

# MRC Compression of Compound Documents Using Threshold Segmentation, Iterative Data-filling and H.264/AVC-INTRA

A. Zaghetto, R. L. de Queiroz\*  
 Universidade de Brasília, ENE  
 Brasília, DF, Brasil  
 zaghetto@image.unb.br, queiroz@ieee.org

D. Mukherjee  
 Hewlett Packard Labs  
 Palo Alto, CA, USA  
 debargha.mukherjee@hpl.hp.com

## Abstract

The Mixed Raster Content (MRC) ITU document compression standard (T.44) specifies a multi-layer representation of a compound document. It is expected that higher compression can be achieved if more efficient compression standards are used to compress each layer. In this paper we present an MRC compound document codec that uses H.264/AVC operating in INTRA mode to encode Background/Foreground layers and JBIG2 to encode the binary Mask layer. We also present a threshold-based layer segmentation algorithm and an iterative data-filling algorithm for redundant regions. The result is an unrivaled performance for compressing compound documents as demonstrated by our experiments.

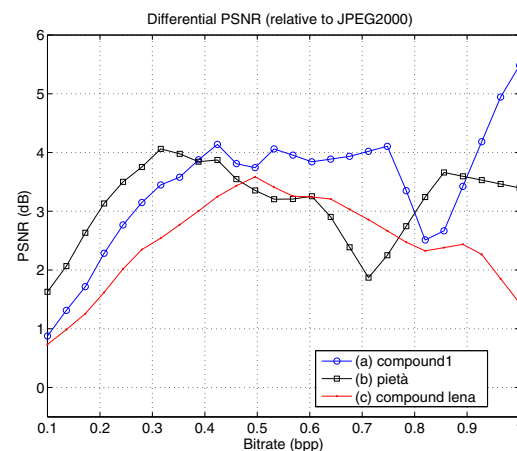
## 1. Introduction

The Mixed Raster Content (MRC) ITU document compression standard (T.44) [1]-[3] specifies a multi-layer representation of a compound document. In this paper we present a basic 3-layer MRC codec that uses the H.264/AVC [4] operating in INTRA mode to encode Background/Foreground layers and JBIG2 [5] to encode the binary Mask layer. The main objective is to show that MRC coding based on H.264/AVC and JBIG2, combined with appropriate layer segmentation and data-filling procedures, can yield better compression rates than schemes that use other state-of-the-art still image coders.

### 1.1 H.264/AVC-INTRA

The newest video coding standard, the H.264/AVC [4], has been well explained in the

\*This work was supported by HP Brasil and by CNPq under grant 47.3696/2007-0.



**Figure 1. Differential PSNR (relative to JPEG2000) plots comparing AVC-I against JPEG2000 for images shown in (a) Fig. 2 (compound1); (b) Fig. 8 (a) (pietà); and (c) Fig. 8 (b) (compound lena). PSNR gains surpass the mark of 4 dB in some cases.**

literature [6]-[11]. Many papers have illustrated its performance showing comparative results against coders such as MPEG-2. All results point to at least a factor of two improvement over previous standards. The reasons why the AVC is so good are many small improvements over previous methods. Each improvement brings a small coding gain, adding up to significant gains. Apart from the factor-of-two improvement over other standards, there are a few unexpected advantages that come with the AVC package.

H.264/AVC is a video compression standard and it was not conceived to be applied as a still image compression tool. Nevertheless, the many coding ad-

vances brought into H.264/AVC, not only set a new benchmark for video compression, but they also make it a formidable compressor for still images [12], [13]. One of the components of these advances is the intra-frame macroblock prediction method, which, combined with the context-adaptive binary arithmetic coding (CABAC), turns the H.264/AVC into a powerful still image compression engine. If we set our H.264/AVC implementation to work on a sole INTRA frame it will behave as a still image compressor. We refer to this coder as AVC-I. The big surprise is that it also outperforms previous state-of-art coders such as JPEG2000 [14]. This is a surprise to many because it was not meant to be an image coder at all. However, results are consistent and unison. Gains of the AVC-I over JPEG2000 are typically in the order of 0.25dB to 0.5dB in PSNR (peak signal-to-noise ratio) for pictorial images [12], [13], [15].

For compound images (mix of text, line-art and picture) the PSNR gains are more substantial, even surpassing the mark of 4 dB improvement in some cases, as shown in Fig. 1.

