JPEG2000-Matched MRC Compression of Compound Documents

Debargha Mukherjee, Christos Chrysafis[1], Amir Said
Imaging Systems Laboratory
HP Laboratories Palo Alto
HPL-2002-165
June 6th , 2002*

JPEG2000, MRC, JBIG, wavelets, mask, scanned document

The Mixed Raster Content (MRC) ITU document compression standard (T.44) specifies a multilayer decomposition model for compound documents into two contone image layers and a binary mask layer for independent compression. While T.44 does not recommend any procedure for decomposition, it does specify a set of allowable layer codecs to be used after decomposition. While T.44 only allows older standardized codecs such as JPEG/JBIG/G3/G4, higher compression could be achieved if newer contone and bi-level compression standards such as JPEG2000/JBIG2 were used instead. In this paper, we present a MRC compound document codec using JPEG2000 as the image layer codec and a layer decomposition scheme matched to JPEG2000 for efficient compression. JBIG still codes the mask. Noise removal routines enable efficient coding of scanned documents along with electronic ones. Resolution scalable decoding features are also implemented. The segmentation mask obtained from layer decomposition, serves to separate text and other features.

* Internal Accession Date Only Approved for External Publication

[1] Divio Inc.
To be published in and presented at IEEE International Conference on Image Processing, 22-25 September 2002, Rochester, New York

JPEG2000-MATCHED MRC COMPRESSION OF COMPOUND DOCUMENTS

Debargha Mukherjee, Christos Chrysafis, Amir Said

Compression and Multimedia Technologies Group
Hewlett Packard Laboratories
Palo Alto, CA 94304

ABSTRACT

The Mixed Raster Content (MRC) ITU document compression standard (T.44) specifies a multilayer decomposition model for compound documents into two contone image layers and a binary mask layer for independent compression. While T.44 does not recommend any procedure for decomposition, it does specify a set of allowable layer codecs to be used after decomposition. While T.44 only allows older standardized codecs such as JPEG/JBIG/G3/G4, higher compression could be achieved if newer contone and bi-level compression standards such as JPEG2000/JBIG2 were used instead. In this paper, we present a MRC compound document codec using JPEG2000 as the image layer codec and a layer decomposition scheme matched to JPEG2000 for efficient compression. JBIG still codes the mask. Noise removal routines enable efficient coding of scanned documents along with electronic ones. Resolution scalable decoding features are also implemented. The segmentation mask obtained from layer decomposition, serves to separate text and other features.

1. INTRODUCTION

Content-adaptivity is the key to efficient compression of rasterized compound documents – comprising a combination of text, graphics and images. In a *layered* codec, this adaptivity is implicit. A redundant representation of a given document is first obtained by intelligently decomposing it into multiple layers, and then standard *non-adaptive* coders are used to code the layers independently. It is the decomposition that contains the magic, and not the layer coders, although decomposition must use knowledge of the layer coders.

Mixed Raster Content (MRC) is an ITU standard T.44 [1], [2], [3], specifying both a layered imaging model, as well as the exact syntax how the layered representation is to be conveyed in a coded bit-stream. While it also specifies the set of allowable layer codecs to be used after decomposition (JPEG [4], JBIG [5], G3, G4), it is reticent about how the representation may be obtained. The MRC imaging model, in its basic mode, decomposes a compound document into 3 layers – an image *background layer* (BG), an image *foreground layer* (FG), and a binary *mask layer*. This is a redundant representation because each original 24 bit color pixel, is represented using two color pixels (from the foreground and background layers) and a binary value (from the mask layer), to make a total of 49 bits. In order to reconstruct the image, a pixel is taken either from the reconstructed foreground layer or from the reconstructed background layer, depending on the value of the corresponding bit in the decompressed mask layer. The spirit of the layered representation is that even though it is redundant initially, if the decomposition is intelligent enough, the three layers when compressed individually can yield a very compact and high quality representation of the compound document.

The apparently simple model has proved to be extremely powerful for compression of compound documents. The model has been successfully employed in several commercial products, such as DjVu [6] and Digipaper [7]. However, they are not ITU standard compliant because they use proprietary layer encoders not supported by the ITU standard [2]. It is only recently that Mukherjee et. al [8] presented a fully T.44 compliant JPEG-matched MRC codec.

