

# Rate-Distortion Based Segmentation for MRC Compression

Hui Cheng<sup>a</sup>, Guotong Feng<sup>b</sup> and Charles A. Bouman<sup>b</sup>

<sup>a</sup>Sarnoff Corporation, Princeton, NJ 08543-5300, USA

<sup>b</sup>Purdue University, West Lafayette, IN 47907-1285, USA

## ABSTRACT

Effective document compression algorithms require scanned document images be first segmented into regions such as text, pictures and background. In this paper, we present a document compression algorithm that is based on the 3-layer (foreground/mask/background) MRC (mixture raster content) model. This compression algorithm first segments a scanned document image into different classes. Then, each class is transformed to the 3-layer MRC model differently according to the property of that class. Finally, the foreground and the background layers are compressed using JPEG with customized quantization tables. The mask layer is compressed using JBIG2. The segmentation is optimized in the sense of rate-distortion for the 3-layer MRC representation. It works in a closed loop fashion by applying each transformation to each region of the document and then selecting the method that yields the best rate-distortion trade-off. The proposed segmentation algorithm can not only achieve a better rate-distortion trade-off, but also produce more robust segmentations by eliminating those misclassifications which can cause severe artifacts. At similar bit rates, our MRC compression with the rate-distortion based segmentation can achieve a much higher subjective quality than state-of-the-art compression algorithms, such as JPEG and JPEG-2000.

**Keywords:** Segmentation, Document Compression, MRC Compression, Rate-Distortion.

## 1. INTRODUCTION

To achieve high quality document reproduction and rendering, paper documents must be scanned at a minimum of 400-600 dpi (dots per inch). A single page of a color document scanned at 400-600 dpi requires approximately 45-100 Megabytes of storage. Consequently, practical systems for processing color documents require document compression methods that achieve high compression ratios with very low distortion.

Since document images contain well defined regions with distinct characteristics, such as text, line graphics, continuous-tone pictures, halftone pictures and background, they are also referred as mixture raster content (MRC). Traditional compression algorithms, such as JPEG, tend to perform poorly on document images, because these algorithms assume that the input image is spatially homogeneous. Therefore, new compression approaches need to be developed for MRC applications.

Most existing MRC compression algorithms can be crudely classified as block-based approaches and layer-based approaches. Block-based approaches<sup>1-4</sup> segment non-overlapping blocks of pixels into different classes, and compress each class differently according to its characteristics. On the other hand, layer-based approaches<sup>5-7</sup> partition a document image into different layers, such as the background layer and the foreground layer. Then, each layer is coded as an image independently from other layers. Most layer-based approaches use the 3-layer (foreground/mask/background) representation proposed in the ITU's Recommendations T.44 for mixed raster content (MRC). The foreground layer contains the color of text and line graphics, and the background layer contains pictures and background. The mask is a bi-level image which determines, for each pixel in the reconstructed image, if the foreground color or the background color should be used.

The performance of a document compression system is directly related to the segmentation algorithm used to produce the binary mask. A good segmentation can not only lower the bit rate, but also lower the distortion.

---

H. Cheng: E-mail: hcheng@sarnoff.com, Telephone: 1 (609) 734-2492, Visual Information Systems, Sarnoff Corporation, Princeton, NJ 08543-5300, USA

G. Feng and C.A. Bouman: E-mail: {fengg, bouman}@ecn.purdue.edu, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285, USA

On the other hand, those artifacts which are most damaging are often caused by misclassifications. Some segmentation algorithms which have been proposed for document compression use features extracted from the discrete cosine transform (DCT) coefficients to separate text blocks from picture blocks.<sup>2,8</sup> Other segmentation algorithms are based on the features extracted directly from the input document image.<sup>5,9</sup> However, most of these algorithms segment a document image based solely on the document image. They do not consider the compression algorithms used for each class and the rate-distortion trade-off preferred by a user. Therefore, we refer to these algorithms as direct segmentation algorithms.

One approach to designing a good document coder is to optimized the operational rate-distortion.<sup>3,4,10</sup> In fact, operational rate-distortion methods have come into wide use for image and video coders.<sup>11</sup> In previous work, de Queiroz applied this technique to finding optimal thresholds for block segmentation.<sup>3</sup> Cheng and Bouman used the rate-distortion optimization criteria to compute the document segmentation that produced approximately the best quality/bit rate trade-off for each document begin compressed.<sup>4,10</sup> However, this method used a block based method rather than the more standard layer based approach of the MRC standard.

In this paper, we present rate-distortion based segmentation algorithm which supports the standard 3-layer MRC format and is based on conventional JPEG compression for the foreground and background layers. The algorithm first segments  $8 \times 8$  non-overlapping blocks of pixels into different classes, such as text, picture and background. Then, each block is represented differently using a 3-layer MRC model according to the property of that class. The  $8 \times 8$  block segmentation is computed by optimizing the actual rate-distortion performance for the image being coded. It works by first applying each class to each region of the image, and then selecting the class for each region which approximately maximizes the rate-distortion performance. The optimization is based on the measured distortion and an estimate of the bit rate for each class. Compared with direct image segmentation algorithms, the rate-distortion based segmentation has several advantages. First, it produces more robust segmentations. Intuitively, misclassifications which cause severe artifacts are eliminated because all possible classes are tested for each block of the image. In addition, it allows us to control the trade-off between the bit rate and the distortion by adjusting a weight. For each weight set by a user, an approximately optimal segmentation is computed to achieve the best rate-distortion trade-off.

