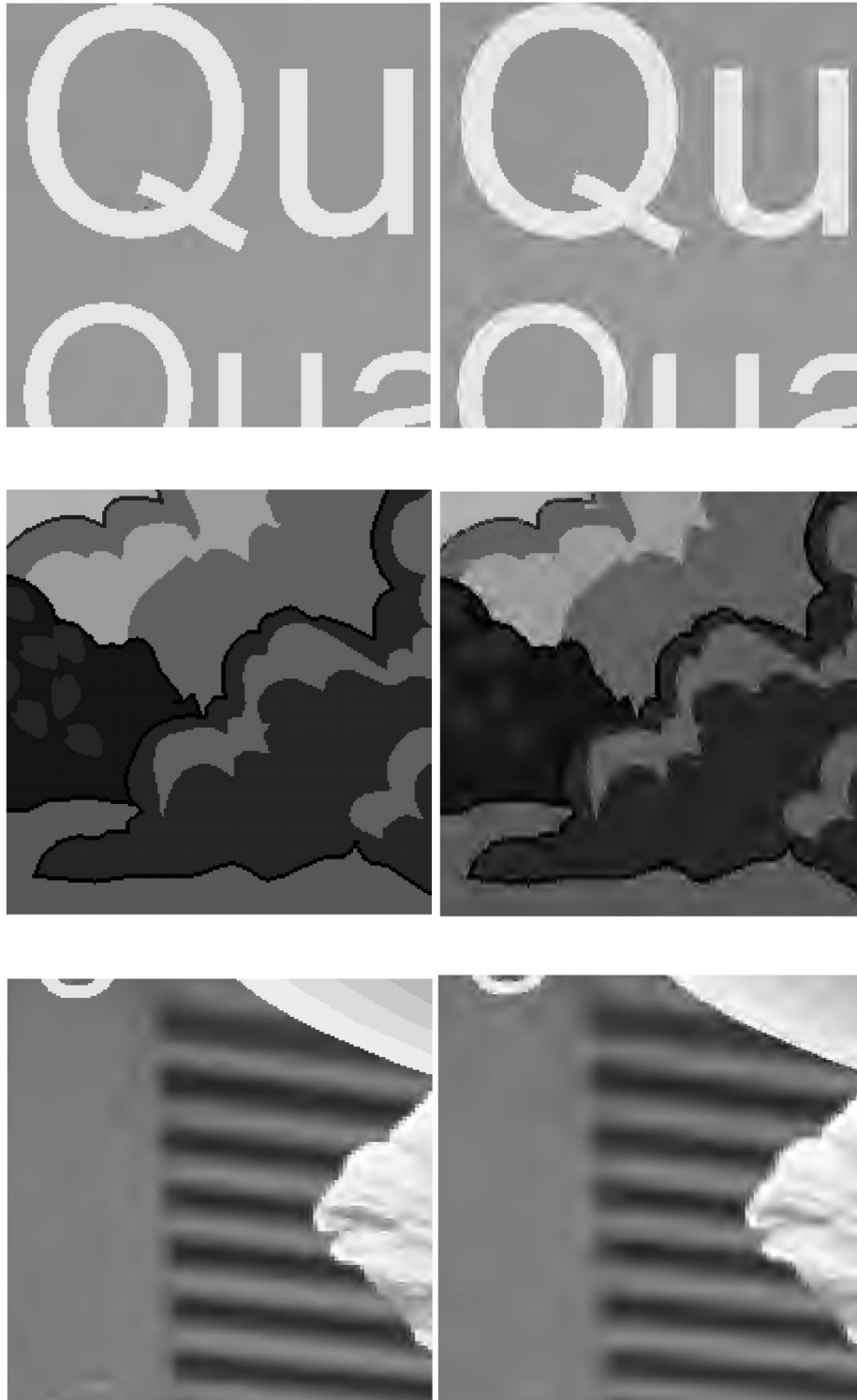


FIGURE 5: Comparison of portions of decoded cmpnd2 images by our two-class coder at 0.133 bpp, PSNR = 37.45 dB (left) and by DjVu coder at 0.138 bpp, PSNR = 29.73 dB (right).

