

# Segmentação por Limiarização de Macroblocos e Data-filling Iterativo para Compressão de Documentos Compostos Utilizando o H.264

Alexandre Zaghetto e Ricardo L. de Queiroz

**Resumo**—O padrão MRC (*Mixed Raster Content*) para compressão de documentos (ITU-T.44) especifica uma representação multicamada para documentos compostos. Espera-se que melhores taxas de compressão possam ser alcançadas quando algoritmos de pré-processamento e padrões de compressão mais eficientes são utilizados. O presente artigo propõe a utilização de um algoritmo de segmentação por limiarização de macroblocos para decompor o documento original em três camadas (*Foreground*, *Background* e máscara). O preenchimento das regiões “irrelevantes” (*data-filling*) é realizado por meio de uma técnica iterativa baseada em wavelets. As camadas *Background/Foreground* são comprimidas com o H.264/AVC operando no modo INTRA e a máscara binária é codificada com o JBIG2. O resultado é um desempenho sem igual, como demonstrado pelos experimentos.

**Palavras-Chave**—H.264/AVC, JPEG-2000, Segmentação de página, *Data-filling*, Compressão de documentos, *Mixed Raster Content*.

**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 pre-processing and compression algorithms are used. In this paper we propose using a macroblock-based segmentation algorithm to decompose the original document into three layers (*Foreground*, *Background* and mask). The proposed framework also uses an iterative wavelet-based data-filling algorithm. Background and Foreground layers are compressed using H.264/AVC operating in INTRA mode. JBIG2 is used to encode the binary mask. The result is an unrivaled performance for compressing compound documents as demonstrated by our experiments.

**Keywords**—H.264/AVC, JPEG-2000, Page segmentation, *Data-filling*, Document compression, *Mixed Raster Content*.

## I. INTRODUÇÃO

O padrão MRC (*Mixed Raster Content*) para compressão de documentos (ITU-T.44) [1] especifica uma representação multicamada para documentos compostos. O presente artigo propõe a utilização de um algoritmo de segmentação por limiarização de macroblocos como parte do codificador MRC de três camadas baseado nos padrões H.264/AVC [8] e JBIG2 [19], apresentado em trabalho anterior [2]. O objetivo principal é apenas mostrar que tal codificador, aliado a técnicas de segmentação de página e preenchimento de regiões “irrelevantes” apropriadas, pode atingir melhores taxas de compressão que esquemas baseados em codificadores considerados o estado da arte em compressão de imagens estáticas.

Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília, Brasília, DF. E-mails: zaghetto@image.unb.br, queiroz@ieee.org. Este trabalho foi parcialmente financiado pela HP Brasil.

### A. H.264/AVC-INTRA

O mais novo padrão de codificação de vídeo, o H.264/AVC [8], já foi amplamente descrito pela literatura [9]-[14]. Muitos artigos ilustraram seu desempenho, apresentando inúmeros resultados que o comparam a outros codificadores. Todos os resultados apontam para um aumento de desempenho de no mínimo o dobro em relação a outros padrões de compressão de vídeo. A razão que o torna tão eficiente são pequenas melhorias que individualmente contribuem modestamente para o aumento substancial do ganho geral. Além desse aumento de desempenho, o AVC apresenta algumas vantagens adicionais.

O H.264/AVC é um padrão de compressão de vídeo e não foi concebido para ser aplicado como uma ferramenta de compressão de imagens estáticas. Apesar disso, os vários avanços incorporados ao H.264/AVC não apenas estabelecem um novo patamar em compressão de vídeo, mas também fazem dele um excelente codificador de imagens estáticas [15], [16]. Um desses avanços é o método de predição INTRA, o qual, combinado com o codificador aritmético binário adaptativo ao contexto, transforma o H.264/AVC em um poderoso compressor de imagens estáticas. Se o H.264/AVC for ajustado para codificar apenas um único quadro de vídeo, ele se comporta como um codificador de imagens. Vamos nos referir a esse codificador como AVC-I. A grande surpresa é que o AVC-I supera codificadores de imagens considerados o estado da arte, como o JPEG-2000 [17]. A surpresa advém do fato de tal codificador não ter sido inicialmente concebido para essa finalidade específica. Os ganhos do AVC-I sobre o JPEG-2000 tipicamente são da ordem de 0.25dB a 0.5dB em PSNR (*peak signal-to-noise ratio*) para imagens pictoriais [15], [16], [18].