We test our algorithm on both scanned and noiseless synthetic document images. Experimental results show that, in the same range of compression ratios, the 3-layer MRC with using the proposed rate-distortion based segmentation results in a much higher subjective quality than well-known compression algorithms, such as JPEG and JPEG-2000, especially in text and graphic regions,

## 2. 3-LAYER MRC COMPRESSION

As shown in Fig. 1, the 3-layer MRC model represents a document image using three layers: a foreground layer, a background layer and a mask layer. The mask layer is a binary image. It is used to determine, for each pixel in the reconstructed image whether the foreground color or the background color should be used. Let  $(u, v)$  be a 2-D vector that defines a pixel location. Denote the foreground as  $F$ , the background as  $B$ , and the binary mask as  $M$ . Then, the image reconstructed from a 3-layer MRC model,  $G$ , can be written as

$$G(u, v) = M(u, v)F(u, v) + (1 - M(u, v))B(u, v)$$

Ideally, the foreground layer should contain colors of text, and the background layer should contain continuous-tone, halftone pictures and background colors. Therefore, both the foreground and the background layers have significant spatially redundancy and can be compressed aggressively. On the other hand, the mask layer contains the contours of text and other fine image structures. Although the mask layer needs high spatial resolution to accurately represent text contours and fine image structures, it has only two colors, and can be compressed effectively using token based compression algorithm, such as JBIG2 [2]. Both the foreground layer and the background layer can be compressed using any compression algorithm. However, for real-time copying and scanning applications, we compress both layers using JPEG, but with different quantization tables.

To use a 3-layer MRC model, a document image needs to be first segmented into foreground and background. Since in this paper, JPEG is used to compress both the foreground and the background, the segmentation of

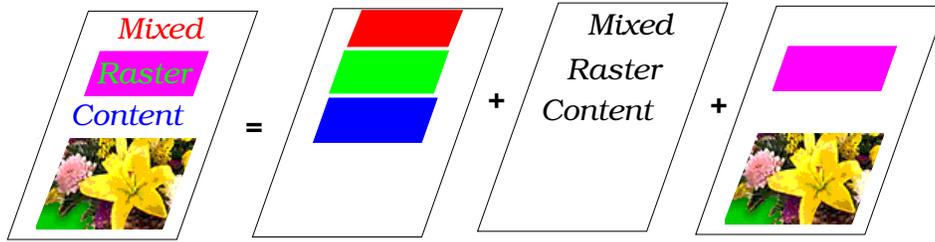


Figure 1: Illustration of 3-layer MRC representation.

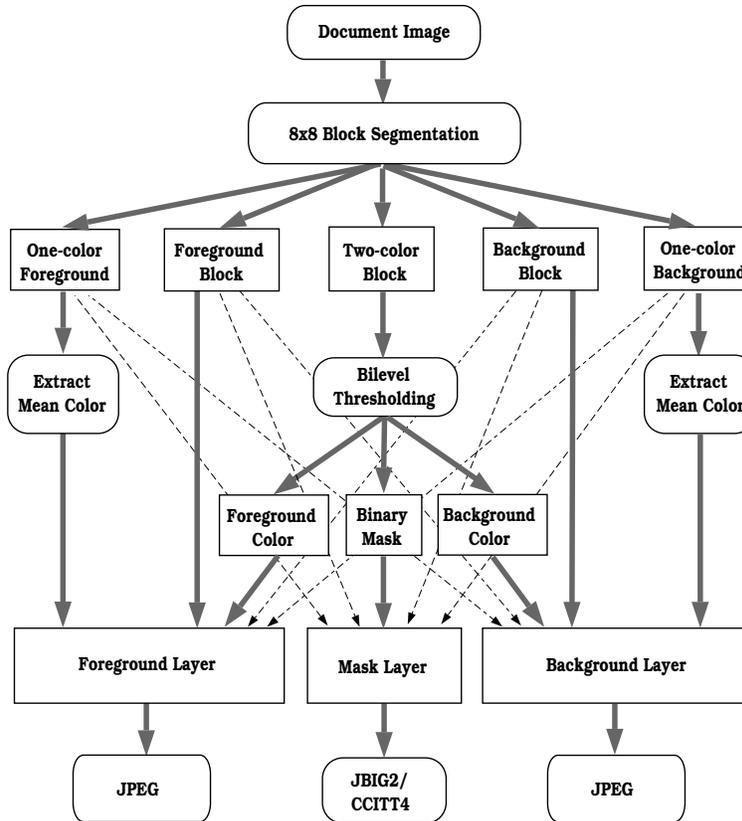


Figure 2. Flow diagram of the rate-distortion optimized 3-layer MRC compression system. For example, for a foreground block, the corresponding block in the background is set with the mean color of the previous block and the mask block is set to be 1.

the whole image can be simplified to the segmentation of pixels into  $8 \times 8$  blocks. For each  $8 \times 8$  block of pixels, there are three possibilities: (1) all pixels belong to the foreground, (2) all pixels belong to the background or (3) some of pixels belong to the foreground and others belong to the background. If all pixels of an  $8 \times 8$  block belong to foreground, the block is called a Foreground block. If all pixels of an  $8 \times 8$  block belong to background, the block is called a Background block. If some pixels of the block belong to foreground and others belong to background, we call the block a Two-color block. In addition, if a Background block can be represented with only one color with acceptable distortion, it is called an One-color Background block. If a Foreground block can be represented with only one color, it is called an One-color Foreground block.

Two-color blocks are effective in compressing text or line graphics. Text and line graphics need to be coded with high spatial resolution, but they can tolerate low color resolution. Therefore, for each Two-color block, a bilevel thresholding is used to extract two colors (one foreground color and one background color) and a binary

mask. Finally, a Two-color block is represented with a foreground block with a constant color, a background block with a constant color and a binary  $8 \times 8$  mask.

Background blocks should be from background regions. Blocks of continuous-tone or halftone pictures that can code well at the JPEG quality factor used for background are also classified as Background blocks. The background layer is often compressed aggressively with customized quantization tables. If the background is uniform, One-color Background blocks can be used to represent the whole block with only one color, further improving the compression.

However, in order to achieve high quality reconstruction, some “difficult” regions within continuous-tone, halftone pictures and graphics need to be compressed at a higher quality level than what is used for background. Therefore, for foreground, different quantization tables with much lower quantization steps than those used for background blocks are used for both luminance and chrominance. Other than the regions that can not be compressed well enough in background layer, foreground also contains colors of text, line art and other detailed document regions. However, the color of text and line art is often similar over large scale. Therefore, they have few high frequency components, and can be compressed with lower quantization steps without significantly increasing the bit rate of the foreground.