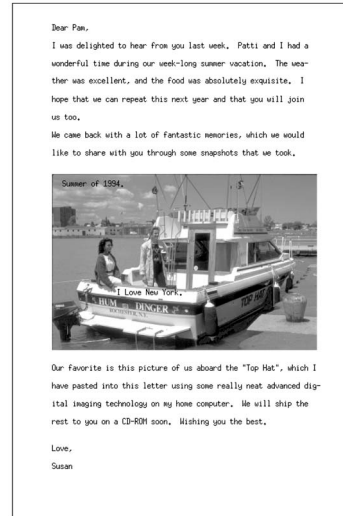
Without a doubt the AVC-I sets a new level of performance that will likely influence future still image coding standards.

## 1.2 JBIG2

JBIG2 [5] is an international standard for lossy and lossless compression of bi-level images developed by the Joint Bi-level Image Experts Group. Published in 2000 as the ITU-T Recommendation T.88, and in 2001 as ISO/IEC Standard 14492, JBIG2 outperforms JBIG1 [16] and Group 4/ MMR [17] by at least a factor of 2 and 3, respectively [18]. Typically, a JBIG2 coder decomposes a bi-level image into regions (text, halftone and generic) and encodes each region using a different method.

## 1.3 Compound Documents

Electronic documents are basically represented in two forms: vectorial or raster [3]. It is not much of a challenge to compress vectorized documents since each object can be compressed individually and the whole file can be compressed losslessly. The real challenge is to compress rasterized documents. Compound documents are assumed here as raster documents which contain a mix of text, line-art and picture. Compression algorithms are developed with a particular image type, characteristic and application in mind and no single algorithm is best across all types of images or applications. When compressing text and line-art, it is im-

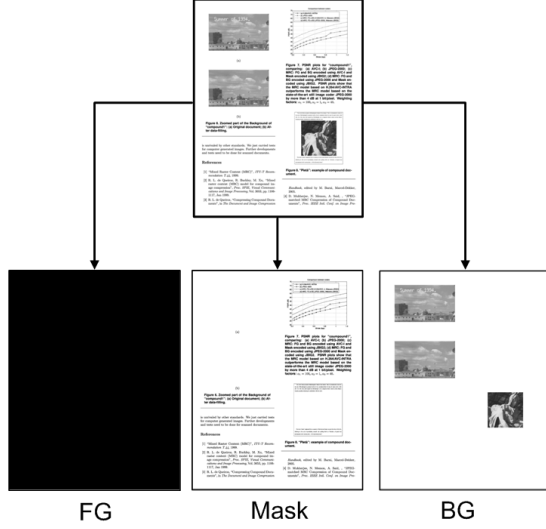


**Figure 2. "compound1": example of a compound document, assumed here as a raster document which contain a mix of text and pictorial contents.**

portant to preserve the edges and shapes of characters accurately to facilitate reading. The human visual system, however, works differently for typical continuous-tone images, better masking high frequency errors [19]. Figure 2 shows an example of a compound document.

## 1.4 Mixed Raster Content

Compound raster documents have typically been compressed as a single image. However, different compression algorithms may be applied to each of the regions of the document. That is the way multiple-coder based algorithms work. The basic 3-layer MRC model represents an image as two image layers (Foreground or FG and Background or BG) and a binary image layer (Mask or M), which determines if a pixel belongs to BG or FG [1]-[3]. Figure 3 illustrates the described model. Once the original single-resolution image is decomposed into layers, each layer can be processed and compressed using different algorithms. Background and Foreground processing operations can include a resolution change and a data-filling procedure. The compression algorithm used for a given layer would be matched to the layer's content, allowing for improved compression while reducing distortion visibility [20]-[22]. The compressed layers are then packed and delivered to the decoder. At the decoder, each plane is retrieved, decompressed, processed and the image is composed using the MRC imaging model.



**Figure 3. Illustration of MRC imaging model. The original document is represented using 3 layers: Foreground (FG), Background (BG) and Mask.**

## 2. Segmentation

The first step of MRC compression is the layer segmentation algorithm [3]. This paper uses a variation of the block-thresholding segmentation algorithm proposed by De Queiroz [23], which will be described next.