### B. JBIG2

O JBIG2 [19] é um padrão internacional para compressão de imagens binárias desenvolvido pelo *Joint Bi-level Image Experts Group*. Publicado em 2000 sob a denominação ITU-T *Recommendation T.88*, e em 2001 como ISO/IEC *Standard 14492*, o JBIG2 supera o JBIG1 [20] e o Group 4/ MMR [21] por um fator de 2 e 3, respectivamente [22]. Tipicamente, um codificador JBIG2 decompõe uma imagem binária em regiões (texto, halftone e genérica) e codifica cada região utilizando métodos diferentes.

### C. Documentos Compostos

Documentos eletrônicos são basicamente representados de duas maneiras: vetorial e *raster* [4]. Não constitui grande

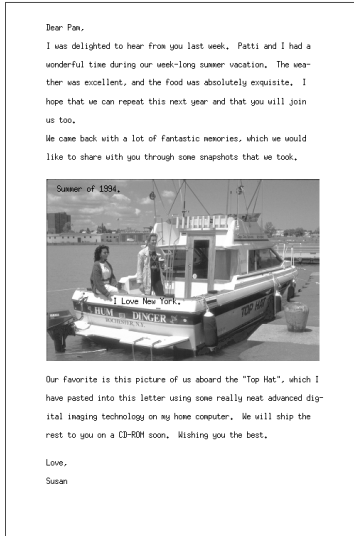


Fig. 1. "Compound1": exemplo de documento composto, considerado como sendo um documento representado na forma *raster* e que contém uma mistura de texto e imagem.

desafio comprimir documentos vetorizados, uma vez que cada objeto pode ser comprimido individualmente e o arquivo todo pode ser comprimido sem perdas. O real desafio é comprimir documentos na forma *raster*. Documentos compostos serão considerados aqueles que são representados na forma *raster* e que contêm uma mistura de texto e imagem.

Os algoritmos de compressão são desenvolvidos tendo-se em mente um tipo particular de imagem e aplicação. Nenhum algoritmo é o mais adequado para todos os tipos de imagens e aplicações. Quando se comprime texto, é importante preservar as bordas e os formatos dos caracteres para que a leitura seja facilitada. O sistema visual humano, por outro lado, trabalha diferentemente para imagens em tons contínuos, melhor mascarando erros de alta frequência [23].

A Fig. 1 mostra um exemplo de documento composto.

#### D. Mixed Raster Content

Documentos compostos são tipicamente codificados como uma imagem única. No entanto, diferentes algoritmos de compressão podem ser aplicados a cada região do documento. É assim que um algoritmo baseado em múltiplos codificadores funciona. O modelo MRC (*Mixed Raster Content*) básico representa um documento por meio de duas camadas de imagem (*Foreground* ou FG e *Background* ou BG) e uma máscara binária, que determina se um *pixel* pertence ao FG ou BG [1]-[7]. A Fig. 2 ilustra o modelo descrito. Após a decomposição do documento, cada camada pode ser processada e comprimida por meio de algoritmos específicos. O processamento do FG e do BG pode incluir mudança de resolução e *data-filling*. O algoritmo de compressão utilizado para uma determinada camada deve ser adequado ao seu conteúdo, possibilitando o aumento da compressão enquanto a distorção é reduzida. Após codificadas, as camadas são empacotadas e enviadas ao decodificador. No decodificador a imagem é recomposta por meio do processo inverso.

