

ITERATIVE PRE- AND POST-PROCESSING FOR MRC LAYERS OF SCANNED DOCUMENTS

Alexandre Zaghetto and Ricardo L. de Queiroz

Department of Electrical Engineering, Universidade de Brasilia, Brazil
zaghetto@image.unb.br, queiroz@ieee.org

ABSTRACT

The Mixed Raster Content (MRC) document compression standard (ITU T.44) specifies a multi-layer multi-resolution representation of a compound document. The model is very efficient for representing sharp text and graphics onto a background. However, since the mask layer is binary, it is difficult to deal with scanned data and soft edges. The edge transitions do not fully belong to the foreground neither to the background, and cause some “halo” to the object edges using the MRC model. This paper presents an algorithm that builds an edge sharpening map and iteratively parameterizes the original edge “softness” at the encoder. The generated map and the “softness” parameters are, then, used to reconstruct the original soft edges at the decoder. Experimental results are presented, showing that the method can yield 1.5 dB gains in PSNR, in the compression ranges of interest.

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

1. INTRODUCTION

Compound or mixed documents are bitmaps containing a mix of text, graphics and pictures. These compound documents often challenge traditional compression mechanisms, which are generally developed with a particular image type and application in mind. The mixed raster content (MRC) imaging model [1]-[3], has been proposed as a multi-layer multi-resolution 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). 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. The compressed layers are then packaged for delivery to the decoder. MRC has been proposed and/or accepted for several standards [2], [4], [5].

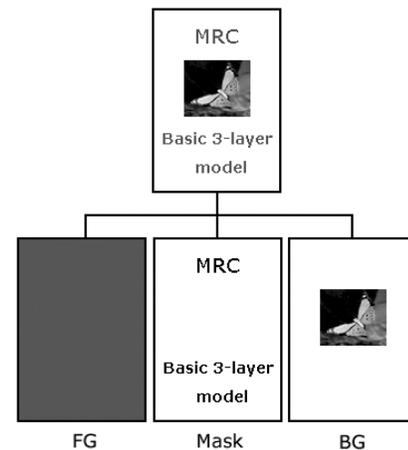


Fig. 1. Illustration of the MRC imaging model.

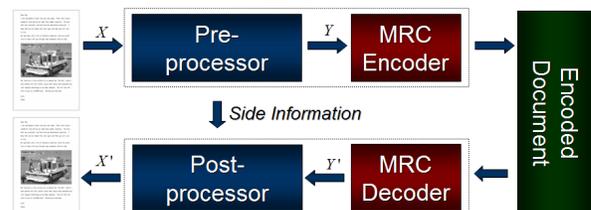


Fig. 2. Proposed framework for dealing with scanned data in MRC.

This paper proposes techniques of pre- and post-processing for MRC layers of scanned documents, thus not addressing the MRC encoding and decoding problem. Fig. 2 describes the complete scheme in which our solution is inserted. First, the original document X is input to a pre-processor that outputs a 3-layer MRC representation of the document, Y . The pre-processor also constructs an edge sharpening map and estimates the original edge “softness”, both considered as side information. The layers are, then, MRC encoded, generating the encoded document. At the decoder, the encoded version of the document is decoded by the MRC decoder, resulting in Y' , a reconstructed version of MRC layers, which, together with the side information is used by the post-processor to assemble the reconstructed soft edge version X' of the document.

Since layer decomposition is not the main focus of this

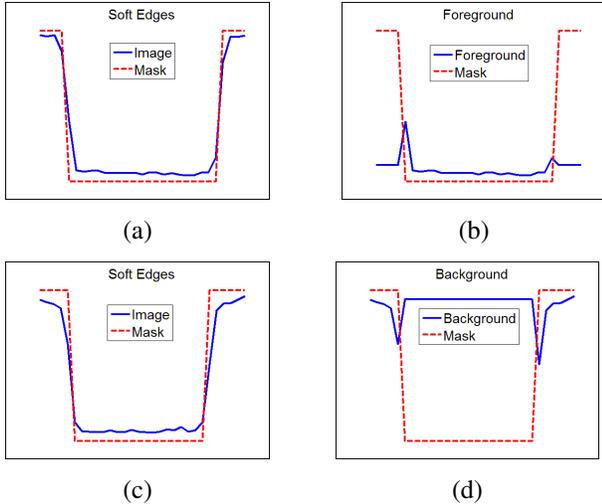


Fig. 3. 1D examples of segmenting soft edges. In (a) and (c) the soft edges and their respective masks are indicated with a dotted line, which determines FG/BG. In the FG (b) or BG (d) planes, the unused parts are replaced by constant values, while the edge regions are not, causing the unwanted spikes.

paper, we used Fan’s text segmentation algorithm [8] in our tests for convenience.