The details of compression of each of these five classes are described in the following subsections. The flow diagram of our compression algorithm is shown in Fig. 2. Throughout this paper, we use  $y$  to denote the original image and  $x$  to denote its  $8 \times 8$  block segmentation. Also,  $y_i$  denotes the  $i$ -th  $8 \times 8$  block in the image, where the blocks are taken in raster order, and  $x_i$  denotes the class label of block  $i$ , where  $0 \leq i < L$ , and  $L$  is the total number of blocks. The set of class labels is then  $\mathcal{N} = \{Two, OnB, OnF, Fgd, Bgd\}$ , where *Two*, *OnB*, *OnF*, *Fgd*, *Bgd* represent Two-color, One-color Background, One-color Foreground, Foreground and Background blocks, respectively.

## 2.1. MRC Representation of One-color Background and One-color Foreground Blocks

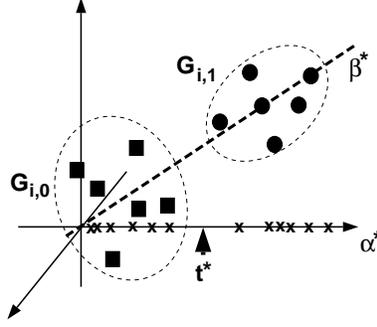
Each One-color Background block and One-color Foreground block is represented by a 24-bit color. For One-color Background blocks, we first extract the mean color of each block. Then, set all pixels of the corresponding block in the background layer with the mean color, and set all pixels of the corresponding block in the foreground layer with the mean color of the previous block in raster order. The corresponding block in the mask layer is set to 0 indicating that the whole block belongs to the background layer.

Similarly, we set all pixels of One-color Foreground blocks in the foreground layer to the mean colors of the corresponding blocks in the original image, and set all pixels of One-color Foreground blocks in the background layer with the mean color of the previous block in raster order. The corresponding block in the mask layer is set to 1 since the whole block belongs to the foreground layer.

## 2.2. MRC Representation of Two-color Blocks

The Two-color class is designed to compress blocks which can be represented well by two colors, such as text blocks. Since Two-color blocks need to be coded with high spatial resolution, but can tolerate low color resolution, each Two-color block is represented by two 24-bit colors and a binary mask. The bilevel thresholding algorithm that we use for extracting the two colors and the binary mask uses a minimal mean squared error (MSE) thresholding followed by a spatially adaptive refinement. The algorithm is performed on two block sizes. First,  $8 \times 8$  blocks are used. But sometimes an  $8 \times 8$  block may not contain enough samples from both color regions for a reliable estimate of the colors of both regions and the binary mask. In that case, a  $16 \times 16$  block centered at the  $8 \times 8$  block will be used instead.

The minimal MSE thresholding algorithm is illustrated in Fig. 3. For a Two-color block  $y_i$ , we first project all colors of  $y_i$  onto the color axis  $\alpha^*$  which has the largest variance among three color axes. The thresholding is done only on  $\alpha^*$ . Since we are mainly interested in high quality document images where text is sharp and the noise level is low, the projection step significantly lowers the computation complexity without sacrificing the quality of the bilevel thresholding. For a threshold  $t$  on  $\alpha^*$ ,  $t$  partitions all colors into two groups. Let  $E_i(t)$  be the MSE, when colors in each group are represented by the mean color of that group. We compute the value  $t^*$  which minimizes  $E_i(t)$ . Then,  $t^*$  partitions the block into two groups,  $G_{i,0}$  and  $G_{i,1}$ , where the mean color



**Figure 3.** Minimal MSE thresholding. We use  $\alpha^*$  to denote the color axis with the largest variance, and  $\beta^*$  to denote the principle axis.  $t^*$  is the optimal threshold on  $\alpha^*$ , and x's are the samples projected on  $\alpha^*$ .

of  $G_{i,0}$  has a larger  $l_1$  norm than the mean color of  $G_{i,1}$ . Let  $c_{i,j}$  be the mean color of  $G_{i,j}$ , where  $j = 0, 1$ . Then,  $\|c_{i,0}\|_1 > \|c_{i,1}\|_1$  is true for all  $i$ . We call  $c_{i,0}$  the background color of block  $i$ , and  $c_{i,1}$  the foreground color of block  $i$ . The binary mask which indicates the locations of  $G_{i,0}$  and  $G_{i,1}$  is denoted as  $b_{i,m,n}$ , where  $b_{i,m,n} \in \{0, 1\}$ , and  $0 \leq m, n \leq 7$ .

The minimal MSE thresholding usually produces a good binary mask. But  $c_{i,0}$  and  $c_{i,1}$  are often biased estimates. This is mainly caused by the boundary points between two color regions since their colors are a combination of the colors of the two regions. Therefore,  $c_{i,0}$  and  $c_{i,1}$  need to be refined. Let a point in block  $i$  be an internal point of  $G_{i,j}$ , if the point and its 8-nearest neighbors all belong to  $G_{i,j}$ . If a point is not an internal point of either  $G_{i,0}$  or  $G_{i,1}$ , we call it a boundary point. Also, denote the set of internal points of  $G_{i,j}$  as  $\tilde{G}_{i,j}$ . If  $\tilde{G}_{i,j}$  is not empty, we set  $c_{i,j}$  to the mean color of  $\tilde{G}_{i,j}$ . When  $\tilde{G}_{i,j}$  is empty, we can not estimate  $c_{i,j}$  reliably. In this case, if the current block size is  $8 \times 8$ , we will enlarge the block to  $16 \times 16$  symmetrically along all directions, and use the same bilevel thresholding algorithm to extract two colors and a  $16 \times 16$  mask. Then, the two colors extracted from the  $16 \times 16$  block are used as  $c_{i,0}$  and  $c_{i,1}$ , and the middle portion of the  $16 \times 16$  mask is used as  $b_{i,m,n}$ . If  $\tilde{G}_{i,j}$  is empty, and the current block is a  $16 \times 16$  block,  $c_{i,j}$  will be used as it is without refinement.

For a Two-color block, the corresponding pixels in background are set to the background color  $\{c_{i,0} | x_i = Two\}$ , and the corresponding pixels in foreground are set to the foreground color  $\{c_{i,1} | x_i = Two\}$ . The mask values are set to  $b_{i,m,n}$ .

### 2.3. MRC Representation of Foreground and Background Blocks

For a Foreground block, copy the original block to the foreground, set pixels of the background to the mean color of the previous background block in raster order, and set the block in the mask to 1.