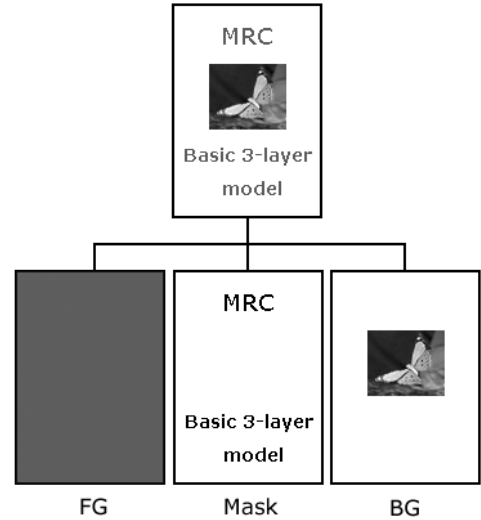


Fig. 2. Ilustração do modelo de imageamento MRC. O documento original é representado utilizando-se três camadas: *Foreground* (FG), *Background* (BG) e *Máscara*.

## II. SEGMENTAÇÃO

O primeiro passo da compressão MRC é a segmentação [4]. No presente artigo, consideramos o modelo MRC básico de três camadas, que utiliza uma representação baseada em FG, BG e máscara. O algoritmo de segmentação descrito a seguir é uma variação do algoritmo baseado em limiarização por blocos proposto por De Queiroz, Fan e Tran [24].

Como o FG e o BG serão codificados por macroblocos (blocos de  $16 \times 16$  *pixels*), desejamos encontrar uma máscara de macrobloco  $m_n(i, j)$ , para cada um dos  $n$  macroblocos do documento. Devido ao fato de áreas que contêm bordas de texto normalmente apresentarem histogramas bimodais, uma abordagem simplificada seria encontrar os macroblocos bimodais e agrupar seus *pixels* ao redor das duas modas. Independentemente do método utilizado para se testar a bimodalidade ou realizar o agrupamento, os *pixels* são divididos por algum tipo de limiarização.

Na limiarização por macrobloco, a máscara  $m_n(i, j)$  é dada por,

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

onde  $t_n$  é o limiar do bloco,  $x_n(i, j)$  representa o macrobloco original e  $u(k)$  é a função degrau discreta (igual a 1 para  $k \geq 0$  e 0 caso contrário).

Em um macrobloco há 256 *pixels* e, por isso, 256 é o número de possíveis limiares. Para cada macrobloco, seleciona-se um conjunto de  $l \leq 256$  limiares  $t_n$  organizados em ordem crescente e procura-se minimizar a seguinte função de custo,

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

onde  $\alpha_i$  são pesos, e  $V_{BG}$  e  $V_{FG}$  são as variâncias dos *pixels* dos macroblocos do BG e do FG, respectivamente.  $N_t$  é o número de transições horizontais de  $m_n(i, j)$  (a primeira coluna da máscara do macrobloco atual é utilizada como

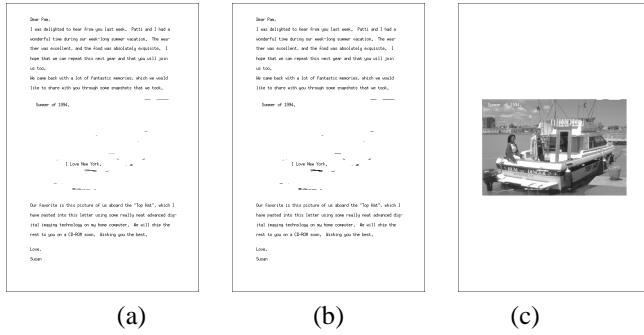


Fig. 3. Decomposição em camadas do documento “compound1” antes do procedimento de *data-filling*: (a) FG; (b) Máscara; (c) BG.

referência para a última coluna da máscara do macrobloco anterior). Valores de  $V_{BG}$  e  $V_{FG}$  mais elevados indicam uma menor compressão dos planos BG e FG, o que contribui para o aumento do custo. Em relação à máscara  $m_n(i, j)$ , quanto maior o número de transições  $N_t$ , menor a compressão, ou seja, maior o custo.

