Mixed Raster Content (MRC) Model for Compound Image Compression

Ricardo de Queiroz, Robert Buckley and Ming Xu
Corporate Research & Technology, Xerox Corp.
[queiroz@wrc.xerox.com, rbuckley@crt.xerox.com, ming_xu@wb.xerox.com]

Keywords: mixed raster content, text, compression

ABSTRACT

This paper will describe the Mixed Raster Content (MRC) method for compressing compound images, containing both binary text and continuous-tone images. A single compression algorithm that simultaneously meets the requirements for both text and image compression has been elusive. MRC takes a different approach. Rather than using a single algorithm, MRC uses a multi-layered imaging model for representing the results of multiple compression algorithms, including ones developed specifically for text and for images. As a result, MRC can combine the best of existing or new compression algorithms and offer different quality-compression ratio tradeoffs. The algorithms used by MRC set the lower bound on its compression performance. Compared to existing algorithms, MRC has some image-processing overhead to manage multiple algorithms and the imaging model. This paper will develop the rationale for the MRC approach by describing the multi-layered imaging model in light of a rate-distortion trade-off. Results will be presented comparing images compressed using MRC, JPEG and state-of-the-art wavelet algorithms such as SPIHT. MRC has been approved or proposed as an architectural model for several standards, including ITU Color Fax, IETF Internet Fax, and JPEG 2000.

1.Introduction

With recent advances in data processing systems and in electronic imaging and scanning devices, documents are now present in a wide spectrum of printing systems. From offset printers to home desktop computers, documents in digital form became common place. Frequently, documents are available as bitmaps and may contain text, graphics and pictures. As typical documents are often generated at a reasonably high resolution, document image sizes are invariably large and commonly consume several megabytes for storage. Furthermore, the final destination for those documents are frequently parties other than those who generated them. Thus, it is desirable to possess the ability to transmit those large document images. Storage or transmission of large amounts of data is often costly and image compression is a necessity. Many standard compression algorithms are available today and in common use commercially. More are continually being developed to improve on existing methods or to meet special requirements. As a rule, any one compression algorithm was developed with a particular image type and characteristic, and a particular application in mind. For a different image type or application, a given algorithm either does not apply or does not perform as well as some other, better-tailored algorithm. No single algorithm is best across all image types or applications.

Different image classes require different coding fidelity. When compressing text, it is important to preserve the edges and shapes of characters accurately to facilitate reading. Once the text is binarized, its compression is typically lossless since coding errors in text are easily perceived. The human visual system, however works differently for typical continuous-tone images because of the richness of patterns and frequency contents. Errors in high frequency components are better masked and lossy compression is usually employed, since lossless compression is often ineffective in this case. In terms of image resolution, text requires much higher resolution than pictures. Actually, roughly speaking, text requires few bits per pixel but many pixels per inch, while pictures require many bits per pixels but fewer pixels per inch. Image compression has been very intensively studied around the world and we cannot possibly reference adequately the most notable algorithms. However, in terms of international standards the notable algorithms for binary image compression are MH[1], MMR[2], JBIG[3] and the forthcoming JBIG-2[4]. Multi-level compression algorithm standards are JPEG[5] and the forthcoming JPEG-2000[6].

The differences between text and con-tone images become significant when it comes to compressing a compound image, such as the scanned image of a page containing text and pictures. For the text, lossless binary compression is suitable; for the pictures, lossy continuous-tone compression is preferable. The Mixed Raster Content (MRC) proposal allows both to be used within a single raster image. The idea is to decompose the raster into several image layers each one containing a distinct image class. Hence, different layers can be compressed more efficiently with off-the-shelf standard single-purpose compression systems. MRC is not a file format or compression system, but an imaging model that can be used as an architectural framework for image compression systems, which, in their turn can be wrapped into a file format.

Document compression is frequently linked to facsimile systems, in which large document bitmaps are compressed before transmission over telephone lines. The facsimile systems that most people are familiar with today are black-and-white (binary images) and conform to international standards set by the ITU-T (Telecommunication Standardization sector of the International Telecommunication Union, formerly known as the CCITT). These standards specify the protocols and bi-level coding procedures that sending and receiving stations use to (i) establish and manage a real-time connection, (ii) agree on a data representation and transmission parameters, and (iii) transmit and receive image data. These standards provide three key benefits: acceptable levels of performance and quality, a common reference point for equipment manufacturers and implementers and interoperability among users. Together with the ubiquity of the public switched telephone network (PSTN), these standards have led to the explosive growth in Group 3 black-and-white facsimile that has occurred since 1980. The same convenience and ease of use for color facsimile requires wider use of color scanners, displays and printers; faster modems and communication channels to handle the increased data volume; and equivalent standards for color facsimile. These enablers are already being put in place. For example, the ITU-T last year approved V.34 for facsimile, which supports data rates up to 33.6 Kbps, and it is now available commercially in fax machines. There is now a focus on new standards to provide enhanced color facsimile services over both the PSTN and the Internet. MRC use for color fax is defined in a forthcoming ITU standard[7,8]. MRC is part of RFC 2301[9] or TIFF-FX (TIFF for Fax eXtended), the IETF file format proposal for Internet Fax. MRC has also been proposed as the JPEG 2000 Architectural Framework[10].