As the FG/BG planes will be encoded by macroblocks (16x16 pixels block), we want to find each macroblock mask  $m_n(i, j)$ . Because line-art/text-edge areas should have histograms which are bimodal, one simple approach is to find the bimodal macroblocks and to cluster the pixels around each of its modes. Whatever method is used to perform clustering or test bimodality, the pixels will be divided by some sort of threshold. In macroblock thresholding the mask is found as:

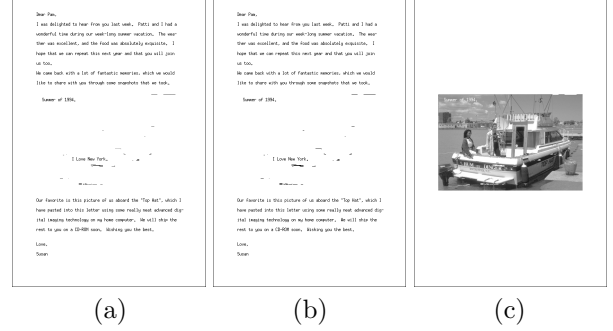
$$m_n(i, j) = u(t_n - x_n(i, j) - 1), \quad (1)$$

where  $t_n$  is the block's threshold,  $x_n(i, j)$  represents the original image macroblock and  $u(k)$  is the discrete step function (equals 1 for  $k \geq 0$  and 0 otherwise).

In a macroblock there are 256 pixels and therefore up to 256 thresholds. For each macroblock, we select a set of  $n \leq 256$  sorted thresholds,  $t_n(k)$ , and seek to minimize the following cost function:

$$J_n = \alpha_1 V_{BG} + \alpha_2 V_{FG} + \alpha_3 Nt, \quad (2)$$

where  $\alpha_i$  are weighting factors,  $V_{BG}$  and  $V_{FG}$  are



**Figure 4. Layer decomposition of “compound1” before data-filling procedure: (a) Foreground; (b) Mask; and (c) Background.**

the variances of pixels in the BG and FG layer macroblocks, respectively.  $N_t$  is the number of horizontal transitions of the mask block (the first column of the current block uses as reference the last column of the previous block). For a given threshold, a mask macroblock  $m_n(i, j)$  is obtained and we define two sets:

$$X_{FG} \equiv \{x_n(i, j) | m_n(i, j) = 0\} \quad (3)$$

$$X_{BG} \equiv \{x_n(i, j) | m_n(i, j) = 1\}.$$

We define  $n_{FG}$  and  $n_{BG}$  as the number of pixels in the set  $X_{FG}$  and  $X_{BG}$ , respectively, where obviously  $n_{FG} + n_{BG} = 256$  and, then, variances are computed as:

$$V_{FG} = \frac{\sum x_n(i, j)^2}{n_{FG}} - \left( \frac{\sum x_n(i, j)}{n_{FG}} \right)^2 \quad (4)$$

$$V_{BG} = \frac{\sum x_n(i, j)^2}{n_{BG}} - \left( \frac{\sum x_n(i, j)}{n_{BG}} \right)^2$$

which can be efficiently implemented.

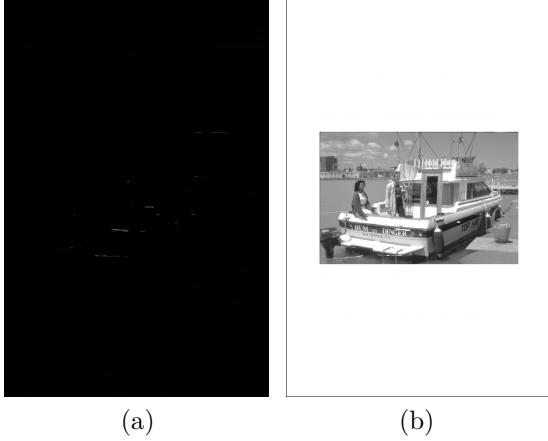
As for the weights, without loss of generality we can normalize one of them (e.g.,  $\alpha_2 = 1$ ). The choice of the other two weights is empirical.

Figure. 4 shows the layer decomposition of “compound1” before the data-filling procedure.

## 3. Data-filling

Once the image is segmented there will be “don’t care” regions on BG and FG 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 [3], [24], [25]. There are many methods for the replacement (data-filling). This paper

uses the iterative wavelet-based plane filling [26], which we describe next.