Para cada um dos  $l$  limiares  $t_n$ , obtém-se  $m_n(i, j)$ , e dois conjuntos são definidos,

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

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

Define-se  $n_{FG}$  e  $n_{BG}$  como o número de *pixels* dos conjuntos  $X_{FG}$  e  $X_{BG}$ , respectivamente. Obviamente,  $n_{FG} + n_{BG} = 256$ . As variâncias são calculadas pelas equações,

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

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

as quais podem ser implementadas de forma eficiente. Quanto aos pesos  $\alpha_i$ , sem perda de generalidade, pode-se normalizar um deles (p. ex.,  $\alpha_2 = 1$ ). A escolha dos outros dois pesos é realizada de forma empírica.

O limiar  $t_n$  que resultar no menor custo  $J$  é escolhido, a máscara  $m_n(i, j)$  é definida e a separação em camadas para o macrobloco é realizada. Tal procedimento é repetido para cada um dos  $n$  macroblocos do documento.

A Fig. 3 mostra a decomposição em camadas do documento “compound1”, antes do procedimento de *data-filling*.

### III. DATA-FILLING

Após a decomposição em camadas, *pixels* associados ao BG serão marcados como “irrelevantes” no FG. Da mesma forma, *pixels* associados ao FG serão marcados como “irrelevantes” no BG. Esses *pixels* podem ser substituídos por qualquer valor com o objetivo de se melhorar a compressão [4], [25], [26]. Existem vários métodos destinados a esse fim. O presente artigo utiliza um método de preenchimento iterativo baseado em wavelets [27], que será descrito a seguir.

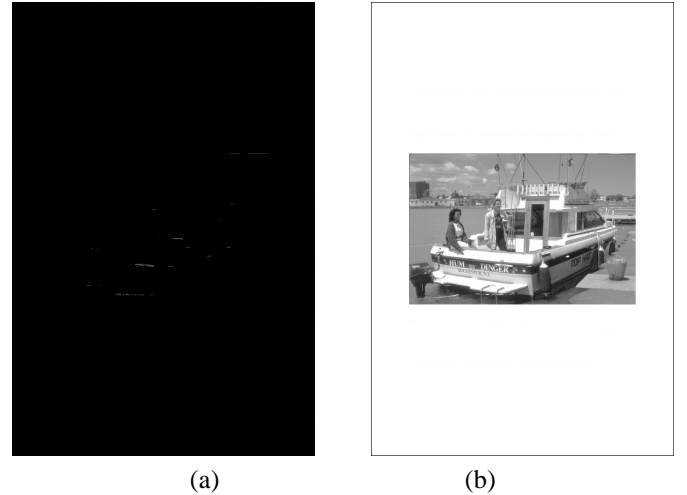


Fig. 4. (a) FG e (b) BG do documento “compound1” após o procedimento de *data-filling*.

Definiremos  $F$  e  $B$  como sendo as posições dos *pixels* onde a máscara do documento indica FG e BG, respectivamente. Inicialmente, calcula-se as médias,

$$m_{BG} = \text{mean}(x(i, j) | (i, j) \in B) \quad (5)$$

$$m_{FG} = \text{mean}(x(i, j) | (i, j) \in F),$$

onde  $x(i, j)$  representa a imagem original.

Seja  $I_0$  a camada FG inicial com os *pixels* “irrelevantes” substituídos por  $m_{FG}$ . Seja  $\tilde{I}_n$  a versão codificada e decodificada de  $I_n$ , utilizando-se um codificador a uma dada taxa de *bits* alvo. Se planejamos utilizar um codificador baseado em wavelets,  $\tilde{I}_n$  pode ser aproximado por,

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

onde  $W$  denota a Transformada Wavelet,  $\text{round}(\cdot)$  é o operador de arredondamento e  $Q$  é o passo de quatinização dos coeficientes da transformada. Então, para  $n = 0$  até  $n = \nu$ , onde  $\nu$  limita o número de ciclos em 3 ou 4, computa-se,

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

O algoritmo é interrompido após  $\nu$  ciclos ou quando,

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

onde  $\xi$  é um limiar de tolerância. Em outras palavras, o algoritmo é interrompido quando o preenchimento das regiões “irrelevantes” converge. O mesmo processo é aplicado ao BG, substituindo-se a notação do FG pela do BG, e vice-versa. A Fig. 4 mostra as camadas do documento “compound1” após o processo de *data-filling* descrito anteriormente.