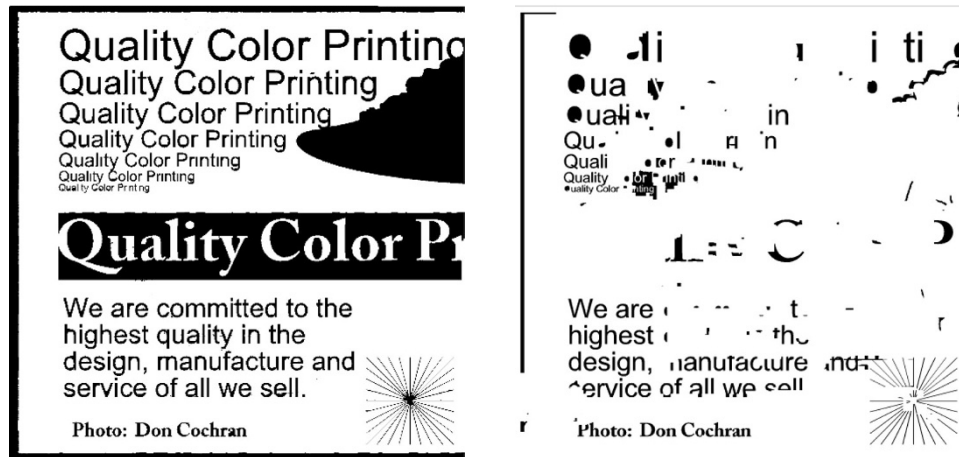


FIGURE 6: Comparison of the classification map generated by our algorithm (left) and the mask layer generated by DjVu algorithm (right) for cmpnd2.

## REFERENCES

- [1] R. L. de Queiroz, Z. Fan, and T. D. Tran, "Optimizing block-thresholding segmentation for multilayer compression of compound images," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1461–1471, 2000.
- [2] L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. Le Cun, "High quality document image compression with DjVu," *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 410–425, 1998.
- [3] H. Cheng and C. Bouman, "Document compression using rate-distortion optimized segmentation," *Journal of Electronic Imaging*, vol. 10, no. 2, pp. 460–474, 2001.
- [4] H. Cheng, C. A. Bouman, and J. P. Allebach, "Multiscale document segmentation," in *IS&T 50th Annual Conference*, pp. 417–425, Cambridge, Mass, USA, May 1997.
- [5] A. A. Zlatopolsky, "Automated document segmentation," *Pattern Recognition Letters*, vol. 15, no. 7, pp. 699–704, 1994.
- [6] M. Nadler, "A survey of document segmentation and coding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 28, pp. 240–262, 1984.
- [7] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.
- [8] D. Taubman and A. Zakhor, "Multirate 3-d subband coding of video," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 572–588, 1994.
- [9] Z. Xiong, K. Ramchandran, and M. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Processing*, vol. 6, no. 1, pp. 677–693, 1997.
- [10] D. Taubman, "High-performance scalable image compression with EBCOT," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.
- [11] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, Mass, USA, 1990.
- [12] R. L. de Queiroz, "On data filling algorithms for MRC layers," in *Proc. IEEE International Conference on Image Processing (ICIP '00)*, vol. 2, pp. 586–589, Vancouver, British Columbia, Canada, September 2000.
- [13] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, NY, USA, 2002.
- [14] P. Ausbeck, "The piecewise-constant image model," *Proceedings of the IEEE*, vol. 88, no. 11, pp. 1779–1789, 2000.
- [15] X. Li, "Embedded coding of palette images in the topological space," in *Data Compression Conference (DCC '02)*, p. 462, Snowbird, Utah, USA, April 2002.
- [16] S. Forchhammer and O. R. Jensen, "Content layer progressive coding of digital maps," in *Data Compression Conference (DCC '00)*, pp. 233–242, Snowbird, Utah, USA, March 2000.
- [17] R. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 719–746, 1992.
- [18] J. Li and S. Lei, "Arbitrary shape wavelet transform with phase alignment," in *Proc. IEEE International Conference on Image Processing (ICIP '98)*, pp. 683–687, Chicago, Ill, USA, October 1998.
- [19] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 5, pp. 725–743, 2000.
- [20] P. Y. Simard and H. S. Malvar, "A wavelet coder for masked images," in *Data Compression Conference (DCC '01)*, pp. 93–102, Snowbird, Utah, USA, March 2001.
- [21] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.*, vol. 29, no. 2, pp. 511–546, 1997.
- [22] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.
- [23] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

**Xin Li** received the B.S. degree with highest honors in electronic engineering and information science from University of Science and Technology of China, Hefei, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2000. He was a member of technical staff with Sharp Laboratories of America, Camas, Wash, from August 2000 to December 2002. Since January 2003, he has been a faculty member in Lane Department of Computer Science and Electrical Engineering. His research interests include image/video coding and processing. Dr. Li received the Best Student Paper Award at the Conference of Visual Communications and Image Processing, San Jose, Calif, in January 2001.