Similarly, for a Background block, copy the block to background, set pixels of the foreground to the mean color of the previous foreground block in raster order, and set the block in the mask to 0.

### 2.4. Compression of 3-Layer MRC

The foreground and background layers are both compressed using JPEG. For the experiments, the background layer is compressed using quantization tables similar to the standard JPEG quantization tables at quality level 20; however, the quantization steps for the DC coefficients in both luminance and chrominance are set to 15. The foreground layer is compressed using the standard JPEG quantization tables at quality level 30.

The mask layer is compressed by a JBIG2 coder using the lossless soft pattern matching technique.<sup>12</sup>

### 3. RATE DISTORTION BASED SEGMENTATION FOR MRC

In order to segment each  $8 \times 8$  block of pixels into the five classes discussed in section 2, we propose a rate-distortion optimized segmentation. A number of segmentation algorithms have been proposed to segment a document image into foreground and background.<sup>2,3,5,8-10</sup> Most of these algorithms are direct segmentation algorithms. Direct segmentation algorithms segment a document image based solely on the document image. In contrast, the rate-distortion based method works in a closed loop fashion by applying each coding method to each region of the document and then selecting the method that yields the best rate-distortion trade-off. The rate-distortion based method insures that each block is coded using the method which is best suited for it. This results in more robust segmentations which yield a better rate-distortion trade-off at every quality level. The rate-distortion approach proposed in this paper is closely related to the approach introduced in.<sup>10</sup> However, the previous approach is designed for a block based document compression system called the multilayer compression system, not the 3-layer MRC representation.

Let  $R(y|x)$  be the number of bits required to code  $y$  with block segmentation  $x$ , and  $D(y|x)$  be the total distortion resulting from coding  $y$  with segmentation  $x$ . Then, the rate-distortion based segmentation,  $x^*$ , is

$$x^* = \arg \min_{x \in \mathcal{N}^L} \{R(y|x) + \lambda D(y|x)\}, \quad (1)$$

where  $\lambda$  is a non-negative real number which controls the trade-off between bit rate and distortion. In our approach, we assume that  $\lambda$  is a constant controlled by a user which has the same function as the quality level in JPEG. In addition, since the segmentation is only used to guide the compression and not used in the reconstruction. The block segmentation map does not need to be sent to the decoder. Therefore, no bits are required for the segmentation map.

To compute the rate-distortion based segmentation, we need to estimate the number of bits required for coding each block as each class, and the distortion of coding each block as each class. For computational efficiency, we assume that the number of bits required for coding a block only depends on the image data and class labels of that block and the previous block in raster order. We also assume that the distortion of a block can be computed independently from other blocks. With these assumptions, (1) can be rewritten as

$$x^* = \arg \min_{\{x_0, x_1, \dots, x_{L-1}\} \in \mathcal{N}^L} \sum_{i=0}^{L-1} \{R_i(x_i|x_{i-1}) + \lambda D_i(x_i)\}, \quad (2)$$

where  $R_i(x_i|x_{i-1})$  is the number of bits required to code block  $i$  using class  $x_i$  given  $x_{i-1}$ , and  $D_i(x_i)$  is the distortion produced by coding block  $i$  as class  $x_i$ . After the rate and distortion are estimated for each block using each coder, (2) can be solved by a dynamic programming technique similar to that used in.<sup>13</sup>

An important aspect of our approach is that we use a class-dependent distortion measure. This is desirable because, for document images, different regions, such as text, background and pictures, can tolerate different types of distortion. For example, errors in high frequency bands are less important in background and picture regions, but they can cause severe artifacts in text regions.

In the following sections, we specify how to compute the rate and distortion terms for the 3-layer MRC model. The expressions for rate are often approximate due to the difficulties of accurately modeling high performance coding methods such as JBIG2. However, our experimental results indicate that these approximations are accurate enough to consistently achieve good compression results.

#### 3.1. Bit Rate Estimate

Although the five different classes, (*Two, OnB, OnF, Fgd, Bgd*), are transformed to the 3-layer MRC model differently, they are all represented by one  $8 \times 8$  block in the foreground, one block in the background and one in the mask layer. Therefore, the number of bits required for coding any block consists of the number of bits required for the foreground, the number of bits for the background and the number of bits for the mask.

The bits required for coding either foreground or background block  $i$  can be further divided into two parts: the bits required for coding the luminance of block  $i$ , denoted as  $R_i^l(x_i|x_{i-1})$ , and the bits for coding the chrominance, denoted as  $R_i^c(x_i|x_{i-1})$ . Therefore,

$$R_i(x_i|x_{i-1}) = R_i^l(x_i|x_{i-1}) + R_i^c(x_i|x_{i-1}).$$

Let  $\alpha_i^d(x_i)$  be the quantized DC coefficients of the luminance using the quantization table specified by class  $x_i$ , and  $\alpha_i^a(x_i)$  be the vector which contains all 63 quantized AC coefficients of the luminance of block  $i$ . Using the standard JPEG Huffman tables for luminance,  $R_i^l(x_i|x_{i-1})$  can be computed as

$$R_i^l(x_i|x_{i-1}) = r_d [\alpha_i^d(x_i) - \alpha_{i-1}^d(x_{i-1})] + r_a [\alpha_i^a(x_i)],$$

where  $r_d[\cdot]$  is the number of bits used for coding the difference between two consecutive DC coefficients of the luminance component, and  $r_a[\cdot]$  is the number of bits used for coding AC coefficients. The formula for calculating  $r_d[\cdot]$  and  $r_a[\cdot]$  is specified in the JPEG standard.<sup>14</sup> Notice that  $R_i(x_i|x_{i-1})$  is the exact number of bits required for coding the luminance component using JPEG.