With the emergence of newer standardized coders such as JPEG2000 [9] and JBIG2, it is but natural that MRC should incorporate them as possible layer codecs, besides the already supported older standards such as JPEG and JBIG. In anticipation of the inclusion of JPEG2000 in MRC, in this work, we build on [8] and develop a more efficient MRC codec that uses JPEG2000 as the image layer codec, while otherwise remaining compliant with the ITU standard T.44. We develop an analysis procedure matched to JPEG2000 for 3-layer decomposition of a compound document, leading to a compact bit-stream that would be compliant with the ITU standard T.44 if the JPEG2000 embedded bit streams were replaced with JPEG. JBIG [5] is still used for the binary mask layer, although JBIG2 can yield improvements for text intensive documents. In performance, this JPEG2000-MRC codec achieves compression ratios higher than [8] but at somewhat higher complexity. Technologies like DjVu achieve higher compression ratios, but at the expense of fidelity of image representation and significantly higher encoding complexity.

2. JPEG2000-MATCHED MRC

A schematic of our JPEG2000-MRC encoder is shown in Figure 1. The algorithm works on independent stripes of image data rather than a full image in order to maintain a tractable run-time memory requirement. While the standard specifies exactly most of the components in the figure, it does not specify or even recommend any scheme for the central box, corresponding to the stripe analysis and decomposition routine. Indeed, for every stripe there are literally zillions of possible decompositions and associated stripe encoding parameters, and it is impossible to try each of them out in order to obtain the optimal one in a rate-distortion sense. The challenge is to obtain a near-optimal decomposition in terms of compactness of the coded bit-stream and quality of the reconstructed image, while staying within a reasonable complexity constraint.

In order to accomplish this task efficiently, the analysis

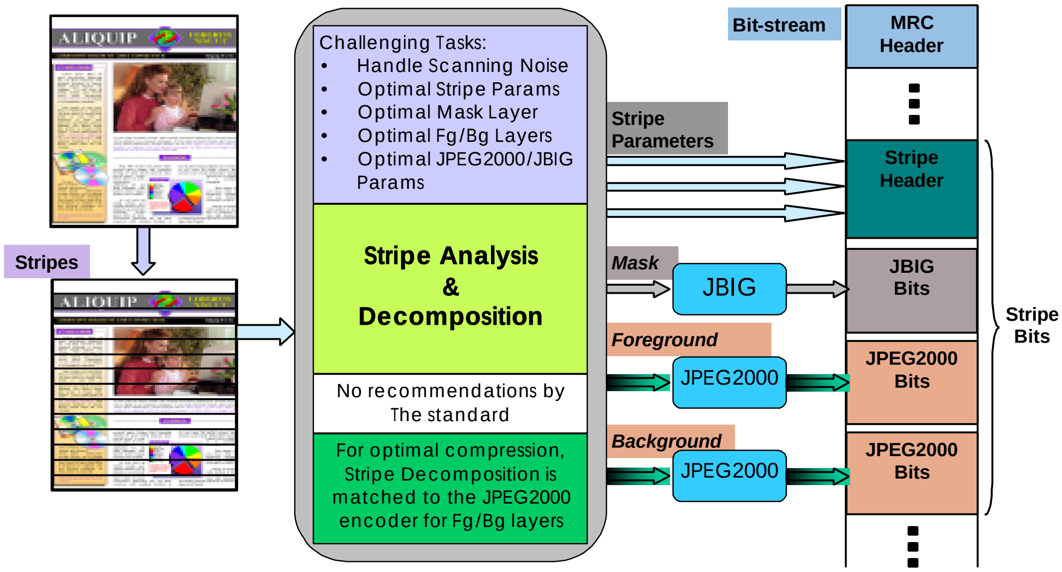


Figure 1. Schematic of the MRC encoder

algorithm should not only analyze the input stripe thoroughly, but also consider the characteristics of the particular coders that are to be used to code the decomposed layers after the analysis. A decomposition algorithm that is optimal for coder A for the foreground and background layers may not be optimal for coder B, and vice versa. Thus, the segmenter used in [8] for block-based JPEG, is quite inappropriate for use with the wavelet based JPEG2000 encoder, and will lead to expansion rather than compression.

The performance of any MRC codec is dependent less on a good match between the decomposition scheme and the mask coder, than on a good match between the decomposition scheme and the image layer coder. In our implementation, the mask layer is still coded with JBIG, but JBIG2 may also be used without needing to change the decomposition scheme.

3. MRC STRIPE SEGMENTER

3.1. Analysis Tasks

The MRC syntax allows for a set of parameters to be transmitted for each stripe. In order to obtain a compact bit-stream, the analysis routine should not only obtain the best decomposition, but also decide on these parameters in an optimal manner. Many of these parameters are related to the fact that the foreground and background layers may be of smaller size than the stripe itself. Parameters *spatial offsets* and *sizes* define the size and position of the coded foreground and background layers, while two other parameters *foreground base color* and *background base color*, are used to fill up the foreground and background layers in portions outside the coded regions as specified by the offset and size parameters.