We are concerned with a distributed scenario such as the web, where a document is stored once, somewhere, and many users are to retrieve and decode the document at a later time. Hence, we do not assume any complexity restrictions.

2. THE PROBLEM WITH SOFT EDGES

In MRC, typically, the imaged text has about the same shape and size as that in the mask plane. When the image is 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 background in one plane and all the foreground in another. An illustration is shown in Fig. 3, where two soft edges are depicted in Fig. 3(a) and (c) along with the sharp corresponding mask (in dotted lines). The mask determines what is BG or FG. Note that the transition is included in both planes. If, in the FG plane, we replace the parts corresponding to BG with the average of the FG plane for increased compression [6], we would not end up with a line, but with a spike as shown in Fig. 3(b). Same thing occurs with the BG plane in Fig. 3(d).

The effect in images is that of a halo around the text in the FG/BG planes, as illustrated in Fig. 7(a) and (c). This effect is very damaging to compression. Furthermore, since it occurs inside the “useful” region, we cannot do much in this regard using only data-filling techniques.

3. EDGE SHARPENING

The halo cannot be removed with data-filling and we are forced to change the data itself. In effect, removing the halo

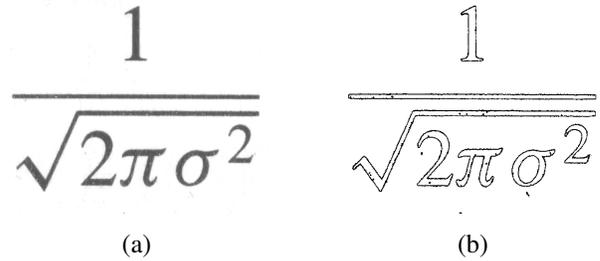


Fig. 4. (a) Zoomed part of original scanned material; (b) pixels to be changed.

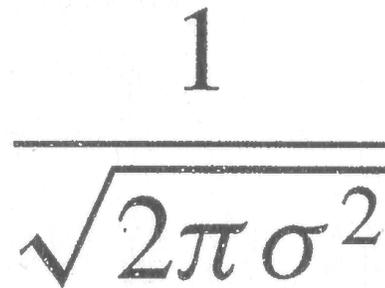


Fig. 5. Resulting image after processing. Note the sharper edges that coincide with the mask edges.

forcefully is equivalent to changing the original image itself to make transitions sharper. We cannot assume that all image edges can cause the problem, nor that all mask transitions are subject to cause halo in scanned material. The effect occurs when transition in the mask coincides with the edges of the image. This means that a sharp mask transition is used to model a soft image transition, thus causing the unwanted spikes. Hence, the first step is to estimate where the halo will possibly occur. Our approach is to find the mask transitions using a simple Sobel operator. These transitions are dilated by a $D1 \times D1$ 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 composing 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:

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

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

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

$$\begin{aligned} C_{FG} &= \{x(i, j) | (i, j) \in (F \cap E), |x(i, j) - m_{FG}| > \epsilon\} \\ C_{BG} &= \{x(i, j) | (i, j) \in (B \cap E), |x(i, j) - m_{BG}| > \epsilon\}, \end{aligned} \quad (2)$$

where ϵ is a tolerance value. For the image in Fig. 4(a), and

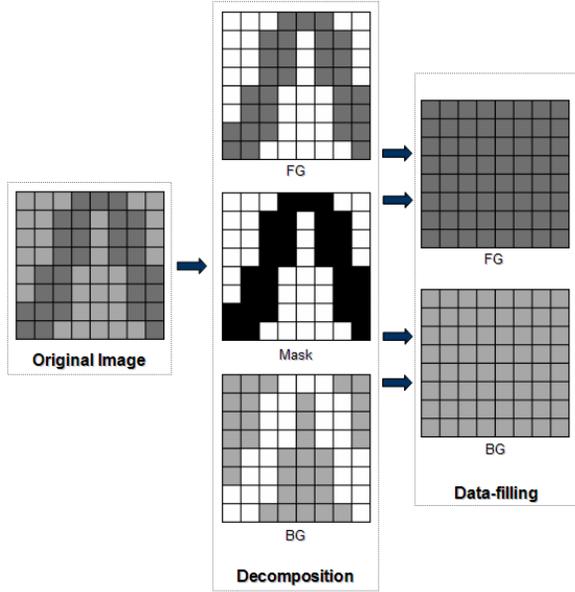


Fig. 6. Diagram of an MRC decomposer, based on a segmenter and plane filling algorithms.

for $\epsilon = 16$ (out of 256 gray levels), the map of the pixels to be changed, i.e. $C = C_{FG} \cup C_{BG}$, is shown in Fig. 4(b). 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} . The result is shown in Fig. 5, where we can note how sharp the transitions are. Note that the process is not the same as sharpening the image. Only regions that might be affected by the halo are processed. General FG/BG is left untouched.