**Figure 5. (a) Foreground and (b) Background of “compound1” after data-filling procedure.**

Let  $F$  and  $B$  represent the pixel positions where the  $M$  indicates FG or BG respectively. First, we compute averages as:

$$\begin{aligned} m_{BG} &= \text{mean}(x(i, j) | (i, j) \in B) \\ m_{FG} &= \text{mean}(x(i, j) | (i, j) \in F), \end{aligned} \quad (5)$$

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

Let  $I_0$  be the starting FG plane with “don’t care” pixels replaced by  $m_{FG}$ . Also, let  $\tilde{I}_n$  be the compressed and decompressed version of  $I_n$  using a given coder at a target bit rate. If we plan to use a wavelet coder,  $\tilde{I}_n$  can be approximated as:

$$\tilde{I}_n = W^{-1}(\text{round}(W(I_n)/Q) * Q), \quad (6)$$

where  $W$  denotes the Wavelet Transform of 5 or 6 levels,  $\text{round}(\cdot)$  is a rounding operator, and  $Q$  is a step size to quantize the wavelet coefficients. It is expected to use quite large  $Q$  numbers, like those that would yield very high compression ratios. Then, for  $n = 0$  until  $n = \nu$ , where  $\nu$  limits the number of cycles to the maximum of let us say 3 or 4, we compute

$$I_{n+1}(i, j) = \begin{cases} I_n(i, j) & : (i, j) \in F \\ \tilde{I}_n(i, j) & : (i, j) \in B. \end{cases} \quad (7)$$

We stop the loop either after  $\nu$  cycles or when

$$\text{mean}(|I_n(i, j) - I_{n-1}(i, j)|) < \xi, \quad (i, j) \in B, \quad (8)$$

where  $\xi$  is some tolerance number, i.e. it stops when the filling in the “don’t care” region converges. The same



(a)



(b)

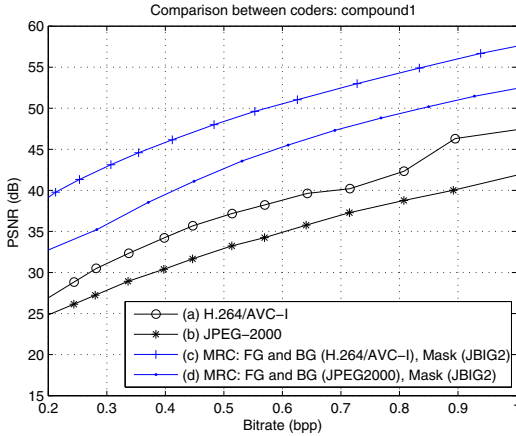
**Figure 6. Zoomed part of the Background of “compound1”: (a) Original document; (b) After data-filling.**

process applies to the BG plane, replacing foreground by background notation and vice-versa. Figure 5 shows BG and FG of “compound1” processed using the above method. Figure. 6 shows a zoomed part of the BG, where the effect of the data-filling algorithm can be observed more clearly.

## 4. Results

The Image “compound1” was encoded using AVC-I, JPEG2000 and using the proposed MRC model. In MRC compression, the Mask was encoded using JBIG2. BG and FG were both encoded using JPEG2000 and AVC-I. PSNR plots are shown in Fig. 7. The documents shown in Figs. 8 (a)-(f) were also compressed using the same encoders. PSNR plots are shown in Figs. 9 (a)-(f).

AVC-I single-coder seems to have an extra capacity of adapting itself to heterodox content [15]. For the compound documents shown in Fig. 2 and Figs. 8 (a) and (b), for example, the PSNR gains are substantial, even surpassing the mark of 4 dB improvement over JPEG2000 in some cases. In spite of this extra ca-



**Figure 7. PSNR plots for “compound1”, comparing: (a) AVC-I; (b) JPEG2000; (c) MRC: FG and BG encoded using AVC-I and Mask encoded using JBIG2; (d) MRC: FG and BG encoded using JPEG2000 and Mask encoded using JBIG2. Weighting factors:  $\alpha_1 = 100$ ,  $\alpha_2 = 1$ ,  $\alpha_3 = 40$ .**

capacity of AVC-I, the multiple-coder MRC model proposed here offers results that outperform the AVC-I single-coder approach, surpassing the mark of at least 4 dB improvement at 1 bit/pixel. PSNR plots shown in Fig. 7 and Figs. 9 (a)-(f) also demonstrate that the MRC model based on AVC-I outperforms the MRC model based on the state-of-the-art still image coder JPEG2000.

