IMPROVED LAYER PROCESSING FOR MRC COMPRESSION OF SCANNED DOCUMENTS

Alexandre Zaghetto and Ricardo L. de Queiroz

Universidade de Brasilia, Brazil
alexandre@cic.unb.br, queiroz@ieee.org

ABSTRACT

The Mixed Raster Content (MRC) document compression is a well documented standard. Its efficiency for representing sharp text and graphics over a background has been extensively presented. Scanned documents, however, are difficult to be dealt with because of soft transitions. In one of our previous works we presented a pre/post-processing algorithm for MRC compression of scanned images that sharpens the edges, before compression, and softens it again, after reconstruction. The present paper uses the same basic idea but describes three new features that improve, in a rate-distortion sense, the compression performance. First, the real bitrate is used to determine the pre/post-processing parameters. Second, the Mask layer and the edge sharpening/softening map have their quality improved and, third, layer resolution change has been explored. Experimental results show that the method not only provides higher subjective quality, but it can yield up to 2.5 dB gains in PSNR against single coder approaches, in the compression ranges of interest.

Index Terms— Document compression, mixed raster content, document model, edge processing.

1. INTRODUCTION

The mixed raster content (MRC) imaging model [1]-[4] has been proposed as a multi-layer representation of a compound document. The basic 3-layer MRC model represents a color image as two color image layers (Foreground or FG and Background or BG) and a binary image layer (Mask or M). The Mask layer describes how to reconstruct the final image from the FG/BG layers, i.e. to use the corresponding pixel from the FG or BG layers when the Mask pixel is 0 or 1, respectively, in that position. An illustration of the imaging model is shown in Fig. 1. Once the original single-resolution image is decomposed into layers, each layer can be processed and compressed using different algorithms. Foreground and Background processing operations may include resolution change and data-filling procedure. The compressed layers are then packaged for delivery to the decoder.

In MRC, typically, the imaged text has about the same shape and size as that in the Mask plane. When the image is
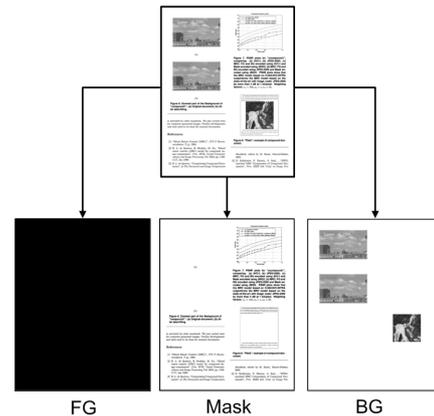
Fig. 1. Illustration of the MRC imaging model. The basic 3-layer MRC model represents a color image as two color image layers (Foreground or FG and Background or BG) and a binary image layer (Mask or M).

scanned, the edges of text and graphics are not as sharp and we refer to them as "soft edges". Since the selector plane is binary and the edge transitions are smooth, it is not possible to contain all the Foreground in one plane and all the Background in another. The effect in images is that of a "halo" around the text in the FG/BG planes, as illustrated in Fig. 3 (a) and (c), which is very damaging to compression.

This paper proposes an improved pre- and post-processing algorithm that sharpens the document before compression by removing the described halo effect and softens it again, after reconstruction, using a Gaussian filter. Regarding compression, we propose a 3-layer MRC codec for compound documents that uses the H.264/AVC [5], [6] operating in INTRA mode (AVC-I) to encode Foreground and Background layers and JBIG2 [7] to encode the binary Mask [8]. Even though the basic scheme used here has been presented in one of our previous works [9], there are three new features that were included in the present method:

- The binary Mask and the halo processing map C are post-processed, improving the rate-distortion efficiency of the CODEC;

- Instead of estimating the number of bits needed to en-

code the document, the real rate R is used to determine more appropriate pre/post-processing parameters, ϵ, h and σ [9].

- Foreground/Background resolution change was included, increasing the compression range of interest.

Since layer decomposition is not the main focus of this paper, we used Fan's text segmentation algorithm [10] in our tests, for convenience.

