

RATE DISTORTION OPTIMIZED DOCUMENT CODING USING RESOLUTION ENHANCED RENDERING

Guotong Feng[†], Hui Cheng[‡], and Charles A. Bouman[†]

[†]Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907-1285
{fengg, bouman}@ecn.purdue.edu

[‡]Visual Information Systems
Sarnoff Corporation
Princeton, NJ 08543-5300
hcheng@sarnoff.com

ABSTRACT

Raster document coders are typically based on the use of a binary mask layer that efficiently encodes the text and graphic content. While these methods can yield much higher compression ratios than natural image compression methods, the binary representation tends to distort fine document details, such as thin lines, and text edges.

In this paper, we describe a method for encoding and decoding the binary mask layer that substantially improves the decoded document quality at a fixed bit rate. This method, which we call resolution enhanced rendering (RER), works by adaptively dithering the encoded binary mask, and then applying a nonlinear predictor to decode a gray level mask at the same or higher resolution. We present experimental results illustrating that the RER method can substantially improve document quality at high compression ratios.

1. INTRODUCTION

Document imaging applications such as scan-to-print, document archiving, and internet fax are driving the need for document compression standards that maintain high quality while achieving high compression ratios. Recently, a number of methods have been proposed for encoding the high resolution text and graphics components of a document using a binary mask that selects between foreground and a background content. Methods such as mix raster content (MRC) [1], DjVu [2], and RDOS [3, 4] use this technique to achieve high document quality at much lower bit rates than are achievable using more conventional natural image coders. However, a disadvantage of such coders is that the binary mask representation can substantially distort fine document details, such as thin lines and text edges.

In this paper, we describe a new method called resolution enhanced rendering (RER) for jointly optimizing the

MRC encoder and decoder to achieve high quality rendering, while maintaining desired compression ratios. The method works by adaptively dithering the binary mask layer to produce intermediate tone levels that can better represent fine document detail. This dithering is performed using a novel adaptive error diffusion algorithm. A tree-based nonlinear predictor is designed into the MRC decoder to reconstruct a gray level mask at the same or higher resolution. This gray level mask is then used to mix the foreground and background content. Both the dithering and nonlinear prediction algorithms are jointly optimized to produce the best quality rendering. The optimization is performed by iteratively optimizing the encoder and decoder to achieve the minimum distortion.

We show how RER can also be used to interpolate text and graphics to a higher spatial resolution, while retaining the favorable bit rate of a lower resolution encoding. This can be done by interpolating the binary mask during the decoding process. This interpolative RER method (IRER) can be used to compress high resolution documents at very low bit rates. This is because IRER compresses much of the document's content at lower spatial resolution. Only the text and graphic content is rendered at the full resolution necessary to preserve its quality.

The RER method has a number of additional advantages. First, it is quite general since it can be applied to any document compression method which encodes text and graphics with a binary mask. Second, it is compatible with standards used for document encoding, such as MRC, because it encodes edge gradations into the spatial attributes of the binary mask.

2. CONVENTIONAL MRC ENCODERS

Typical document encoding algorithms, such as MRC, represent a document using different layers. As shown in Fig. 1, a 3-layer MRC document contains a background layer, a

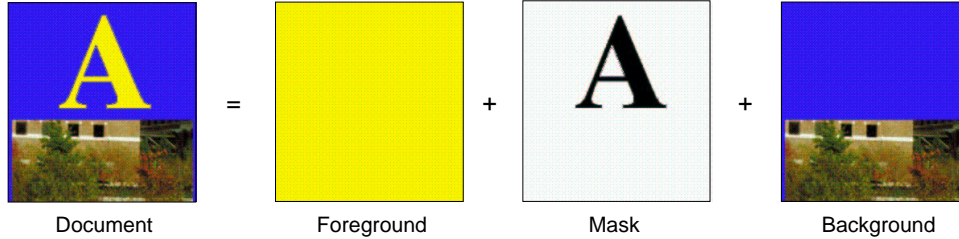


Fig. 1. Typical document coders form text and line art by using a binary mask to choose between foreground and background layers.