Figures 10 and 11 show parts of “mixed2” and “mixed4” documents, respectively, and their reconstructed JPEG2000, AVC-I, MRC (AVC-I/JBIG2) and MRC (JPEG2000/JBIG2) encoded versions, after compression. Notice how the proposed MRC scheme eliminates the ringing effect around text and line-art, while picture regions have their subjective quality improved.

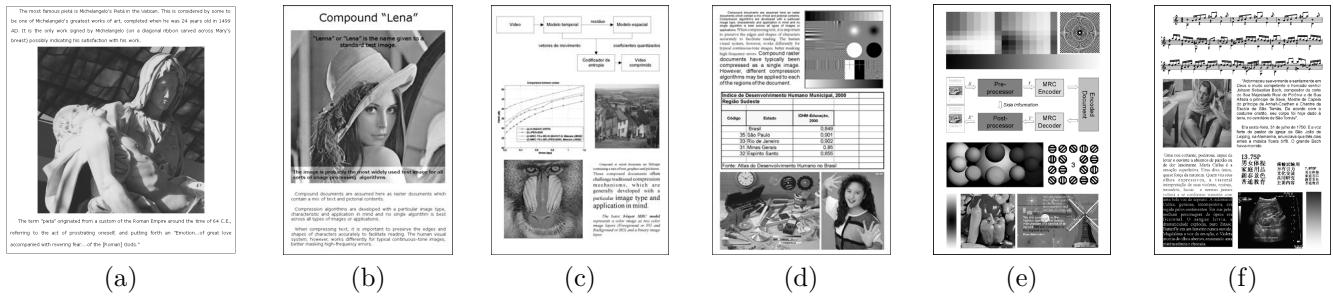
## 5. Conclusions

In this paper, we presented a basic 3-layer MRC codec that uses the H.264/AVC operating in INTRA mode to encode BG/FG layers and JBIG2 to encode the binary Mask layer. We also present a threshold-based layer segmentation algorithm and an iterative data-filling algorithm. Results show that in most cases the MRC model achieves better performance than single coder approaches, such as JPEG2000 and AVC-I. Furthermore, using AVC-I to compress BG and FG yields better results than schemes based on JPEG2000.

Without a doubt MRC schemes based on AVC-I set a new level of performance that is unrivaled by other standards. We just carried tests for electronic computer generated documents. Further developments and tests need to be done for scanned documents.

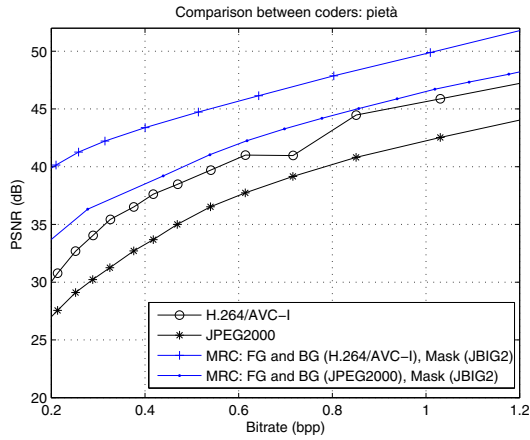
## References

- [1] “Mixed Raster Content (MRC)”, *ITU-T Recommendation T.44*, 1999.
- [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. L. de Queiroz, “Compressing Compound Documents”, in *The Document and Image Compression Handbook*, edited by M. Barni, Marcel-Dekker, 2005.
- [4] “Final Draft International Standard of Joint Video Specification”, *ITU-T Recommendation H.264, ISO/IEC 14496-10 AVC*, March 2003.
- [5] “Information technology - Coded representation of picture and audio information - Lossy/lossless coding of bi-level images”, *ITU-T Recommendation T.88*, March 2000.
- [6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, Wiley, 2003.
- [7] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, “Overview of the H.264/AVC video coding standard”, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 560-576, July 2003.
- [8] T. Stockhammer, M. M. Hannuksela and T. Wiegand, “H.264/AVC in wireless environments”, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 657-673, July 2003.
- [9] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini and G. J. Sullivan, “Rate-constrained coder control and comparison of video coding standards”, *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 688-703, July 2003.
- [10] G. J. Sullivan, P. Topiwala and A. Luthra, “The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions”, *Proc. of SPIE Conference on Applications of Digital Image Processing XXVII*, Special Session on Advances in the New Emerging Standard: H.264/AVC, August, 2004.