In this paper we present the MRC imaging model in the context of its potential for image compression, by examining the redundant-layers model as well as the decomposition approach in terms of a compression-by-distortion trade-off.

2.Mixed Raster Content (MRC) Imaging Model

MRC uses a multi-layered, multi-resolution imaging model to encode compound raster images. The basic 3-layer MRC model represents a color raster image as two multi-level or color image layers (Foreground and Background) and a binary image layer (Mask). The Mask layer describes how to reconstruct the final images from the other two layers. When the Mask layer pixel value is 1, then the corresponding pixel from the Foreground layer is selected for the final image; when it is 0, then the corresponding pixel from the Background layer is selected. An illustration of the imaging model is shown in Fig. 1. The foreground plane is essentially poured through the mask plane onto the background plane. The basic 3-layer model is MRC's most common form. The imaging model, however is composed of basic elementary plane pairs: foreground and mask. Given a background, a foreground plane is imaged onto it through the mask plane composing a new background image. Another foreground layer can be images onto this new background through another mask plane and the process can be repeated several times. The extended MRC model, then, allows for several planes while relying on foreground-mask pairs.

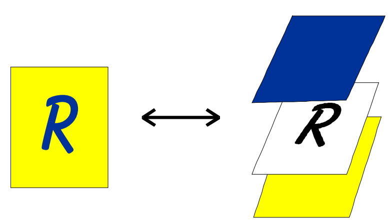

Figure 1 - MRC imaging model for basic 3 plane configuration. Foreground plane in poured into background plane through the mask or selector plane.

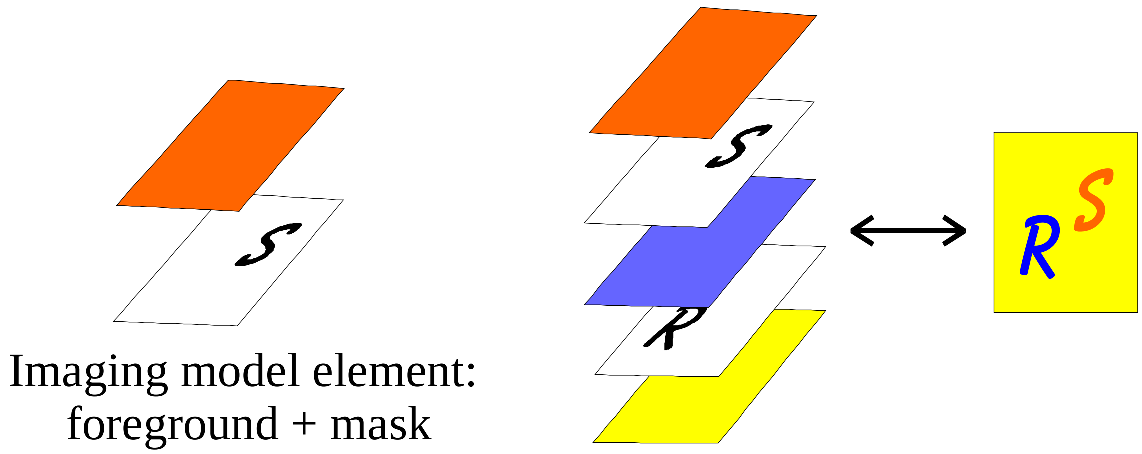

Figure 2 – Basic element of MRC imaging model is the combination of mask and foreground planes. The foreground plane is poured into the background through the mask plane. As a result the MRC model accommodates several object layers.

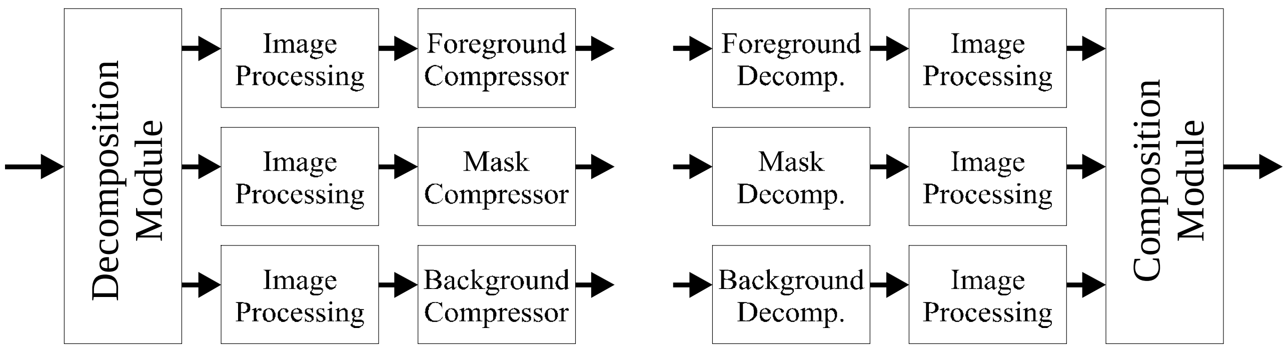

Figure 3 – Block diagram of plane decomposition, compression and composing processes for a basic 3-layer representation.