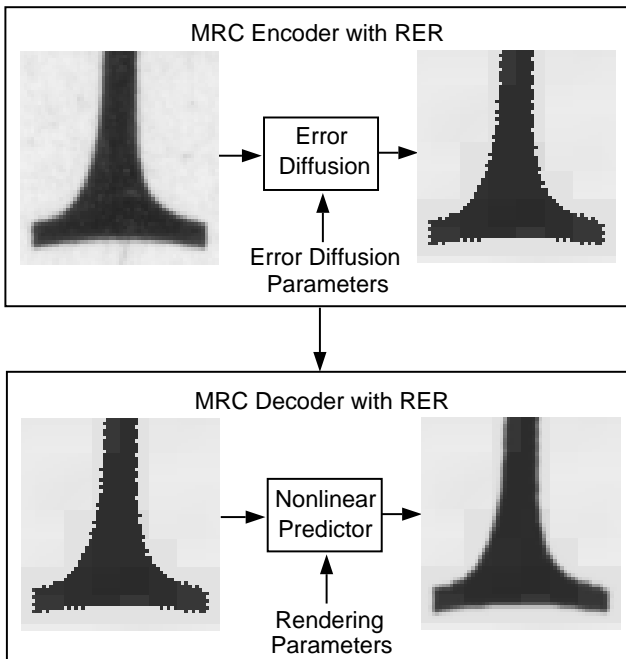


Fig. 2. MRC encoder and decoder with RER.

foreground layer, and a binary mask layer. At each pixel, the value of the binary mask is used to select between the foreground and background pixels. Typically, the foreground and background layers are compressed using natural image coders such as a JPEG; whereas the binary mask is encoded with a lossless binary encoder such as JBIG or JBIG2.

In this work, we use the rate distortion optimized segmentation (RDOS) algorithm proposed in [3] as our baseline document encoder. The RDOS algorithm classifies each 8×8 block of pixels into one of four classes: “picture block”, “two-color block”, “one-color block” and “other block”. Each class corresponds to a different coding method as described in [3]. Importantly, each two-color block is further segmented into foreground and background pixels. This high resolution segmentation forms the binary mask that is encoded using JBIG2.

3. RESOLUTION ENHANCED RENDERING

Fig. 2 illustrates how the resolution enhanced rendering (RER) algorithm adds edge detail while retaining the binary mask layer used in typical document coders. First, the RER encoder segments the foreground and background using an adaptive error diffusion method. This error diffusion method dithers the binary mask along the edge of the character to represent the gradual transitions of true raster scanned text characters. The error diffusion algorithm uses the local value of the mask to adapt the error diffusion weights so that error is diffused along the 1-D mask boundary.

The RER decoder uses the binary mask, together with the foreground and background colors to estimate the true pixel values. This estimation is done using a nonlinear tree-structured predictor as described in [5]. Importantly, this predictor is trained to identify the characteristic patterns of the RER encoder. This substantially improves estimation accuracy.

The RER encoder and decoder are jointly optimized to maximize the quality of the decoded document. In each iteration of this optimization, the parameters of the encoder or decoder are alternatively fixed, while the parameters of the other one are optimized. We found that joint optimization of the encoder and decoder performs substantially better than independent optimization of these two functions. Importantly, two different sets of documents are used for training the encoder and decoder. We have found this improves the robustness of the training procedure.