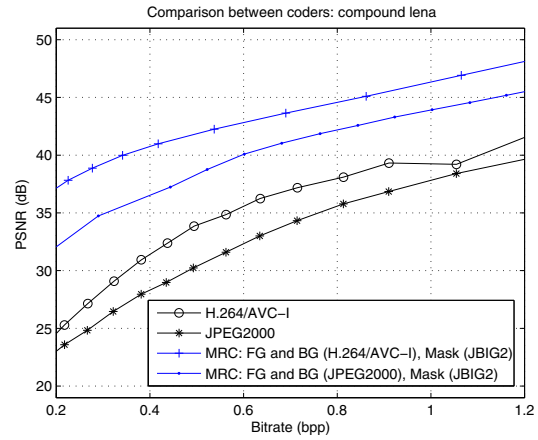


**Figure 8. Examples of compound documents: (a) “pietà”; (b) “compound lena”; (c) “mixed1”; (d) “mixed2”; (e) “mixed3”; and (f) “mixed4”.**

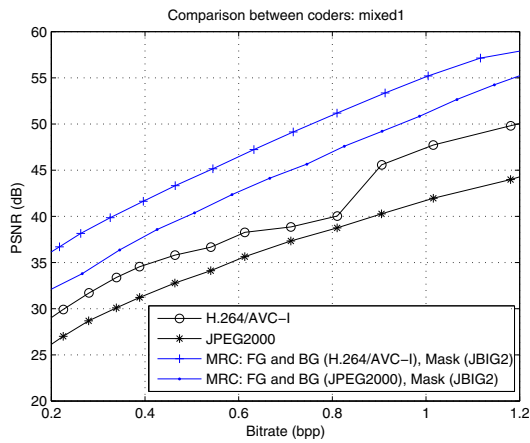
- [11] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer and T. Wedi, “Video coding with H.264/AVC: tools, performance, and complexity”, *IEEE Circuits and Systems Magazine*, Vol. 4, No. 1, pp. 7-28, March 2004.
- [12] D. Marpe, V. George and T. Wiegand, “Performance comparison of intra-only H.264/AVC and JPEG2000 for a set of monochrome ISO/IEC test images”, *Contribution JVT ISO/IEC MPEG and ITU-T VCEG, JVT M-014*, Oct. 2004.
- [13] D. Marpe, V. George, H. L. Cycon and K. U. Barthel, “Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode”, *Wavelet Applications in Industrial Processing*, in *Proc. of SPIE*, Vol. 5266, pp. 129-137, 2004.
- [14] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*, Kluwer Academic, 2002.
- [15] R. L. de Queiroz, R. S. Ortis, A. Zaghetto and T. A. Fonseca, “Fringe benefits of the H.264/AVC”, *Proc. of Intl. Telecom. Symp.*, Fortaleza, Brazil, pp. 208-212, Sep. 2006.
- [16] “Information technology - Coded representation of picture and audio information - Progressive bi-level image compression”, *ITU-T Recommendation T.82*, March 1995.
- [17] “Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus”, *ITU-T Recommendation T.6*, Nov. 1988.
- [18] “JBIG Maui Meeting Press Release: New document compression standard quadruples compression of today’s fax standards and runs at unprecedented speeds”, *ISO/IEC JTC1/SC29/WG1*, Dec. 1999.
- [19] D. Marr, *Vision*, San Francisco, CA: Freeman, 1982.
- [20] D. Mukherjee, N. Memon and A. Said, “JPEG-matched MRC Compression of Compound Documents”, *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Vol. 3, Thessaloniki, Greece, pp. 434-437, Oct. 2001.
- [21] D. Mukherjee, C. Chrysafis and A. Said, “JPEG2000-matched MRC compression of compound documents”, *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Vol. 3, pp. 73-76, 2002.
- [22] G. Feng and C.A. Bouman, “High-quality MRC document coding”, *IEEE Trans. on Image Processing*, Vol. 15, No. 10, pp. 3152-3169, Oct. 2006.
- [23] R. L. de Queiroz, Z. Fan and T. D. Tran, “Optimizing block-thresholding segmentation for multilayer compression of compound images”, *IEEE Trans. on Image Processing*, Vol. 9, No. 9, pp. 1461-1471, Sep. 2000.
- [24] R. L. de Queiroz, “On data-filling algorithms for MRC layers”, *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Vancouver, Canada, Vol. II, pp. 586-589, Sep. 2000.
- [25] G. Lakhani and R. Subedi, “Optimal filling of FG/BG layers of compound document images”, *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Atlanta, USA, pp. 2273-2276, Oct. 2006.
- [26] R. L. de Queiroz, “Pre-processing for MRC layers of scanned images”, *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, Atlanta, USA, pp. 3093-3096, Oct. 2006.