Once the original single-resolution image is decomposed into layers, each layer can be processed and compressed using different algorithms as shown in Fig. 3. The image processing operations can include a resolution change or color mapping. The compression algorithm and resolution used for a given layer would be matched to the layer's content, allowing for improved compression while reducing distortion visibility. The compressed layers are then packaged in a format, such as TIFF-FX or as an ITU-T MRC data stream for delivery to the decoder. At the decoder, each plane is retrieved, decompressed, processed (which might include scaling) and the image is composed using MRC imaging model.

A page may be represented as one, two, three or more layers, depending on its content. For example, a page consisting of a picture could use the background layer only. A page containing black-and-white text could use the mask layer, with the foreground and background layers defaulted to black and to white.

Layers may contain different dimensions and have offsets associated with them. If a plane contains only a small object, the effective plane can be made of a bounding box around the object. The reduced image plane is then imaged onto the larger reference plane, starting from the given offset (top, left) with given size (width, height). This avoids representing large blank areas and improves compression.

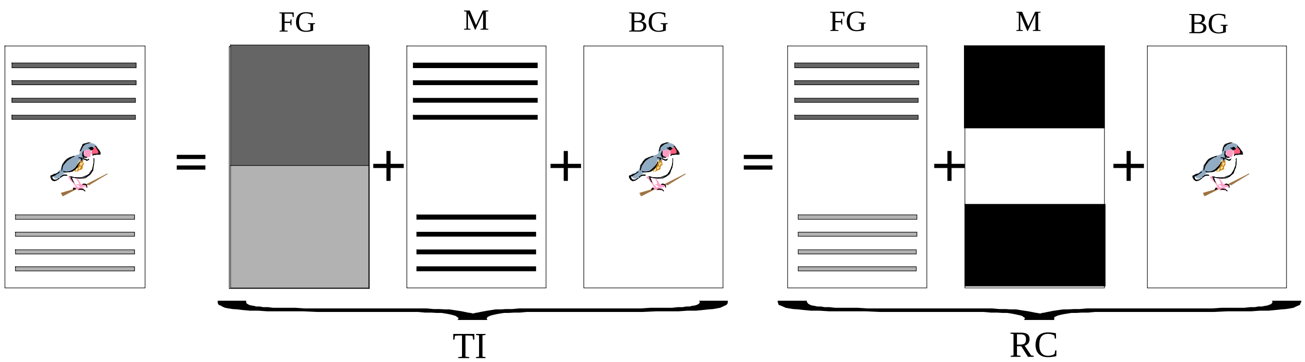

Figure 4 – Typical decomposition approaches yielding the same reconstructed image (in the absence of processing or compression) in which BG=background plane; FG=foreground plane; M=mask plane; TI=transition identification decomposition; RC=region classification decomposition.

3.Decomposition and compression analysis

We will discuss the basic 3-layer decomposition, the reader can easily infer the extension to multiple planes. The degrees of freedom in MRC-based compression are the decomposition process (the Decomposition module in Fig. 3) and the compressors and their associated parameters for each plane. The encoder and decoder would agree a priori on the compressors, which would be part of the standard employing MRC as architectural framework. Decomposition affects the operation of the encoder, but not that of the decoder. Typical approaches to decomposition are illustrated in Fig. 4. The basic approaches are: region classification (RC) and transition identification (TI). In RC decomposition, regions containing text and graphics are identified and represented in a separate (foreground) plane. The whole region is represented in the foreground plane including the spaces in between letters and such. The mask is very uniform with large patches indicating the text and graphics regions, while the background contains the remaining regions, i.e. the document background itself, complex graphics and/or continuous tone pictures. The TI decomposition relies on identifying text and graphics objects. As in the RC approach, the text and graphics regions are identified and the background region is the same. In the TI case, however, the mask and foreground planes represent text and graphics in a different manner. The concept is that the text or graphics ink is poured through the mask plane onto the background. For this, the mask should have the contours of text elements, as illustrated in Fig. 4. Thus, the mask image layer would contain text characters, line art and filled regions, while the foreground layer contains the colors of the shapes in the mask layer, i.e. the color of text letters and graphics.

In both cases the background plane is suitable for continuous tone image compressors such as JPEG or JPEG-2000, unless continuous tone images are not present in the original document and the document was not scanned, in which case other compressors may be more efficient. In the RC case, the mask is very uniform and can be compressed very well. However, the foreground can contain many edges and continuous tone details. Therefore, it may not be very well compressed with typical continuous tone coders such as JPEG. In the TI case, text objects and edges are placed in the binary mask layer, thus, being efficiently encoded using standard binary coders such as MMR, JBIG and JBIG-2. The foreground plane would typically contain large uniform patches and, thus, can also be very efficiently coded even with coders such as JPEG. Besides the foreground plane can be subsampled without much loss in image quality. In other words, edges are moved from the continuous tone plane to the binary one.