Since the two chrominance components are subsampled  $2 \times 2$ , we approximate the number of bits for coding the chrominance components of an  $8 \times 8$  block  $i$ ,  $R_i^c(x_i|x_{i-1})$ , as follows. Let  $j$  be the index of the  $16 \times 16$  block which contains block  $i$ . Also, let  $\beta_{j,k}^d(z_j)$  be the quantized DC coefficient of the  $k$ -th chrominance component using the chrominance quantization table of class  $z_j$ , and  $\beta_{j,k}^a(z_j)$  be the vector of the quantized AC coefficients. Then, we assume that

$$R_i^c(x_i|x_{i-1}) = \frac{1}{4} \sum_{k=0}^1 \{r'_d [\beta_{j,k}^d(x_i) - \beta_{j-1,k}^d(x_{i-1})] + r'_a [\beta_{j,k}^a(x_i)]\},$$

where  $r'_d(\cdot)$  is the number of bits used for coding the difference between two consecutive DC coefficients of the chrominance components, and  $r'_a(\cdot)$  is the number of bits used for coding AC coefficients of the chrominance components. Notice that we split the bits used for coding the chrominance equally among the four corresponding  $8 \times 8$  blocks of the input document image.

The bits used for coding the mask are approximated by the entropy of a non-parametric conditional probability mass function. Assume that the number of bits for coding  $b_{i,m,n}$  only depends on its four causal neighbors, denoted as  $V_{i,m,n} = [b_{i,m-1,n-1}, b_{i,m-1,n}, b_{i,m-1,n+1}, b_{i,m,n-1}]^t$ . Define  $b_{i,m,n}$  to be 0, if  $m < 0$  or  $n < 0$  or  $m > 7$  or  $n > 7$ . Then, the number of bits required to code the binary mask is approximated as

$$- \sum_{m=0}^7 \sum_{n=0}^7 \log_2 p_b(b_{i,m,n}|V_{i,m,n}),$$

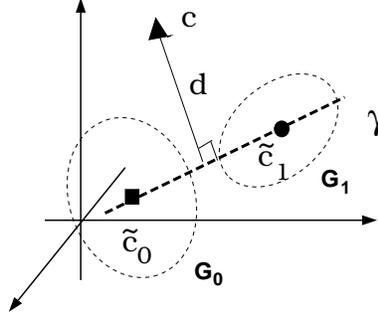
where  $p_b(b_{i,m,n}|V_{i,m,n})$  is the transition probability from the four causal neighbors to pixel  $(m, n)$  in block  $i$ .

### 3.1.1. Distortion

For the four classes (except Two-color blocks): One-color Background, One-color Foreground, Foreground and Background blocks, the total squared error in YCrCb color space is used as the distortion measure. The distortion is computed in the DCT domain, eliminating the need to compute inverse DCT's. Let  $e_i^l(x_i)$  be the quantization error of luminance DCT coefficients of block  $i$  using the luminance quantization table of  $x_i$ , and  $e_{j,k}^c(z_j)$  be the quantization error of DCT coefficients of the  $k$ -th chrominance component of the  $16 \times 16$  block containing block  $i$  using the chrominance quantization table of  $z_j$ . Then, the distortion is approximately given by

$$D_i(x_i) = \|e_i^l(x_i)\|^2 + \sum_{k=0}^1 \|e_{j,k}^c(x_i)\|^2.$$

Here, we approximate the distortion due to the chrominance channels by dividing the chrominance error among the four corresponding  $8 \times 8$  blocks of the luminance channel.



**Figure 4.** Two-color distortion measure.  $\tilde{c}_0$  and  $\tilde{c}_1$  are indexed mean colors of group  $G_0$  and  $G_1$ , respectively.  $\gamma$  is the line determined by  $\tilde{c}_0$  and  $\tilde{c}_1$ . The distance between a color  $c$  and  $\gamma$  is  $d$ . When  $c$  is a combination of  $\tilde{c}_0$  and  $\tilde{c}_1$ ,  $d = 0$ .

However, the distortion measure for Two-color blocks is different from the other four classes. Its distortion measure is designed with the following considerations. In a scanned image, pixels on the boundary of two color regions tend to have a color which is a combination of the colors of both regions. Since only two colors are used for the block, the boundaries between the color regions are usually sharpened. Although the sharpening generally improves the quality, it gives a large difference in pixel values between the original and the reconstructed images on boundary points. On the other hand, if a block is not a Two-color block, a third color often appears on the boundary. Therefore, a desired distortion measure for Two-color coder should not excessively penalize the error caused by sharpening, but should produce a high distortion value, if more than two colors exist. Also, desirable Two-color blocks should have a certain proportion of internal points. If a Two-color block has very few internal points, the block usually comes from background or halftone background, and it can not be a Two-color block. To handle this case, we set the cost to the maximal cost, if the number of internal points is less than or equals to 8. The distortion measure for the Two-color block is defined as follows. Define  $I_{i,m,n}$  as an indicator function.  $I_{i,m,n} = 1$ , if  $(m,n)$  is an internal point.  $I_{i,m,n} = 0$ , if  $(m,n)$  is a boundary point. If  $x_i = Two$ ,

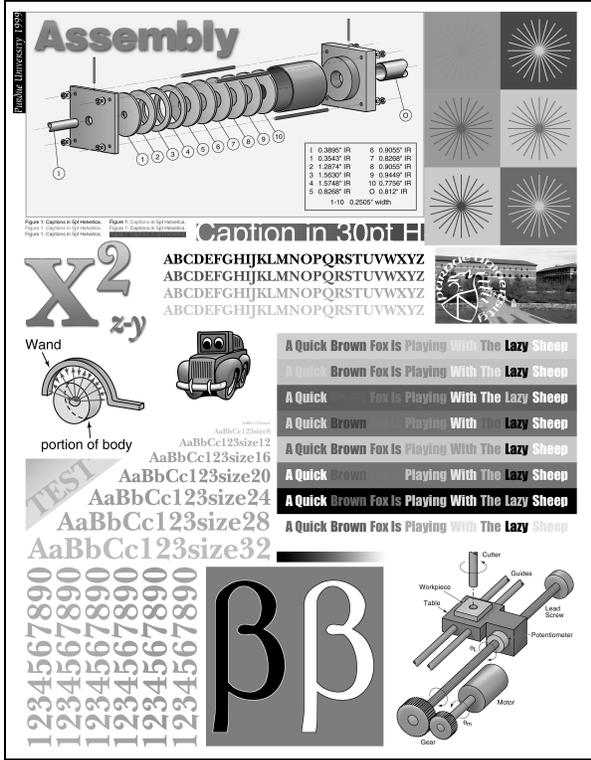
$$D_i(x_i) = \begin{cases} \sum_{m=0}^7 \sum_{n=0}^7 [I_{i,m,n} \|y_{i,m,n} - \tilde{c}_{i,b_{i,m,n}}\|^2 + (1 - I_{i,m,n}) d^2(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1})], & \text{if } \sum_{j=0}^1 |\tilde{G}_{i,j}| > 8 \\ 255^2 \times 64 \times 3, & \text{if } \sum_{j=0}^1 |\tilde{G}_{i,j}| \leq 8 \end{cases}$$