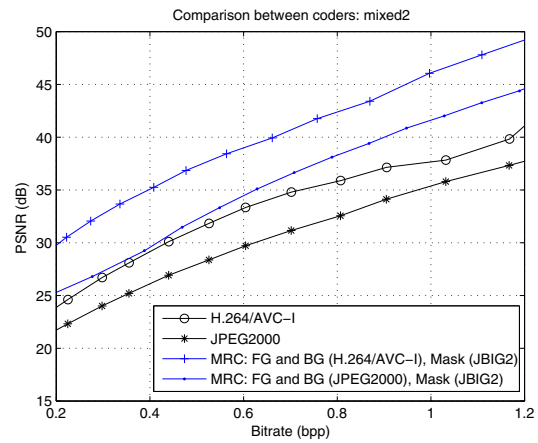
(a)



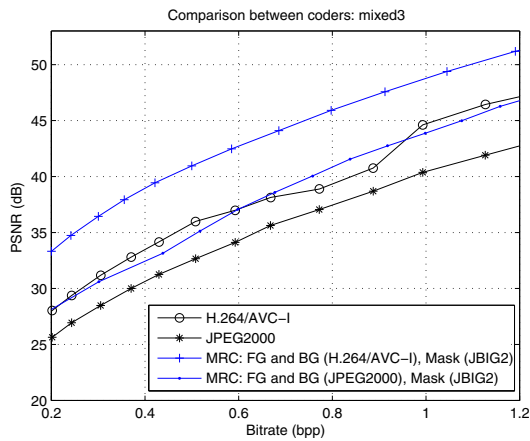
(b)



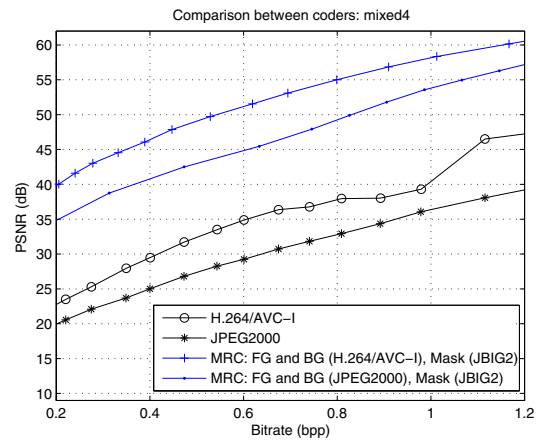
(c)



(d)



(e)



(f)

**Figure 9. PSNR plots for: (a) “pietà” (weighting factors:  $\alpha_1 = 500, \alpha_2 = 1, \alpha_3 = 40$ ); (b) “compound lena” (weighting factors:  $\alpha_1 = 300, \alpha_2 = 1, \alpha_3 = 40$ ); (c) “mixed1” (weighting factors:  $\alpha_1 = 500, \alpha_2 = 1, \alpha_3 = 100$ ); (d) “mixed2” (weighting factors:  $\alpha_1 = 500, \alpha_2 = 1, \alpha_3 = 100$ ); (e) “mixed3” (weighting factors:  $\alpha_1 = 500, \alpha_2 = 1, \alpha_3 = 40$ ); and (f) “mixed4” (weighting factors:  $\alpha_1 = 500, \alpha_2 = 1, \alpha_3 = 40$ ).**



(a)



(b)



(c)



(d)



(e)

**Figure 10. Subjective performance comparison between coders. Part of “mixed2” document: (a) original; (b) AVC-I; (c) JPEG2000 (d) MRC (AVC-I/JBIG2); and (e) MRC (JPEG2000/JBIG2). Bitrate is 0.3 bpp.**



(a)



(b)



(c)



(d)



(e)

**Figure 11. Subjective performance comparison between coders. Part of “mixed4” document: (a) original; (b) AVC-I; (c) JPEG2000 (d) MRC (AVC-I/JBIG2); and (e) MRC (JPEG2000/JBIG2). Bitrate is 0.2 bpp.**