4. DATA-FILLING

Once the image is segmented there will be “don’t care” regions on FG and BG layers. Pixels assigned to the BG will be marked as “don’t care” on the FG, and vice-versa. These pixels can be replaced by anything to enhance compression, as illustrated in Fig. 6. The problem of data-filling over the redundant data has been studied in [6], [7]. This paper uses the iterative wavelet-based plane filling proposed by De Queiroz [9].

Fig. 7(a)-(d) show FG and BG before and after halo processing and data-filling procedures. Note how the pre-processing improved the quality of the FG/BG planes.

5. EDGE SOFTENING

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.

If one can send as side information the map in C , we can blur only the pixels that belong to this map. The result is an image like the one shown in Fig. 8, which was reconstructed using an $h \times h$ Gaussian filter with standard deviation σ ($h = 11$ and $\sigma = 1.5$, in this example). Note how only the edges

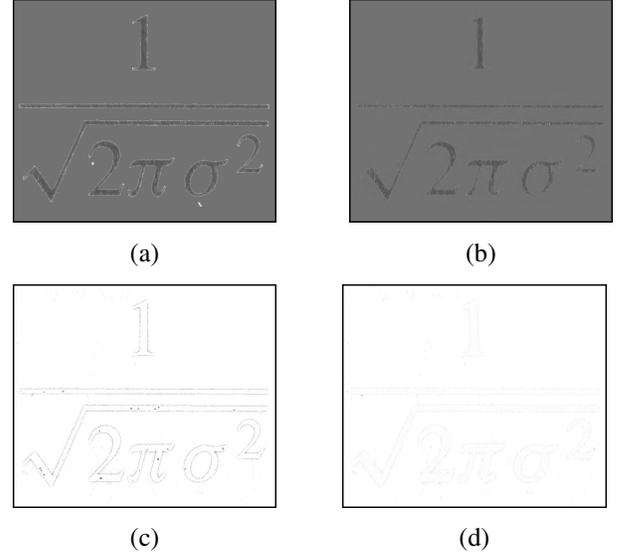


Fig. 7. Original (a) FG and (c) BG; processed (b) FG and (d) BG. Note the halo around the unprocessed FG/BG text.

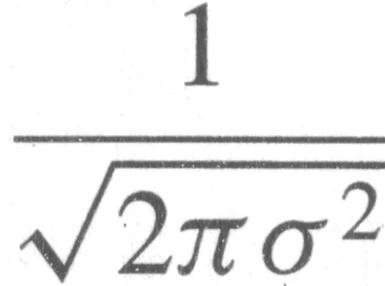


Fig. 8. Resulting image after edge softening. Pixels in C were blurred by a 11×11 Gaussian filter with standard deviation $\sigma = 1.5$.

that were pre-processed were softened again. In general, the image is not blurred.

6. ITERATIVE ESTIMATION OF PRE- AND POST-PROCESSING PARAMETERS

One can notice that in Sec. 3 the tolerance value ϵ , from which we find the map C , was arbitrarily set. The same occurred in Sec. 5 with Gaussian filter parameters h and σ . The quest is to estimate the best value of ϵ , h and σ in a rate-distortion sense. For this, we approximate the solution by minimizing the following cost function:

$$J(\epsilon, h, \sigma) = \alpha_1 V_{FG} + \alpha_2 V_{BG} + \alpha_3 N_t + \alpha_4 D, \quad (3)$$

where α_i are weighting factors, V_{FG} and V_{BG} are the variances of pixels in FG and BG and N_t is the number of horizontal transitions of C . D is the distortion incurred by the pre/post-processing algorithm and is defined as,

$$D = \sum_{ij} |x(i, j) - \tilde{x}(i, j)|, \quad (4)$$

where $x(i, j)$ and $\tilde{x}(i, j)$ represent the original and reconstructed document, respectively.

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

1. Initialize best epsilon, ϵ_{best} , as 32.
2. Initialize previous epsilon, ϵ_{prev} , as 0.
3. While $\epsilon_{best} \neq \epsilon_{prev}$
 - (a) $\epsilon_{prev} = \epsilon_{best}$.
 - (b) Generate map C using ϵ_{prev} .
 - (c) Sharpen the edges using C .
 - (d) For a set of h and σ pairs, find the combination $(h_{best}, \sigma_{best})$ that results in the minimum distortion D .
 - (e) For $\epsilon = 0$ to 255
 - i. Generate map C using ϵ .
 - ii. Sharpen the edges using C .
 - iii. Filter edges using Gaussian filter with parameters $(h_{best}, \sigma_{best})$.
 - iv. Calculate and store cost $J(\epsilon, h_{best}, \sigma_{best})$.
 - (f) Find best ϵ that results in the minimum cost J . Make it ϵ_{best} .
 - (g) Go to step 3.
4. Store parameters $\epsilon_{best}, h_{best}, \sigma_{best}$.