The potential gain of the MRC model for compression can be analyzed under the light of its rate-distortion (RD) characteristics. If the image in Fig. 5(a) is compressed with a generic coder A with fixed parameters except for a compression parameter, it will operate under a given RD curve as shown in Fig. 5(b). Another coder B under the same circumstances is said to outperform coder A if its RD curve is as shown in Fig. 5(c) , i..e shifted to the left. The rationale for MRC is to split the image into multiple planes as shown in Fig. 5(d), and to apply to each plane a coder (C, D, and E) whose RD curves are better than those of coder A. In that case, the equivalent coder may have better RD curves than A, despite the overhead associated with a multi-plane representation.

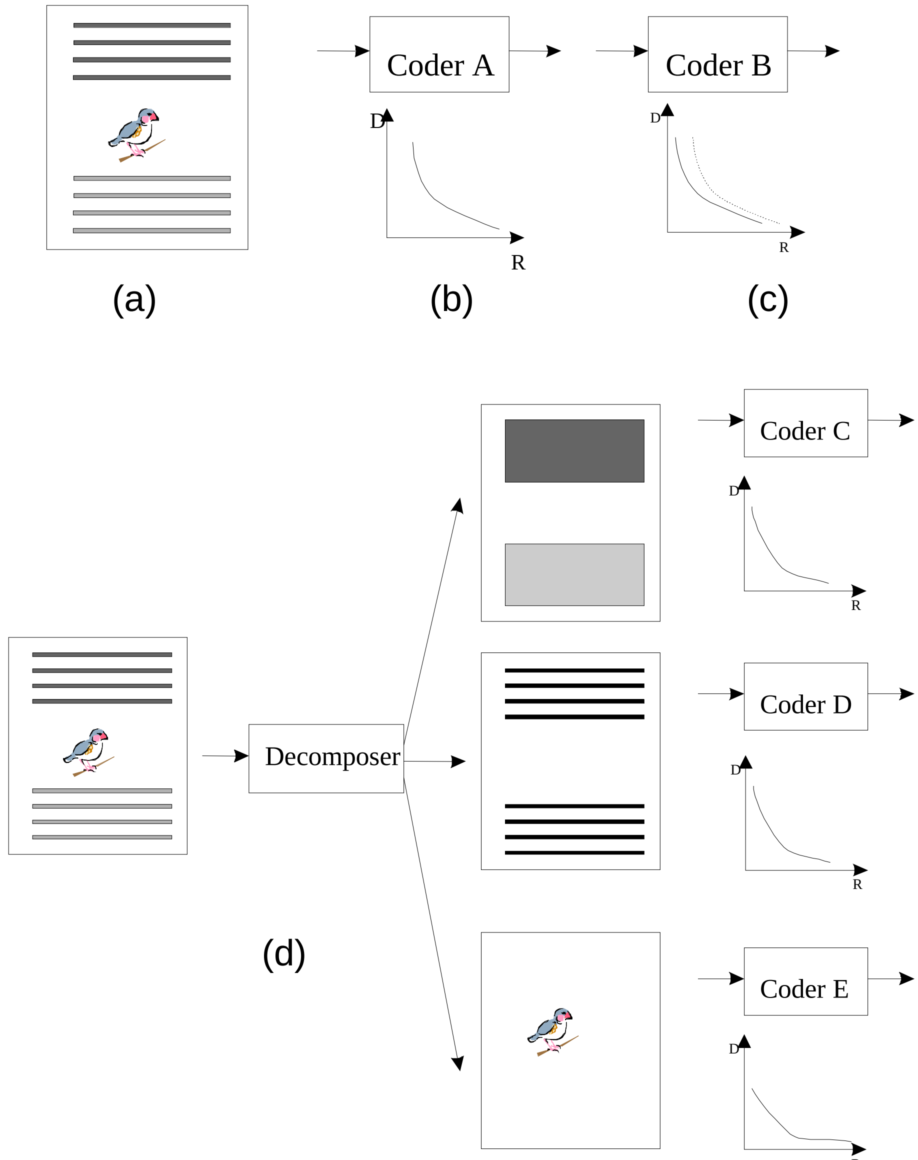

Figure 5 – Interpretation of the multiplane approach as a mean to modify the RD characteristics of coding mixed images. (a) Mixed image; (b) RD curve for given coder A; (c) modified RD curve for resulting coder B; (d) coder B may be achieved by plane decomposition, where each plane undergoes a better tailored coder, thus achieving better RD curves.