2. EDGE SHARPENING AND SOFTENING

In order to remove the halo effect we are forced to change the data itself. The first step is to estimate where the halo will possibly occur. Our approach is to find transitions by applying the Sobel operator to the Mask. The resulting transitions are morphologically dilated by a $d \times d - pixel$ structured element in order to mark a neighborhood. The image pixels that coincide with the dilated Mask transitions are marked as possible processing targets. Let E be the set of pixel locations comprising this region.

The next step is to find pixels which are supposed to cause the halo effect. Let F and B represent the pixel positions where the Mask indicates FG or BG, respectively. We compute averages as:

$$
\begin{aligned}
m_{FG} &= mean(x(i,j)|(i,j) \in F) \\
m_{BG} &= mean(x(i,j)|(i,j) \in B),
\end{aligned}
\tag{1}
$$

where $x(i,j)$ represents an original image pixel.

We mark any pixel in the candidate region whose gray level is far apart from its layer average, i.e.:

$$
C_{FG} = \left\{ \begin{array}{lll} 0 & : & |x(i,j) - m_{FG}| > \epsilon \,|\, (i,j) \in (F \cap E) \\ 1 & : & otherwise \end{array} \right. ,
$$

$$
C_{BG} = \left\{ \begin{array}{lll} 0 & : & |x(i,j) - m_{BG}| > \epsilon \,|\, (i,j) \in (B \cap E) \\ 1 & : & otherwise \end{array} \right. .
\tag{2}
$$

where ϵ is a tolerance value.

Next, we find the pixels marked by C_{BG} whose values are less than $(m_{FG} + \epsilon)$ and transfer them to the FG layer. This means that pixels in the candidate region $(B \cap E)$ that are distant from m_{BG} but are close enough to m_{FG} will not be pre/post-processed. Similarly, we find pixels marked by C_{FG} whose values are greater than $(m_{BG} - \epsilon)$ and transfer them to the BG layer, i.e., those in candidate region $(F \cap E)$ that are distant from m_{FG} but are close enough to m_{BG} will not be affected by pre/post-processing. The mask M, as well as the maps C_{FG} and C_{BG} are updated to accommodate the inter-layer pixel transfer.

For the image in Fig. 2 (a), and for $\epsilon = 16$ (out of 256 gray levels), the mask M, the candidate region for processing E and the map of the pixels to be changed, i.e $C = C_{FG} \cup$
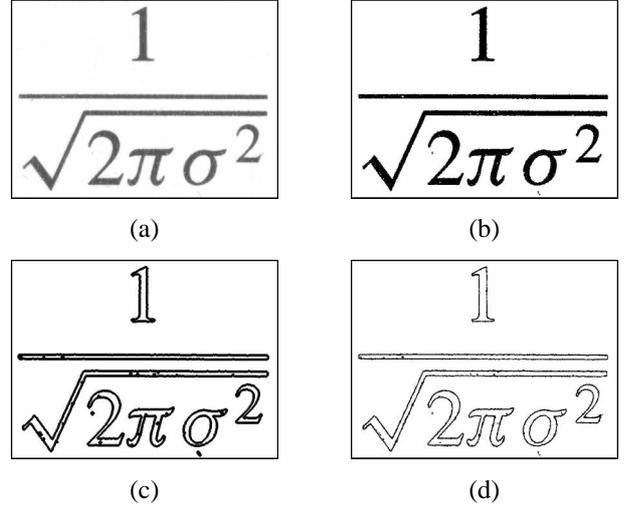

Fig. 2. (a) Original scanned material; (b) mask M; (c) candidate region for processing E; (d) pixels to be changed C.

C_{BG}, are shown in Figs. 2 (b), (c) and (d), respectively. In order to clean up the edge spots, we replace the values of the pixels in C_{FG} by m_{FG} and the values of the pixels in C_{BG} by m_{BG}. This is equivalent to changing the original image itself in order to make transitions sharper. If we send the JBIG2 encoded C map as side information, we can blur only the pixels that belong to this map using an $h \times h$ Gaussian filter with standard deviation σ.