3.1. The RER encoder

Let X_s be a pixel in the raster document at location s . Assume that each pixel also has an associated foreground color, F_s , and background color, B_s . The binary MRC mask then determines whether F_s or B_s will be used to represent the true pixel value X_s . Next define the scalar value λ_s which determines the relative mixture of foreground and background color in the pixel X_s . More specifically, λ_s is given by the value in the interval $[0, 1]$ which minimizes the squared error

$$\|X_s - (B_s + \lambda_s(F_s - B_s))\|^2 \quad (1)$$

Notice that λ_s forms a gray scale image with minimum value 0 and maximum value 1. The approximation error at each pixel is then given by

$$\Delta_s = \min\{\lambda_s, 1 - \lambda_s\} \quad (2)$$

The RER Encoder computes the binary mask by applying an adaptive version of serpentine scan Floyd Steinberg error diffusion. The modified method works by setting the error diffusion weights at each pixel according to the formula

$$w_{s_j} = \frac{\alpha_j \Delta_{s_j} u(\Delta_{s_j} - \tau)}{\sum_{j=0}^3 \alpha_j \Delta_{s_j} + 0.001} \quad (3)$$

where s_j is the j^{th} neighbor of the pixel in the error diffusion algorithm, and α_j are fixed weights that parameterize the algorithm. Notice that this method is specially designed to diffuse errors along text edge transitions.

3.2. The RER decoder

The RER decoder works by using a nonlinear predictor to compute, $\hat{\lambda}_s$, the minimum mean squared error estimate of λ_s . Using this estimate, the reconstructed pixel color is

$$\hat{X}_s = \hat{\lambda}_s F_s + (1 - \hat{\lambda}_s) B_s. \quad (4)$$

Here we assume that the foreground and background colors are the same as used in the RER encoder.

The RER algorithm can be used with or without interpolation of the binary mask. When interpolation is used, the estimated gray scale mask $\hat{\lambda}_s$ has twice the sample resolution of the original gray scale mask λ_s . This is done by designing the nonlinear predictor to estimate four output values for each input pixel location.

The nonlinear predictor works by first extracting the binary mask in a window about the pixel in question. This data forms a binary vector, z_s , which is then used as input to a binary regression tree predictor known as Tree-Based Resolution Synthesis (TBRS) [5, 6]. The TBRS predictor estimates the value of λ_s in a two-step process. First, it classifies the vector z_s into one of M classes using a binary tree classifier. Each class, then has a corresponding linear prediction filter which is used to estimate the value of λ_s from z_s using the equation

$$\hat{\lambda}_s = A_m z_s + b_m$$

where m is the class of the vector z_s , A_m and b_m are the corresponding linear prediction parameters of class m .

The basic idea of TBRS is to use a binary regression tree as a piecewise linear approximation to the conditional mean estimator. This classification step is essential because it can separate out regions of the document corresponding to mask edges of different orientation and shape.

3.3. Training RER

The objective of the training process is to optimize the performance of the RER encoder and decoder by selecting the encoder and decoder parameters to maximize the decoded document quality over a training set of documents. The distortion metric used to measure document quality is mean squared error.

The training process alternated between optimization of the encoder and decoder parameters [4]. So, when optimizing the encoder parameters, the previously obtained decoder parameters were used; and when optimizing the decoder parameters, the previously obtained encoder parameters were used. The training phases for encoder and decoder used different sets of training data. This strategy seemed to produce more robust training results.

4. EXPERIMENTAL RESULTS

In our experiments, we used an encoder training set consisting of a single documents and a decoder training set of 8 full documents. All results were obtained from testing data not contained in either of the training sets. All images were scanned at 600 dpi and 24 bits per pixel on the HP 6100C flatbed scanner. The test page used in these results is available from www.ece.purdue.edu/~bouman.

We tested the performance of both RER and interpolative RER (IRER). For IRER, the decoder produces an output document with twice the resolution of the input document as described in Section 3.2. All our experiments used a 300dpi input document formed by decimation of the 600dpi original document. Decimation was performed using block averaging.