where  $|\tilde{G}_{i,j}|$  is the number of elements in the set  $\tilde{G}_{i,j}$ , and  $d(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1})$  is the distance between  $y_{i,m,n}$  and the line determined by  $\tilde{c}_{i,0}$  and  $\tilde{c}_{i,1}$ . As illustrated in Fig. 4, if a color  $c$  is a combination of  $c_1$  and  $c_2$ ,  $c$  will be on the line determined by  $c_1$  and  $c_2$ ,  $d(c; c_1, c_2) = 0$ . Therefore, for boundary points of Two-color blocks,  $d(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1})$  is small. However, if a third color does exist on a boundary point,  $d(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1})$  tends to be large.

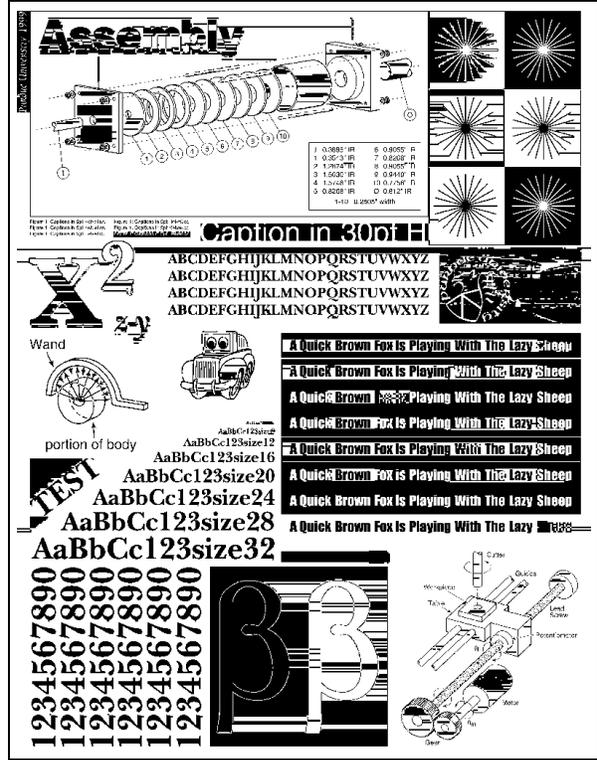
#### 4. EXPERIMENTAL RESULTS

For our experiments, we use an image database consisting of 30 scanned and one synthetic document image. The scanned documents come from a variety of sources, including ASEE Prism and IEEE Spectrum. These documents are scanned at 400 dpi and 24 bits per pixel (bpp) using the HP flat-bed scanner, scanjet 6100C. A large portion of the 30 scanned images contain halftone background and have ghosting artifacts caused by printing on the reverse side of the page. These images are used without pre-processing. The synthetic image shown in Fig. 5(a) \* has a complex layout structure and many colors. It is used to test the ability of a compression algorithm to handle complex document images.

\*To obtain color version of the experimental results, please visit <http://dynamo.ecn.purdue.edu/bouman/publications> or visit <http://min.ecn.purdue.edu/hui>.



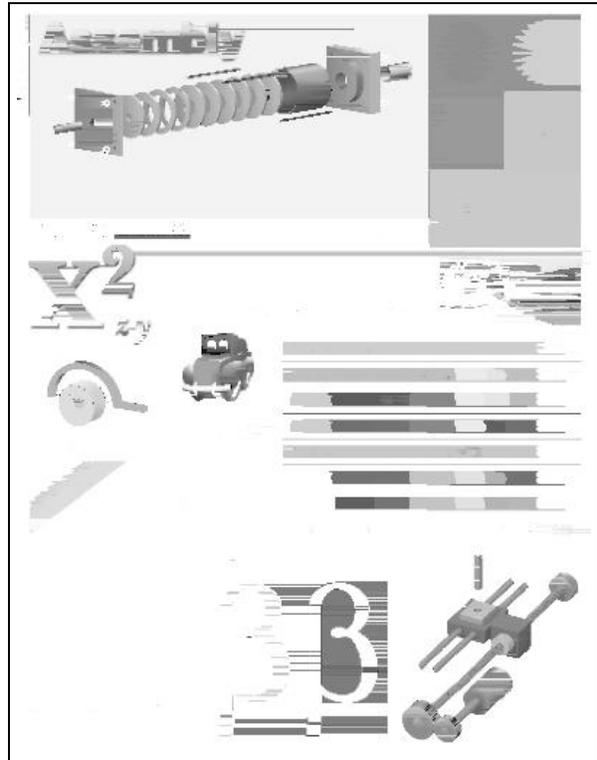
(a)



(b)

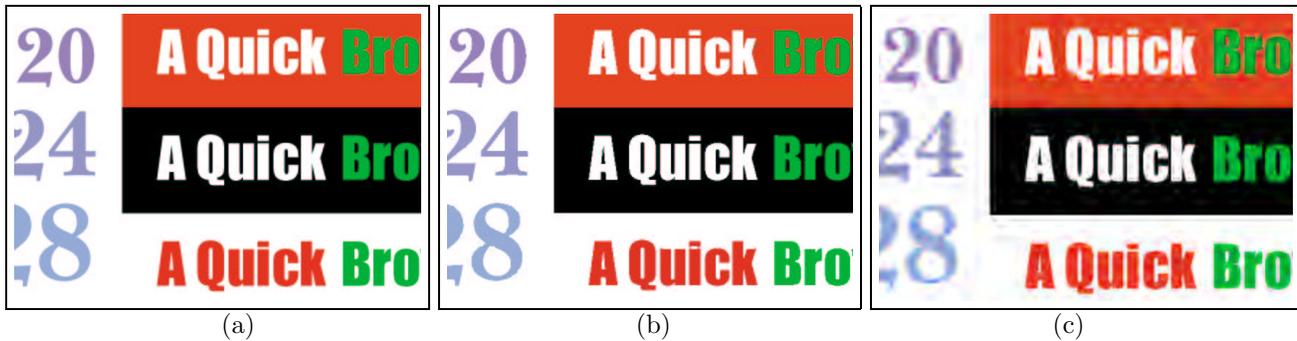


(c)