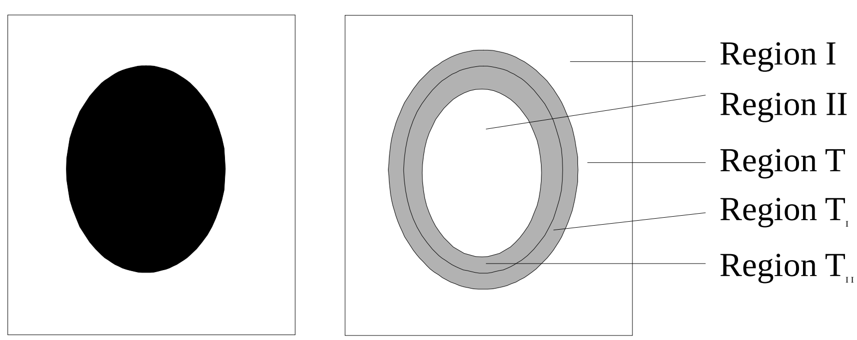

Figure 6 – Diagram of the regions of an image plane with respect to the mask plane shown on the left. Region I comprises the pixels which are to be reconstructed from the background plane while region II comprises those from the foreground plane. We also define a transition region T, which is divided into respective pixels from regions I and II.

The mask controls which pixels from which plane are used in the reconstruction of the image. Fig. 6 shows an example mask plane on the left. Let us label the image regions according to the mask as in Fig. 6. Region I is the region selected for the background plane, region II for the foreground plane and the transition (T) region encompasses some neighbourhood of transition region. The transition region can be further subdivided into subparts TI and TII, belonging to regions I and II respectively. Let the original image be encoded using a single coder S which does not use MRC, spending R_S bits to encode the image with a distortion D_S such that:

$$R_S = R_S^I + R_S^{II} + R_S^T \quad \text{and} \quad D_S = D_S^I + D_S^{II} + D_S^T \ , \tag{1}$$

where the distortion model was chosen to be linear, i.e. overall distortion is the sum of local distortions. If the image is split into the 3 planes (Foreground, Background, Mask) corresponding to the MRC model, then the overall rate and distortion are given by

$$R = R_M + \sum_{\Pi=F,B} \sum_{\Omega=I,II,T} R_\Pi^\Omega \quad \text{and} \quad D = \sum_{\Pi=F,B} \sum_{\Omega=I,II,T_I,T_{II}} D_\Pi^\Omega \tag{2}$$

Note that the mask is encoded without distortion and that redundant pixels, i.e. region II in B plane and region I in F plane do not contribute to overall distortion. Thus

$$D = D_B^I + D_B^{TI} + D_F^{II} + D_F^{TII} \tag{3}$$

If one wants the MRC scheme to outperform the single coder, it is necessary that either or both $R < R_S$ and $D < D_S$. It is sufficient to have

$$R < R_S \text{ and } D < D_S. \tag{4}$$

In a simple coding scenario where the coder for the foreground and background planes is the same as the single coder, we can make the following assumptions: $R_B^I = R_S^I$, $R_F^{II} = R_S^{II}$, $D_B^I = D_S^I$, $D_F^{II} = D_S^{II}$, so that

$$D_S - D = D_S^T - D_B^{TI} - D_F^{TII} \tag{5}$$

$$R_S - R = R_S^T - R_M - R_B^{II} - R_F^I - R_B^T - R_F^T = R_S^T - R_o - R_B^T - R_F^T \tag{6}$$

where R_o is the overhead rate, due to the mask and to redundant data in the continuous planes. Reduction in rate and distortion are achieved iff

$$D_B^{TI} + D_F^{TII} < D_S^T \tag{7}$$

$$R_o + R_B^T + R_F^T < R_S^T \tag{8}$$

So, in the analysis of this simple example, we see that *transition regions are the main regions where compression can be improved by using MRC.* In more detail, improvement comes when:

- distortion in the transition region is less than in the single coder
- the savings in encoding the transition regions (in both B and F) planes compared to the single coder are enough to offset the expenditure of bits to encode the overhead.

In TI decomposition, fortunately, (7) is usually satisfied. In general, $R_B^T < R_S^T$ and $R_F^T < R_S^T$. However the decomposition has to be done in such a way that there will be enough transitions in the image to allow enough savings. Furthermore, the regions chosen to be transitions have to be such that they lead to large savings in bit rate in each plane in order to compensate for the redundant information and the overhead.

In an MRC approach, the planes can be obtained from the original image following the diagram shown in Fig. 7. The pre-processor block can very well replace pixels in redundant regions (e.g. region II and TII in the background plane) with any computer generated data which would reduce the most the distortion and bit rate, i.e. to ensure that (7) and (8) are satisfied.