Once the mask decomposition has derived, the foreground and background layers show holes corresponding to pixels that go to the other layer. These are essentially *don't-care* pixels because they are never used in the reconstruction process at the decoder end. Nevertheless, because the foreground and background are JPEG2000-coded as a solid rectangular image, the holes in them need to be interpolated with some values. Indeed, how the holes are filled up have a significant impact on the efficiency of JPEG2000 compression, and as such, the objective should be to fill the holes with values that make the layer easiest to code, yielding the highest compression. This process is an essential step in layer decomposition and is referred to as layer interpolation.

To summarize, the analysis routine should decide on the following sets of parameters on a per stripe basis:

- The offsets and sizes of the coded FG and BG layers.
- FG and BG base colors.
- The full-resolution binary mask layer.
- Interpolation of the don't-care pixels (holes) in the FG and BG layers.
- JPEG2000 parameters for the FG and BG layers.
- JBIG parameters for the mask layer.

Because many of these parameters are related, a truly optimal analysis scheme needs to jointly optimize all of them. However, because of practicality considerations, a sequential step-by-step approach is adopted, where at each stage some of the parameters are determined while holding the others fixed. The most critical step is the JPEG2000-matched core algorithm that derives the mask, after the layer offsets and base colors have been determined. Additionally, fixed parameters are used for JPEG2000 and JBIG encoding of the image and mask layers. This yields a near-optimal solution, which nevertheless is good for all practical purposes. In the next section, the step-by-step procedure is discussed in detail.

3.2. Analysis Procedure

The flowchart in Figure 2 shows the broad steps involved in the coding operation. Before stripe analysis commences, a *preprocessing* routine may be used to remove noise from scanned documents. For scanned documents, a *text*

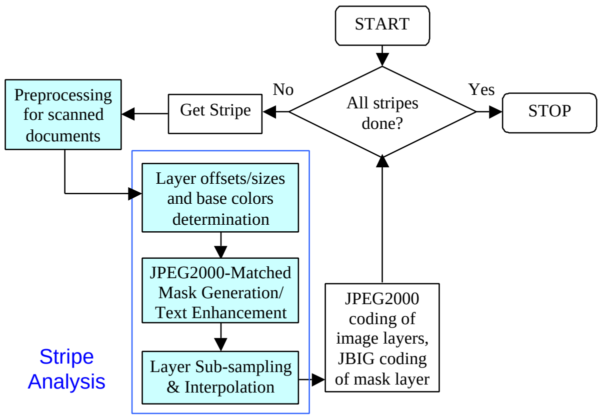


Figure 2. Stripe Analysis Architecture

enhancement feature in the *mask generation* algorithm is also turned on to enhance the quality of text. Both the preprocessing module and the text enhancement feature can be skipped entirely for electronic documents. The main principles behind the three basic components in stripe analysis are described in the next sub-sections.

3.2.1 Layer offsets and base color determination

This is the first step of stripe analysis where the foreground and background layer offsets and sizes, as well as their base colors are determined. If the compound document has margins of constant colors, they can be conveyed very economically by means of the offset and base color parameters allowed by the MRC syntax, without having to code them explicitly with JPEG2000. The objective is to find the thickest margins along the edges of a stripe, consisting of only two colors, so that the residual coded region in the image layers are minimized. The task is accomplished by analysis of rows and columns of the input stripe, starting from the periphery and going inwards. Of the two colors found, the lighter is assigned to the background and the darker is assigned to the foreground, and the mask assignment for these pixels is made accordingly.

3.2.2 Symmetric JPEG2000-matched mask generation

This is the core algorithm in the stripe analysis process that separates the majority of stripe pixels into foreground and background layers. The domain of operation is the reduced region computed in the previous step. The algorithm is designed to mitigate the edges as much as possible within the same layer, so that wavelet based JPEG2000 encoding subsequently on the layers will lead to a very compact coded representation. We further impose strict complexity requirements unlike the full optimization approach in [10].