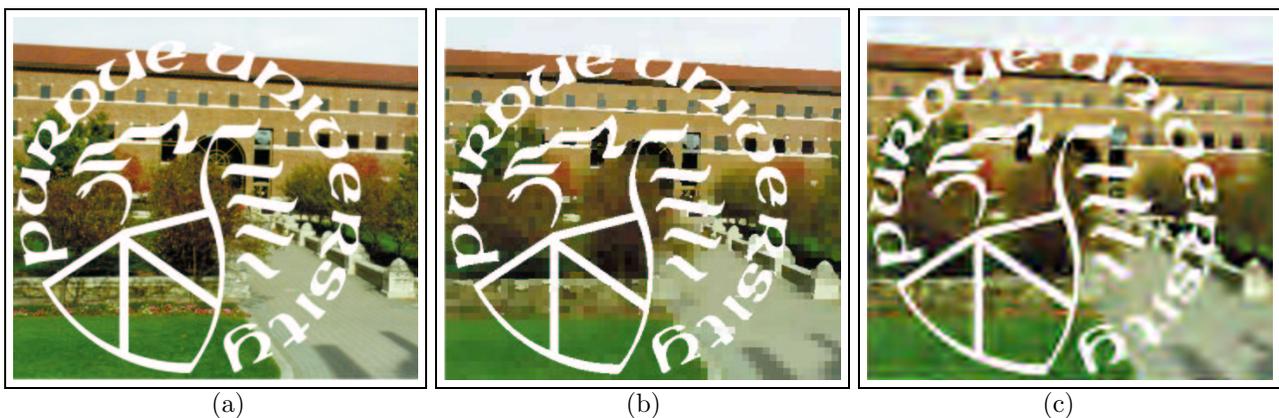


(d)

**Figure 5.** 3-layer MRC representation using rate-distortion based segmentation. (a) Synthetic test image. (b) Mask layer of the 3-layer MRC representation. (c) Foreground layer of the 3-layer MRC representation. (d) Background layer of the 3-layer MRC representation.



**Figure 6.** Compression result I. (a) A portion of the synthetic test image. (b) MRC with rate-distortion segmentation compressed at 0.1085 bpp (221:1 compression), where  $\lambda = 0.1$ . (c) JPEG-2000 compressed at 0.1091 bpp (220:1 compression).



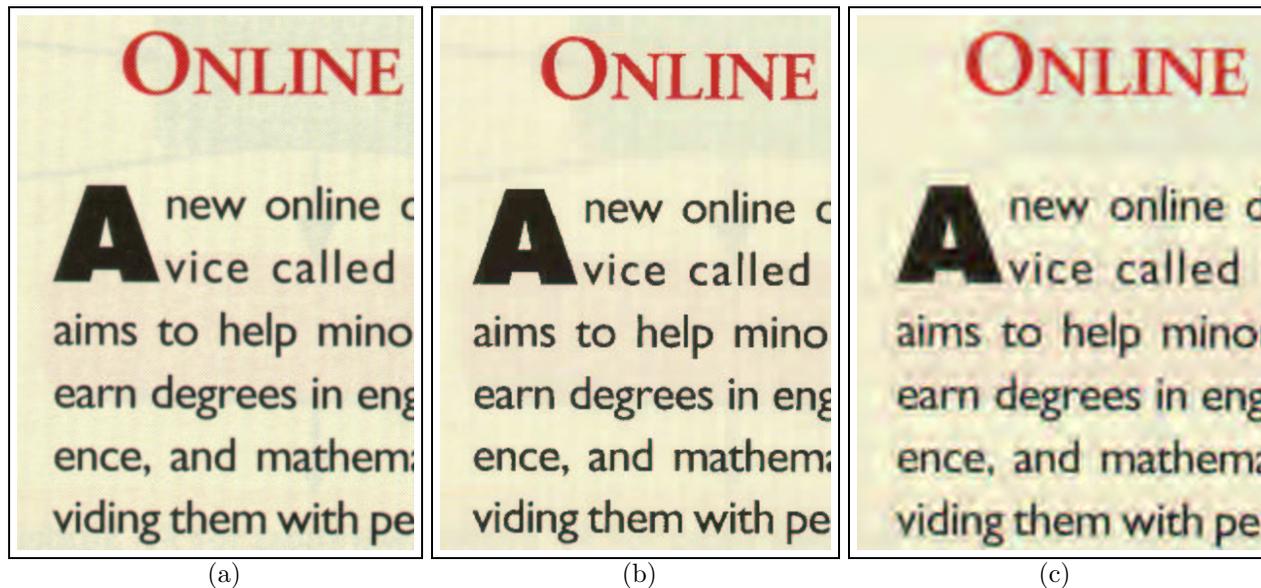
**Figure 7.** Compression result II. (a) A portion of the synthetic test image. (b) MRC with rate-distortion segmentation compressed at 0.1085 bpp (221:1 compression), where  $\lambda = 0.1$ . (c) JPEG-2000 compressed at 0.1091 bpp (220:1 compression).

Fig. 5 shows the experimental result using the synthetic document image. The original image is shown in grayscale in Fig. 5(a). Fig. 5(b), Fig. 5(c) and Fig. 5(d) are the mask layer, the foreground layer and the background layer, respectively, resulting from the proposed MRC compression with the rate-distortion based compression. The image is compressed at 0.1085 bpp (bit per pixel), which achieves a 221:1 compression.