Once the image is segmented, there will be "don't care" regions in the BG and FG layers, i. e., pixels assigned to the BG are marked as "don't care" on the FG, and vice-versa. These pixels can be replaced by anything to enhance compression. The problem of data-filling over the redundant data has been studied [1], [11], [12]. This paper uses an iterative wavelet-based plane filling [13]. Figure 3 shows FG/BG planes before and after halo processing and data-filling procedures. Note how the pre-processing improved the quality of the FG/BG planes.

3. ESTIMATION OF PRE- AND POST-PROCESSING PARAMETERS

Since the edges are sharpened to accommodate the Mask, in order to reconstruct soft edges, we have to somehow estimate the transition of the image edges. The quest is to determine the best values of parameters ϵ, h and σ in a rate-distortion sense. For this, we determine the solution by minimizing the following cost function:

$$
J(\epsilon, h, \sigma) = D + \lambda R,
\tag{3}
$$

where λ is a weighting factor, D is the distortion incurred by the pre-processing, MRC encoding/decoding and post-

(a)   (b)
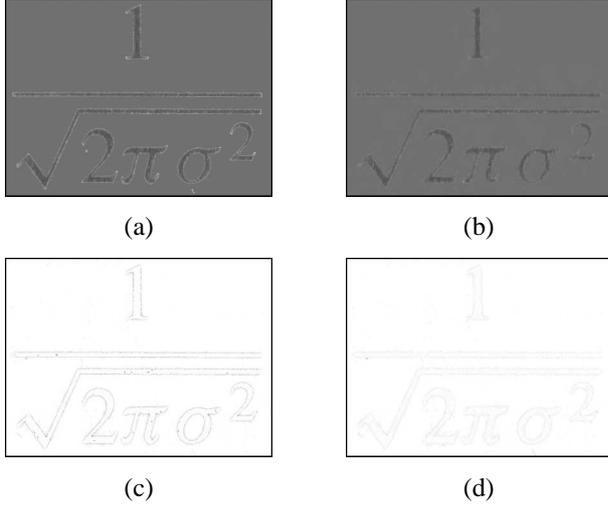
(c)   (d)

Fig. 3. Original (a) FG and (c) BG; processed (b) FG and (d) BG. Note the halo around the unprocessed FG/BG text. Pre-processing improves the quality of FG/BG planes.

processing algorithms, and R is the bitrate for compressing the document layers.

The algorithm that determines the best values for pre-processing parameters, ϵ, h and σ, is described as follows:

ALGORITHM 1

1   $h \leftarrow h_0$;
2   $\sigma \leftarrow \sigma_0$;
3   **for** $\epsilon \leftarrow \epsilon_0$ **to** ϵ_k
4   **do** Generate map C using ϵ;
5   Sharpen the edges using C;
6   Run data-filling algorithm;
7   MRC encode/decode FG, BG and M;
8   Encode/decode C;
9   Filter edges using a Gaussian filter
            with parameters (h_0, σ_0);
10   Calculate and store cost $J(\epsilon, h_0, \sigma_0)$;
11 Find ϵ that results in the minimum cost J and
    make it ϵ_{best};
12 Generate map C_{best} using ϵ_{best};
13 Sharpen the edges using C_{best};
14 Run data-filling algorithm;
15 MRC encode/decode FG, BG and M;
16 Encode/decode C_{best};
17 **for** $h \leftarrow h_0$ **to** h_i
18   **do for** $\sigma \leftarrow \sigma_0$ **to** σ_j
19   **do** Filter edges using a Gaussian filter
                with parameters (h, σ);
20   Calculate distortion D;
21 Find (h, σ) pair that minimizes D and make it
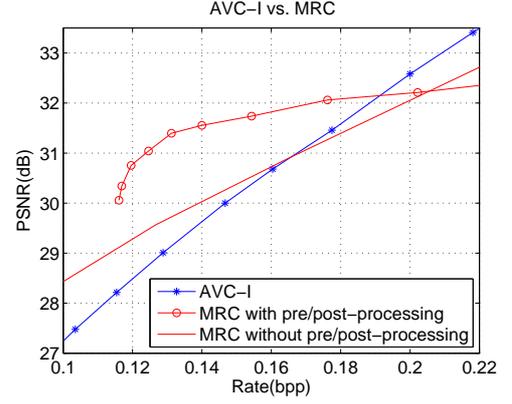 $(h_{best}, \sigma_{best})$;