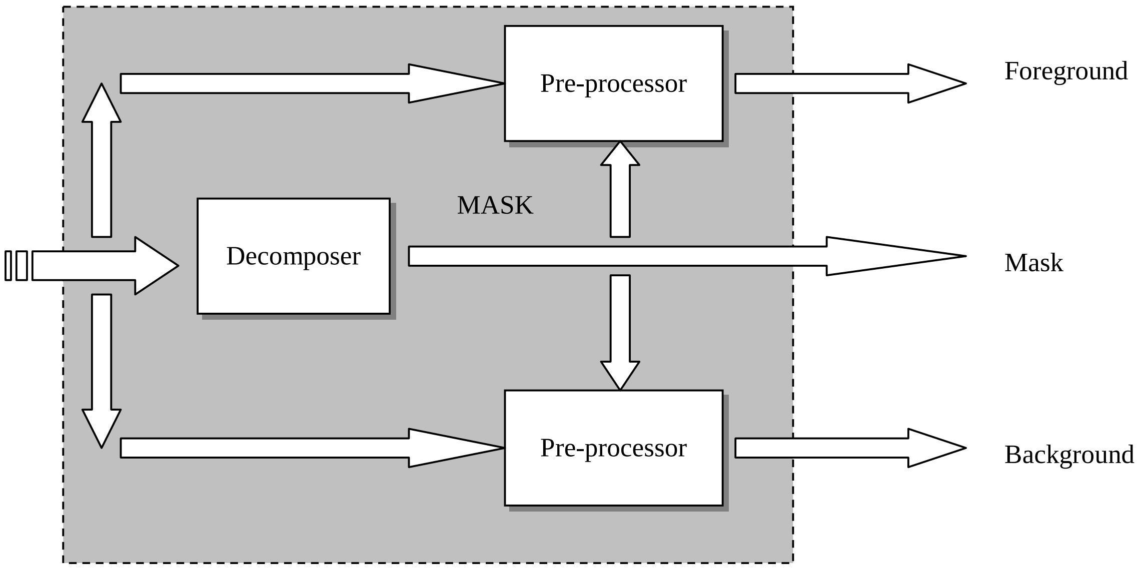

Figure 7 – Decomposer module diagram.

Note that, for a given preprocessing strategy the only output of the decomposer is the mask plane. From that all the planes are calculated in a deterministic way.

In TI decomposition, the transition in the mask occurs for edges in the original image. Hence R_S^T and D_S^T are very high. If, for example, the transition region in each plane is made very smooth, not only will the distortion decrease but the bit-rate can be kept very small. For smooth enough transitions and if we discard the terms R_B^T *and* R_F^T, then the trade-off of MRC can be summarized as:

$$R_o < R_S^T. \tag{9}$$

In other words, MRC is advantageous if the amount of bits saved by not encoding the transition regions is greater than the amount of overhead data (redundant and mask data). Of course the main assumption is that the transition in both planes can be made "smooth" enough to save significantly in both distortion and bit-rate. Also the input image has to contain a sufficient amount of those edges. An image with large text regions is a typical case. If there are only pictorial images, however, it is harder (but not impossible) to make a multiplane MRC outperform the single coder. In the limit, it may be advantageous to place the pictorial image in a single MRC layer, in which case the MRC behaves as a single coder.

In reality, a coding scenario is usually more favorable to MRC than the above example. This is because coders for foreground and background can be selected to outperform the single coder, while the mask plane often compresses very well. For example if the text is placed into the mask (using TI decomposition), techniques such as JBIG, MMR, JBIG-2 exist that compress text well. The foreground contains mainly text color and can be largely subsampled. The background plane contains the pictorial images and paper texture, which are features that do not contain high-resolution details. In that case, moderate subsampling can be carried out before compression. The different nature of the data in each plane allows for very efficient compression with lower error visibility.

4.MRC Standards and File Format

Realizing the value of MRC requires an agreed upon file format for representing and conveying the results of the multi-layered MRC model. MRC was originally approved for use in Group 3 color fax and is described in ITU-T Recommendation T.44. For the storage, archiving and general interchange of MRC-encoded image data, the TIFF-FX file format has been proposed. TIFF-FX (TIFF for Fax eXtended) represents the coded data generated by the suite of ITU recommendations for facsimile, including single-compression methods MH, MR, MMR, JBIG and JPEG, as well as MRC. As IETF RFC 2301, TIFF-FX is a Proposed Internet Standard, currently undergoing interoperability testing. MRC has also been proposed as an architectural framework for JPEG 2000.

5.Compression Example

We performed several tests using MRC technology. We present simple tests to demonstrate its potential for compressing compound (mixed) documents. The performance of MRC will be illustrated with typical pages, containing text and images. Due to space limitations a typical example will be shown for the 24-bits-per-pixel image (8 bits per color separation per pixel) shown in Fig. 8. This image was compressed in a very conservative way to obtain a compression ratio of 70:1. For that we used only “off-the-shelf” coders. The foreground and background planes were compressed using JPEG[5] (standard quantizer and Huffman tables) *without subsampling*. The mask plane was compressed using the ITU-T MMR algorithm. Clearly, there is room for improvement in these algorithms by:

- Applying a more efficient binary coder such as JBIG or JBIG2
- Subsampling the continuous-tone planes before compression. The foreground is extremely redundant and smooth and the background has no sharp edges.
- Applying a more efficient wavelet-based coder to the continuous-tone planes.

However, this simple approach is sufficient to outperform standard and state-of-art single coders. We performed tests also for the JPEG baseline coder[5] and SPIHT[11], a wavelet-based coder.