The idea is to separate pixels in a stripe into foreground and background layers based on local contrast. Decisions are made in units of small decision regions, which are square windows typically of size 6×6 to 8×8, that are traversed in raster scan order for the entire stripe (See Figure 3). Let the (*ij*)th decision region be called $\mathbf{D}_{ij}$. Around $\mathbf{D}_{ij}$ is a larger square analysis region $\mathbf{A}_{ij}$, typically of size 10×10 to 12×12, over which local contrast is computed. For each decision region already covered, the algorithm maintains two histories of average pixel values F_{kl} and B_{kl}, corresponding to average foreground and background region colors respectively in decision region $\mathbf{D}_{kl}$. For each $\mathbf{D}_{ij}$, either all pixels are assigned to the foreground layer, or all pixels are assigned to the background layer, or some pixels are assigned to the foreground layer and some to the background layer. If $\mathbf{A}_{ij}$ is of high contrast, the pixels in $\mathbf{D}_{ij}$ are separated into two groups using the vector 2-means algorithm. All pixels in analysis window $\mathbf{A}_{ij}$ are used for training the classifier, but the classification is finally applied only to pixels in decision region $\mathbf{D}_{ij}$. Of the two cluster means, the lighter (higher luminance) is always assigned to the background layer, while the darker is assigned to the foreground layer. The mask is chosen accordingly. The history values F_{ij} and B_{ij} are assigned to the cluster means. If on the contrary, $\mathbf{A}_{ij}$ is of sufficiently low variance, $\mathbf{D}_{ij}$ is assigned entirely to the foreground or background layer depending on whether its pixels are closer to the value of $(F_{i,j-1} + F_{i-1,j})/2$ or $(B_{i,j-1} + B_{i-1,j})/2$. If $\mathbf{D}_{ij}$ is assigned to foreground, F_{ij} is assigned to be the mean of $\mathbf{A}_{ij}$, while B_{ij} is assigned to be $(B_{i,j-1} + B_{i-1,j})/2$. The scheme automatically maintains smoothness between successive decision regions in the foreground and background layers, to enable good JPEG2000 compression.

Note that unlike some other approaches to mask generation, this is a symmetric approach based purely on a smoothness criterion. Consequently there is no interpretation for the foreground and background layers other than that they are an internal representation. The mask generated however, is

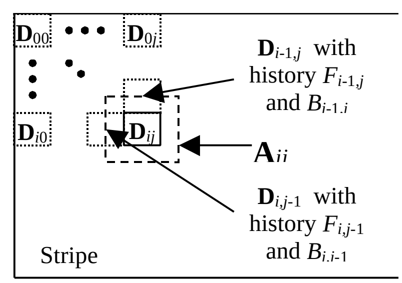


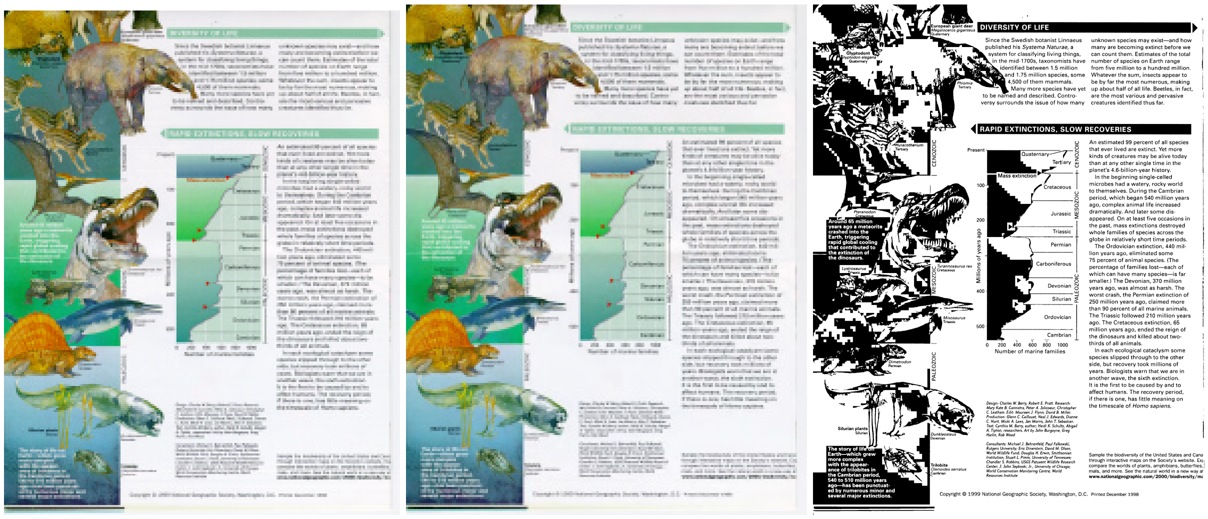
Figure 3. Mask Generation Algorithm

often a meaningful segmentation of the color document.