Fig. 4. AVC-I and MRC performance for a scanned text/graphics image. The proposed MRC with pre/post-processing (ϵ_{best}=28, h_{best}=25, σ_{best}=1.5) outperforms MRC without pre/post-processing by more than 1.5 dB and AVC-I by more than 2 dB, at 0.13 bpp.

Since FG and BG are encoded using AVC-I, a design quantizer parameter, QP_D, needs to be set for the MRC encoder in steps 7 and 15. The H.264/AVC quantizer parameter, QP, may vary from 0 to 51. Since we are interested in very low bitrates, a high QP_D (above 30) is suggested.

MRC imaging model also allows resolution change of FG/BG layers. Resize factor, S, of 1, 1/2 and 1/4 were used. The performance of the codec was evaluated for those values, as described by the following algorithm:
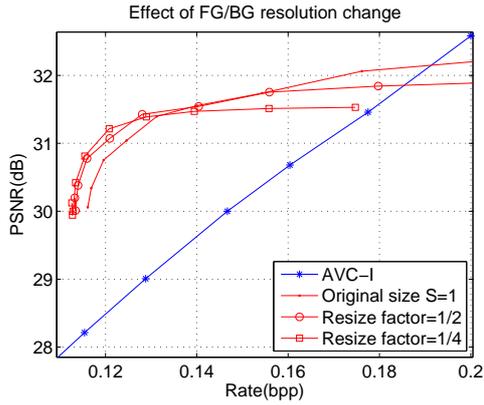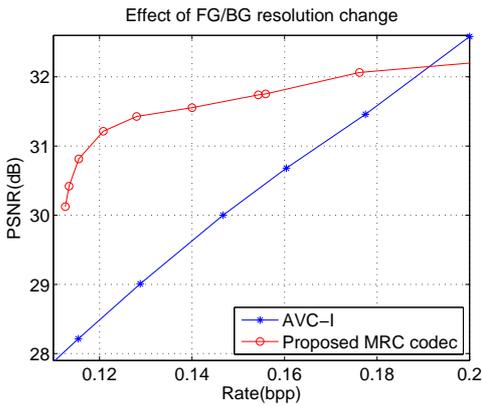
ALGORITHM 2

1   **for** $S \leftarrow 1, 1/2$ and $1/4$
2   **do for** $QP \leftarrow QP_0$ To QP_k
3   **do** Generate rate-distortion points (R, D);
4   Sort (R, D) points along R, in ascending order;
5   $N \leftarrow$ number of (R,D) points;
6   **for** $i \leftarrow 1$ To N
7   **do if** $D_i < D_{i-1}$
8   **then** Select (R_i, D_i) point;

4. RESULTS

Figure 4 shows a PSNR plot comparing AVC-I and MRC performance with and without the pre/post-processing steps for one typical document. No FG/BG resolution change was applied ($S = 1$). MRC with pre/post-processing outperforms MRC without pre/post-processing by more than 1.5 dB and AVC-I by more than 2.0 dB, at 0.13 bpp. However, there is only a short interval wherein there are gains using the proposed scheme (from 0.12 bpp to 0.18 bpp, when compared to AVC-I). This occurs because the method is bounded in PSNR due to the edge sharpening/softening procedure. Also, the achieved bitrate has a lower bound because of the number of bits needed to encode C and M losslessly.

Effect of FG/BG resolution change

(a)

Effect of FG/BG resolution change

(b)

Fig. 5. Effect of FG/BG resolution change: (a) MRC codec performance evaluation for resize factors S = 1, 1/2 and 1/4; (b) PSNR performance after running Algorithm 2 for the curves shown in Fig. 5 (a).