For the example of Fig. 8, Figures 9 and 10 show enlarged portions of the images resulting from compression using different coders at the same compression ratio (70:1). Fig. 9 shows details of the text regions. MRC clearly outperforms the single coders in distortion. Fig. 10 shows the pictorial regions. Note that the distortion is also smaller for this region as compared to SPIHT, for example, although MRC used Baseline JPEG to encode the same region, which is inferior to SPIHT. The reason for that is that by saving bits in text regions there were more bits available to encode the pictorial regions. Hence, the MRC outperformed the single coders in all regions, not only text, while using only inexpensive, standard compression algorithms.

Due to space limitations and the large size of the images, hardcopies of further results will be presented at the conference site. Results will be shown clearly demonstrating MRC’s better performance for compressing several mixed documents, both scanned and computer generated.

6.Conclusions

While the basic MRC decomposition of a compound image starts with a little more than twice the data compared to the original, MRC gives higher compression with the same or better quality than does compressing the original with a single compression algorithm. This is due to the interplay of several factors, however it is mainly due to the fact that coding for each plane can be operated at more favorable R-D curve than it can for the single coder for the whole image. Of course this condition can depend on the image content (compound documents with text/graphics regions), the decomposition approach and the choice of encoders and image processing operations. Actually, the foreground and background layers can have lower spatial resolution than the original; only the layers or the portions of the layers with valid image data are compressed; and more aggressive compression can be applied to the foreground and background layers, independent of any text content, which is compressed with an algorithm designed for text.

Comparisons with single layer coders were carried out and results were presented showing the clear advantage of MRC for compound images. MRC decomposition was discussed in light of RD trade-offs, concluding that with certain approximations and assumptions, the bits saved to encode the transitions have to be sufficient to offset the bits spent on redundant and mask data.

MRC has been proposed and accepted for several standards. The reason for its success is its versatility and the fact that as an imaging model it does not interfere with the coding mechanism, nor with the overall architecture. If, for some reason, it is seen that MRC multi-layer decomposition is not appropriate for a particular image, it is necessary to only chose a single layer and encode it with a proper compressor. Thus, no overhead is incurred. Another advantage of MRC is that it can unify different standard compression mechanisms such as JBIG-2 and JPEG-2000 into one framework, while allowing compatibility with older standards such as JBIG, MMR and JPEG.

MRC-based coders are being developed at the present and MRC flavors are being devised for multiple applications. Examples are XIF[12], DejaVu[13], and check compression algorithms[14].

7.References

1. ITU-T Recommendation T.4, Standardization of group 3 facsimile apparatus for document transmission, July 1996.
2. ITU-T Recommendation T.6, Facsimile coding schemes and coding control functions for group 4 facsimile apparatus, November 1988.
3. ITU-T Recommendation T.82, Information technology - Coded representation of picture and audio information - Progressive bi-level image compression, March 1995. Also ITU-T Recommendation T.85, Application profile for Recommendation T.82 - Progressive bi-level image compression (JBIG coding scheme) for facsimile apparatus, August 1995.
4. JBIG2 Working Draft WD14492, ISO/IEC JTC1/SC29 JBIG Committee, 21 August 1998, http://www.jpeg.org/public/jbigpt2.htm
5. W. P. Pennebaker and J. L. Mitchell, *JPEG: Still Image Compression Standard*, Van Nostrand-Reinhold, New York, NY, 1993.
6. ISO/IEC JTC1/SC29/WG1 N505, "Call for contributions for JPEG 2000 (JTC 1.29.14, 15444): Image Coding System," March 21, 1997.
7. Draft Recommendation T.44, Mixed Raster Content (MRC), ITU-T Study Group 8, Question 5, May 1997.
8. R. Buckley, D. Venable and L. McIntyre, "New developments in color facsimile and internet fax,"
9. IETF RFC 2301. File Format for Internet Fax. L. McIntyre, S. Zilles, R. Buckley, D. Venable, G. Parsons, J. Rafferty. March 1998. ftp://ftp.isi.edu/in-notes/rfc2301.txt.
10. ISO/IEC JTC1/SC29/WG1 N542, "ITU-T Mixed Raster Content model as JPEG 2000 Architectural Framework," June 30, 1997.
11. A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and System for Video Technology*, vol. 6, pp. 243-250, June 1996.
12. XIFF - eXtended Image File Format, Version 2.0, Xerox ScanSoft, http://www.xerox.com/scansoft/xiff/xiff20.htm
13. L. Bottou, P. Haffner, P. Howard, P. Simard, Y. Bengio and Y. LeCun, "High quality document image compression using DjVu," *Journal of Electronic Imaging* , 7(3), pp. 410-425, July 1998.
14. J. Huang, Y. Wang and E. Wong, "Check image compression using a layered coding method," *Journal of Electronic Imaging* , 7(3), pp. 426-442, July 1998.

January 31, 2001

Dear Mom and Dad,

How are both of you doing? I thought I would drop a line to say hi. Fanny, little Danny, and I are doing well. As you can see by the picture, little Danny isn't quite so little! Isn't this letter really great! I took a picture of Danny that was on a Kodak PhotoCD, and I merged it onto this letter using my computer. I then printed the letter using a color inkjet printer I just bought...