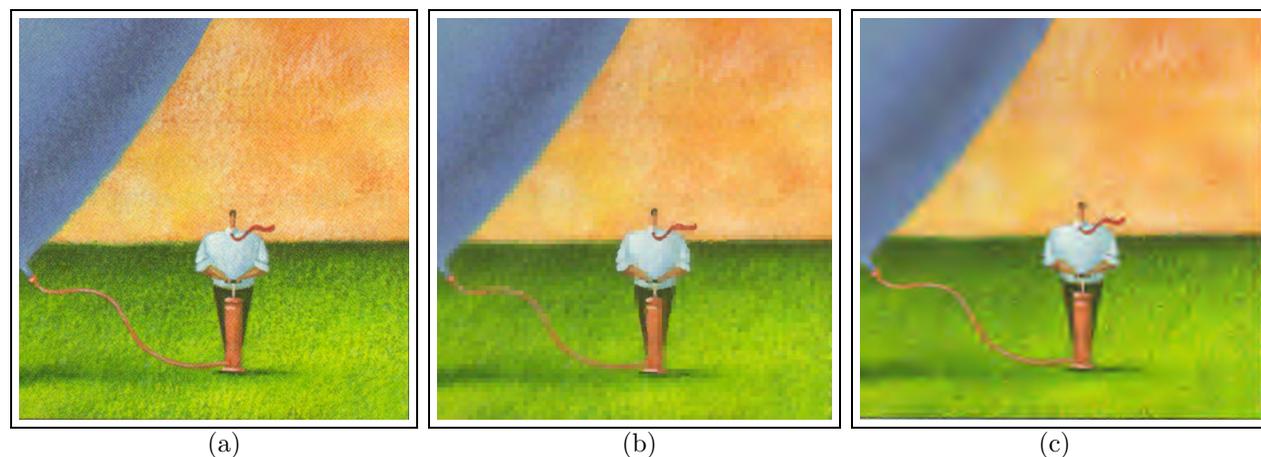
In Fig. 6-Fig. 9, we compare the quality of reconstructed images compressed using the proposed algorithm with the ones compressed using JPEG-2000 verification model 8.6 at similar bit rate. Fig. 6-Fig. 7 are regions from the synthetic test image. Fig. 8-Fig. 9 are regions from a scanned document image. From all four figures, we can see that the proposed algorithm achieves much higher quality than JPEG-2000 at the similar bit rate.

## 5. CONCLUSION

In this paper, we propose a spatially adaptive compression algorithm for document images using the 3-layer MRC model, and a rate-distortion based segmentation algorithm. This algorithm first segments a scanned document image into different classes. Then, each class is represented using 3-layer MRC model differently according to the property of that class. The segmentation is performed by optimizing a rate-distortion performance over the entire image with respect to a rate-distortion trade-off selected by an user. Since each block is tested on all classes, the rate-distortion based segmentation can eliminate severe misclassifications, such as misclassifying a Two-color block as a One-color block. Experimental results show that at similar bit rates, our algorithm can achieve a higher subjective quality than well-known coders such as JPEG-2000.



**Figure 8.** Compression result III. (a) A portion of the original test image I. (b) MRC with rate-distortion segmentation compressed at 0.1807 bpp (133:1 compression), where  $\lambda = 0.1$ . (c) JPEG-2000 compressed at 0.1645 bpp (146:1 compression).



**Figure 9.** Compression result IV. (a) A portion of the original test image I. (b) MRC with rate-distortion segmentation compressed at 0.1807 bpp (133:1 compression), where  $\lambda = 0.1$ . (c) JPEG-2000 compressed at 0.1645 bpp (146:1 compression).

## ACKNOWLEDGMENTS

This work was conducted when Hui Cheng was with Digital Imaging Technology Center, Xerox Corporation. We thank Xerox Foundation for their support of this research. We also thank Dr. Faouzi Kossentini and Mr. Dave Tompkins of Department of Electrical and Computer Engineering, University of British Columbia for providing us the JBIG2 coder. In addition, we thank ASEE, ASEE Prism, IEEE, IEEE Spectrum, and Stanley Electric Sales of America for allowing us to use documents published on ASEE Prism and IEEE Spectrum in this research.

## REFERENCES

1. S. J. Harrington and R. V. Klassen, "Method of encoding an image at full resolution for storing in a reduced image buffer," *US Patent 5,682,249*, October 1997.
2. K. Konstantinides and D. Tretter, "A method for variable quantization in JPEG for improved text quality in compound documents," in *Proc. of IEEE Int'l Conf. on Image Proc.*, **2**, pp. 565–568, (Chicago, IL), October 4-7 1998.
3. M. Ramos and R. L. de Queiroz, "Adaptive rate-distortion-based thresholding: application in JPEG compression of mixed images for printing," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Kobe, Japan), October 25-28 1999.
4. H. Cheng and C. A. Bouman, "Multilayer document compression algorithm," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Kobe, Japan), October 25-28 1999.
5. L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. LeCun, "High quality document image compression with 'DjVu'," *Journal of Electronic Imaging* **7**, pp. 410–425, July 1998.
6. J. Huang, Y. Wang, and E. K. Wong, "Check image compression using a layered coding method," *Journal of Electronic Imaging* **7**, pp. 426–442, July 1998.
7. R. L. de Queiroz, R. Buckley, and M. Xu, "Mixed raster content (MRC) model for compound image compression," in *Proc. of SPIE Conf. on Visual Communications and Image Processing*, **3653**, pp. 1106–1117, (San Jose, CA), February 1999.
8. K. Murata, "Image data compression and expansion apparatus, and image area discrimination processing apparatus therefor," *US Patent 5,535,013*, July 1996.
9. H. Cheng and C. A. Bouman, "Multiscale bayesian segmentation using a trainable context model," *IEEE Trans. on Image Processing* **10**, pp. 511–525, April 2001.
10. H. Cheng and C. A. Bouman, "Document compression using rate-distortion optimized segmentation," *Journal of Electronic Imaging* **10**, pp. 460–474, April 2001.
11. A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Proc. Magazine* **15**, pp. 23–50, November 1998.
12. P. G. Howard, F. Kossentini, B. Martins, S. Forchhammer, and W. J. Rucklidge, "The emerging JBIG2 standard," *IEEE Trans. on Circuits and Systems for Video Technology* **8**, pp. 838–848, November 1998.
13. G. M. Schuster and A. K. Katsaggelos, *Rate-distortion based video compression*, Kluwer Academic Publishers, Boston, 1997.
14. W. B. Pennebaker and J. L. Mitchell, *JPEG: still image data compression standard*, Van Nostrand Reinhold, New York, 1993.