Layer downsampling procedure has been included as an effort to improve the compression range of interest. The rate-distortion effect of this procedure is shown in Fig. 5 (a). Figure 5 (b) shows the resulting rate-distortion performance after running Algorithm 2 for the curves shown in Fig. 5 (a). Note how the bitrate lower bounds are shifted left and the PSNR upper bounds are shifted upwards for bitrates closer to 0.12 bpp.

5. CONCLUSIONS

We have improved previously developed methods that counter balance the effects of soft edges in MRC compression of scanned documents. An MRC encoding and decoding scheme based on H.264/AVC-INTRA and JBIG2 was used and we have shown performance improvements, for low bitrates, over single coder approaches.

Although the proposed method is meant to deliver a re-constructed image which should be as similar as possible to the original scanned one, in some particular applications the post-processing procedure may be turned off. Subjectively, sharpened (pre-processed only) documents may present better quality than re-softened (post-processed) ones. Hence, the decoder might chose between softening or not the text. Furthermore, regular MRC decoders would ignore the C map and decode the sharper version.

The proposed approach improves the reconstruction fidelity in the MRC compression of scanned documents. In effect, our results have shown that the method enables competitive MRC compression of soft-edge document images. The complexity overhead incurred by the proposed pre/post-processing algorithm may constitute a disadvantage. However, since we are concerned with applications where the document is stored once and many users are to retrieve and decode the document at a later time, complexity has not been considered a hindrance..

6. REFERENCES

[1] R. L. de Queiroz, "Compressing Compound Documents", in *The Document and Image Compression Handbook*, edited by M. Barni, Marcel-Dekker, 2005.

[2] R. L. de Queiroz, R. Buckley and M. Xu, "Mixed raster content (MRC) model for compound image compression", *Proc. SPIE, Visual Communications and Image Processing*, Vol. 3653, pp. 1106-1117, Jan 1999.

[3] R. Buckley, D. Venable and L. McIntyre, "New developments in color facsimile and internet fax", *Proc. of IS&T's Fifth Color Imaging Conf.*, pp. 296-300, Scottsdale, AZ, Nov. 1997.

[4] "Mixed Raster Content (MRC)", *ITU-T Recommendation T.44*, 1999.

[5] "Final Draft International Standard of Joint Video Specification", *ITU-T Recommendation H.264, ISO/IEC 14496-10 AVC*, March 2003.

[6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, Wiley, 2003.

[7] "Information technology - Coded representation of picture and audio information - Lossy/lossless coding of bi-level images", *ITU-T Recommendation T.88*, March 2000.

[8] A. Zaghetto, R. L. de Queiroz and D. Mukherjee, "MRC compression of compound documents using threshold segmentation, iterative data-filling and H.264/AVC-INTRA", *Indian Conf. on Computer Vision, Graphics and Image Processing*, Bhubaneswar, India, Dec. 2008.

[9] A. Zaghetto and R. L. de Queiroz, "Iterative pre- and post-processing for MRC layers of scanned documents", *Proc. IEEE Intl. Conf. on Image Processing*, San Diego, CA, USA, pp. 1009-1012, Oct. 2008.

[10] J. Fan, "Text extraction via an edge-bounded averaging and a parametric character model", *Proc. of SPIE Document Recognition and Retrieval X*, Vol. 5010, pp. 8-19, Jan. 2003.

[11] R. L. de Queiroz, "On data-filling algorithms for MRC layers", *Proc. IEEE Intl. Conf. on Image Processing*, Vancouver, Canada, Vol. II, pp. 586-589, Sep. 2000.

[12] G. Lakhani and R. Subedi, "Optimal filling of FG/BG layers of compound document images", *Proc. IEEE Intl. Conf. on Image Processing*, Atlanta, USA, pp. 2273-2276, Oct. 2006.

[13] R. L. de Queiroz, "Pre-processing for MRC layers of scanned images", *Proc. IEEE Intl. Conf. on Image Processing*, Atlanta, USA, pp. 3093-3096, Oct. 2006.