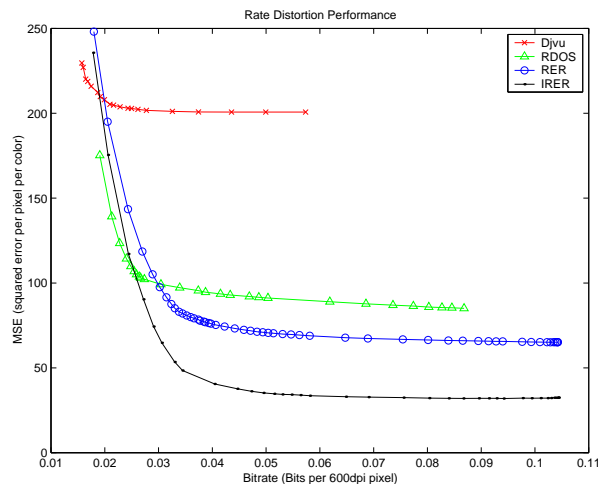


Fig. 4. Rate-distortion performance of RDOS, DjVu, RER, and IRER on a document decoded at 600dpi.

Fig. 3 shows the comparison among the original 600dpi



Fig. 3. Comparison of compression results. (a) A portion of the original 600dpi test image. Results of compression and decompression of a 300dpi original using (b) standard RDOS at 0.2014 bpp; (c) resolution enhanced rendering (RER) at 0.2016bpp; (d) resolution interpolated RER (IRER) at 0.2000 bpp.

image, the image rendered by standard RDOS, the image rendered by the RER enhanced RDOS encoder/decoder pair, and the image rendered by IRER. Notice that the IRER result is a much more accurate rendition of the original 600dpi document. Fig. 3b exhibits objectionable 300dpi “jaggie” artifacts around the text edges. Fig. 3c shows that the RER method eliminates these artifacts but is still limited to 300dpi output. Alternatively, Fig. 3d produces an relatively accurate rendition of the original 600dpi character.

Fig. 4 illustrates the comparison of rate-distortion performance between DjVu, standard RDOS, and RER enhanced RDOS. Each method uses the 300dpi document as input. The IRER method produces a 600dpi output, but DjVu, RDOS and RER produce 300dpi output which are then interpolated to 600dpi using pixel replication. The distortion of each 600dpi decoded document is measured using squared error per pixel per color; and the bit rate is measured in bits per 600dpi output pixel.

The two rate-distortion curves show the substantial improvement achieved by the RER and IRER methods. In particular, the IRER method produces good quality encodings of the 600dpi document, but with very high compression ratios on the order of 500:1. These high compression ratios result from the fact that much of the document (corresponding to background and image content) is actually rendered at 300dpi. Only the text and graphic content is rendered at the 600dpi resolution necessary to preserve its quality.

5. CONCLUSIONS

In this paper, we have proposed the resolution enhanced rendering (RER) method for document encoding and decoding. Our experimental results indicate that the RER and the interpolative RER (IRER) methods can produce good quality 600dpi documents at compression ratios exceeding 500:1.

6. REFERENCES

- [1] 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*, San Jose, CA, Februray 1999, vol. 3653, pp. 1106–1117.
- [2] 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*, vol. 7, no. 3, pp. 410–425, July 1998.
- [3] H. Cheng and C. A. Bouman, “Document compression using rate-distortion optimized segmentation,” *Journal of Electronic Imaging*, vol. 10, no. 2, pp. 460–474, April 2001.
- [4] G. Feng, H. Cheng, and C. Bouman, “High quality mrc document coding,” in *IS&T Conf. on Image Proc., Image Quality and Image Capture Sys.(PICS)*, Montreal Canada, April 21-25 2001, pp. 320–325.
- [5] C. B. Atkins, C. A. Bouman, and J. P. Allebach, “Tree-based resolution synthesis,” in *IS&T Conf. on Image Proc., Image Quality and Image Capture Sys.(PICS)*, Savannah, GA, April 25-28 1999, pp. 405–410.
- [6] C. B. Atkins, *Classification Based Methods in Optimal Image Interpolation*, Ph.D. dissertation, Purdue University, West Lafayette, IN, December 1998.