In general, after 2 or 3 iterations we can obtain satisfactory results. The choice of weights α_i is empirical. However, there are some few guidelines that we can use. The larger α_1 and α_2 , the lower the number of unwanted spikes that cause halo. This reduces the number of bits needed to encode FG and BG, increases the number of bits needed to encode the map C and reduces quality. In the other hand, the larger α_3 and α_4 , the higher the number of unwanted spikes. This means that the quality is increased, the number of bits used to encode C is reduced, but the FG and BG become more complex. Thus, more bits are required to encode these two layers. The choice of the initial ϵ_{best} is also empirical. The only restriction is that it must be different from the initial ϵ_{prev} .

Fig. 9 shows an example of PSNR plot comparing MRC performance with and without the pre/post-processing steps. Since this paper is not addressing the problem of MRC encoding and decoding, we used a codec based on H.264/AVC-INTRA and JBIG2 [10]. The reader may note that the proposed techniques enhances average PSNR in more than 1dB for bitrates between 0.08 and 0.24 bits/pixel. For the compression ratios of interest (80:1 or 0.1 bits/pixel) the gain surpasses the 1.5 dB mark. Higher bitrates are not of interest since regular coders become competitive for scanned material at those rates.

7. CONCLUSIONS

We have developed methods to counter balance the effects of soft edges in MRC compression of scanned images. A

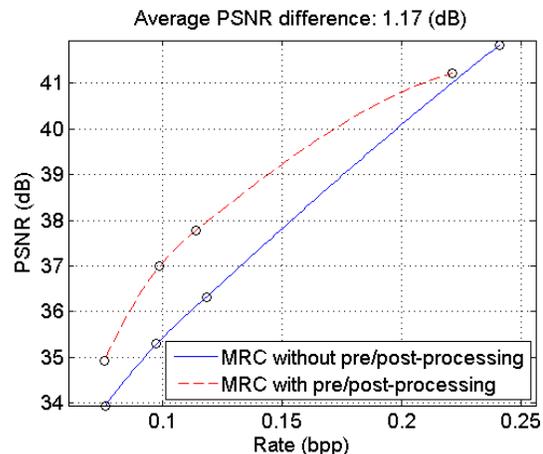


Fig. 9. MRC performance for a scanned text/graphics image. The proposed processing improved average PSNR by more than 1dB ($\epsilon_{best}=40, h_{best}=11, \sigma_{best}=1.5$).

method to sharpen only the edges that might cause the halo effect was presented. We also developed an algorithm that iteratively estimates at the encoder the original edge softness of the image using Gaussian filters and generates a halo location map in a rate-distortion sense. The filter parameters and the map are used to reconstruct soft edges at the decoder. We have shown performance improvements for low bitrates.

Future work will consider using the real rate and distortion, the direct determination of E , and the algorithm convergence. All this will be included in a forthcoming Journal paper.

8. REFERENCES

- [1] 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.
- [2] Draft Recommendation T.44, Mixed Raster Content (MRC), ITU-T Study Group 8, Question 5, May 1997.
- [3] R. de Queiroz, R. Buckley and M. Xu, "Mixed raster content (MRC) model for compound image compression", *Proc. EI'99, VCIP*, SPIE Vol. 3653, pp. 1106-1117, Feb. 1999.
- [4] IETF RFC 2301. File Format for Internet Fax. March 1998. <ftp://ftp.isi.edu/in-notes/rfc2301.txt>.
- [5] ISO/IEC JTC1/SC29 WG1, JPEG 2000 Committee, Working Draft 2.0, June 25, 1999.
- [6] R. de Queiroz, "On data-filing algorithms for MRC layers", *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Vancouver, Canada, Vol. II, pp. 586-589, Sep. 2000.
- [7] R. L. de Queiroz, "Compressing Compound Documents", in *The Document and Image Compression Handbook*, edited by M. Barni, Marcel-Dekker, 2005.
- [8] 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.
- [9] R. L. de Queiroz, "Pre-processing of MRC layers of scanned images", *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Atlanta, USA, pp. 3093-3096, Oct. 2006.
- [10] A. Zaghetto and R. L. de Queiroz, "MRC compression of compound documents using H.246/AVC-I", *Simpósio Brasileiro de Telecomunicações*, Recife, PE, Brazil, Sep. 2007.