3.2.3 Fast Layer Interpolation & Sub-sampling

The purpose of the interpolation routine is to fill up the holes in either image layer by pixels being assigned to the other layer, so that solid rectangular layers that JPEG2000 requires are obtained.

Because the layer coder employed is wavelet-based JPEG2000, smoothness guarantees compactness. Pixels are traversed in raster scan order in each layer, and for each hole pixel encountered, interpolation is performed by a weighted averaging with Gaussian weights, in a P×P sliding window (where P is typically 5 or 7) centered on it. For the already interpolated causal pixels in raster scan order, all pixels are considered in the interpolation. However, for the non-causal pixels, only those that are known to be in the same layer are considered. The holes in the non-causal pixels cannot be considered because their values are not known yet. The



(a) Original scan (21709 KB) (b) Reconstructed document (178 KB) (c) Binary Mask Layer

Figure 4. (a) Original 300 dpi full color National Geographic scan, and (b) Reconstructed document, with file sizes, along with (c) the binary Mask Layer. The JPEG2000-MRC codec uses 2´2 sub-sampling of image layers.

averaging is normalized by the sum of weights actually used for each pixel. The above algorithm generates a very smooth image that a wavelet coder like JPEG2000 can readily compress efficiently, and the redundancy due to each pixel being compressed twice is virtually eliminated.

An alternative method performs a similar averaging but in multiple passes, where in each pass only unfilled holes adjacent to relevant pixels are filled up, until all holes have been covered. While this method yields slightly higher compression than the one-pass method, its complexity is also higher.

Sub-sampling of the foreground and background image layers is an effective means for achieving high compression ratios. The interpolated layers are passed through a low pass filter, and downsampled to yield the layers that are actually input to the JPEG2000 coder. For most documents, 2×2 downsampling produces good quality reproduction.

4. DOCUMENT COMPRESSION RESULTS

Almost all computer-generated 300 dpi compound documents, however complex, can be represented at full-resolution with high quality at 100:1 compression. For simpler documents, the compression ratio is often more than 175:1 KB. With 2×2 sub-sampling of the image layers, the compression ratio easily exceeds 250:1 for most documents.

For scanned (noisy) documents, the coded file-size can be made less than 200 KB by moderate downsampling of the image layers. A test suite of 300 dpi scanned documents yielded an average compression ratio of 125:1 with 2×2 layer sub-sampling, but the quality was superior to algorithms like DjVu.

Figure 4 presents the compression results for a 300 dpi scan from the National Geographic. The mask layer generated by the coding process, is also shown. As seen, the mask serves well to separate text and other features from the document.

5. CONCLUSION

We presented a full-color compound document codec compliant with the ITU standard T.44, using JPEG2000 for the image layers, and JBIG for the mask layer. Our decomposition method, matched to JPEG2000, keeps the encoding complexity low enough to make scan-to-email, scan-to-web, or scan-and-distribute type applications feasible. Resolution scalability based on JPEG2000 is implemented at the decoder to enable viewing documents on machines with varying capabilities.

6. REFERENCES

[1] Peter S. MacLeod, Luc Vincent, Xianing Zhu, "Method and apparatus for compressing color or gray scale documents," *US Patent Number:* 5,778,092; July 7, 1998.

[2] "Mixed Raster Content (MRC)," *ITU-T Recommendation T.44, Study Group-8 Contribution*, 1998.

[3] R. L. de Queiroz, R. Buckley, 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.

[4] W. P. Pennebaker, J. L. Mitchell, "JPEG: Still image compression standard:" Van Nostrand Reinhold, 1993.

[5] Information technology – Coded representation of picture and audio information – Progressive bi-level image compression, *ITU-T Recommendation T.82*, March 1995.

[6] P. Haffner, L. Bottou, P.G. Howard, P. Simard, Y. Bengio, Y. LeCun, "High Quality document image compression with DjVu," *Journal of Electronic Imaging*, pp. 410-25, July 1998.

[7] D. Huttenlocher, P. Felzenszwalb, W. Rucklidge, "Digipaper: A versatile color document image representation," *Proc. Int. Conf. Image Processing*, Kobe, Japan, Oct 1999.

[8] D. Mukherjee, N. Memon, A. Said, "JPEG-Matched MRC compression of compound documents," *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct 2001.

[9] "JPEG 2000 Image Coding System," *ITU-T Recommendation T.800*, ISO/IEC FCD 15444-1, 2000.

[10] R. L. de Queiroz, Z. Fan, T. D. Tran, "Optimizing block-thresholding segmentation for multilayer compression of compound images," *IEEE Trans. Image Proc.*, vol. 9, no. 9, Sept 2000.