Danny's wearing the gorgeous BLUE sweater you gave him last time you were visiting. It just brings out the RED in his lips and cheeks. He definitely gets his good looks from his mother!

Take care of yourselves and write soon.

Love,

Michael

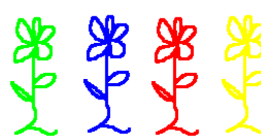

Figure 8 – Example color document image of typical color compound document. This image has 1400x1024 pixels originally in linear RGB.

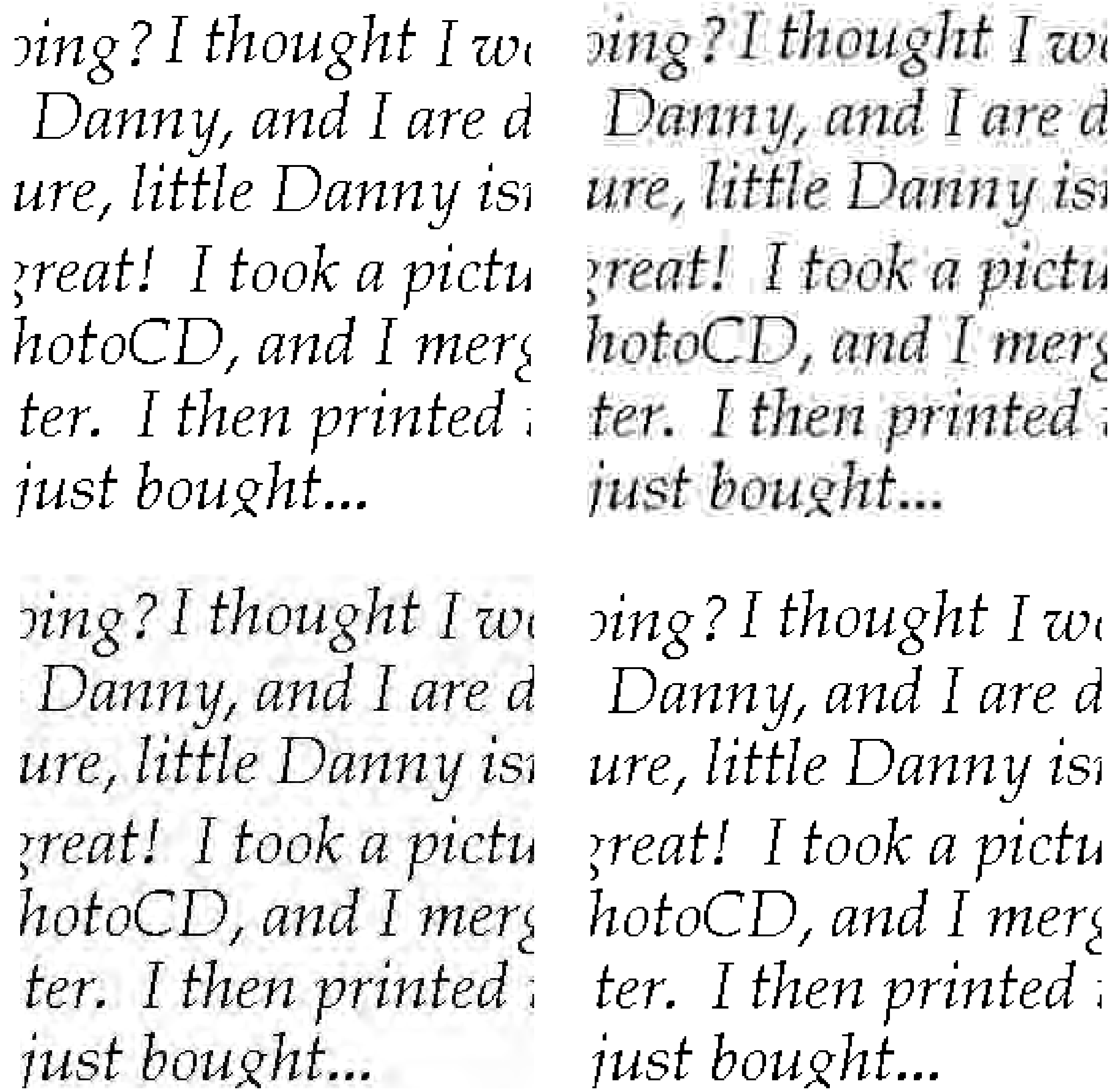

Figure 9 – Enlarged portion of the text region of original and reconstructed images at a compression ratio of 70:1. Top left: original; top right: JPEG baseline; bottom left: SPIHT wavelet-based coder; bottom-right: MRC using JPEG+MMR+JPEG without plane subsampling. Portion size is 256x256 pixels.

Figure10 – Enlarged portion of the pictorial region of original and reconstructed images at a compression ratio of 70:1. Top left: original; top right: JPEG baseline; bottom left: SPIHT wavelet-based coder; bottom-right: MRC using JPEG+MMR+JPEG without plane subsampling. Portion size is 256x256 pixels.