A Fig. 5 mostra a ampliação de uma parte do BG, onde o efeito do algoritmo de *data-filling* pode ser observado em detalhes.



(a)



(b)

Fig. 5. Parte ampliada do BG do documento “compound1”: (a) Documento original; (b) Após o *data-filling*.

#### IV. RESULTADOS

O documento “compound1” foi comprimido com os codificadores AVC-I, JPEG-2000 e o modelo MRC proposto. Na compressão MRC, a máscara foi codificada com o JBIG2 e as camadas BG/FG foram ambas codificadas com o JPEG-2000 e com o AVC-I. A Fig. 6 mostra as curvas PSNR resultantes.

Os documentos da Fig. 7 e 9 também foram comprimidos com os mesmos codificadores. Suas respectivas curvas PSNR podem ser observadas nas Figs. 8 e 10 .

O AVC-I parece ter uma capacidade extra em se adaptar a conteúdos heterodoxos [18]. Para os documentos compostos mostrados nas Figs. 1 e 7, por exemplo, os ganhos em PSNR são substanciais, ultrapassando a marca de 4 dB em relação ao JPEG-2000 a 1 bit/pixel. Apesar dessa capacidade extra do AVC-I, o modelo MRC proposto, baseado em múltiplos codificadores, fornece resultados que ultrapassam a abordagem que utiliza um codificador único em mais de 4 dB a 1 bit/pixel.

As curvas PSNR das Figs. 6, 8 e 10 também mostram que o modelo MRC baseado em AVC-I supera o modelo MRC baseado em JPEG-2000, um codificador considerado o estado da arte em compressão de imagens estáticas

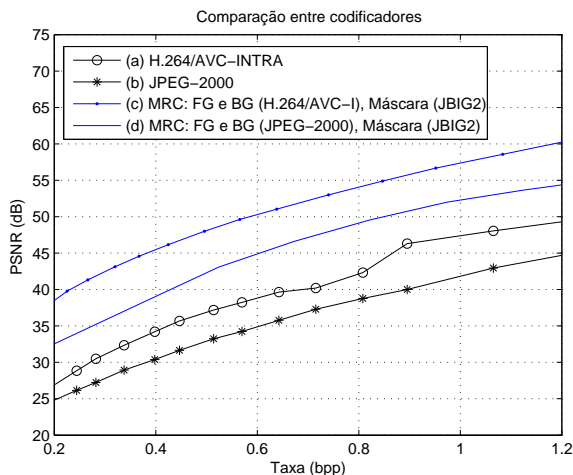


Fig. 6. Curvas PSNR para o documento “compound1”, comparando: (a) AVC-I; (b) JPEG-2000; (c) MRC: FG e BG codificados com AVC-I e Máscara codificada com JBIG2; (d) MRC: FG e BG codificados com JPEG-2000 e Máscara codificada com JBIG2. A curvas mostram que o modelo MRC baseado em H.264/AVC-I supera o modelo MRC baseado em JPEG-2000 em mais de 4 dB a 1 bit/pixel. Pesos considerados:  $\alpha_1 = 100$ ,  $\alpha_2 = 1$ ,  $\alpha_3 = 40$ .

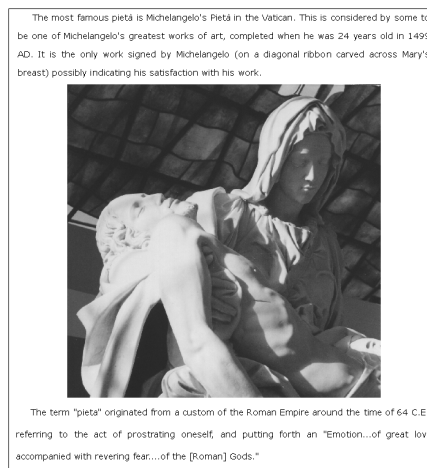


Fig. 7. “Pietà”: exemplo de documento composto.

#### V. CONCLUSÕES

No presente artigo apresentamos um codificador MRC que utiliza o H.264/AVC operando no modo INTRA para codificar as camadas BG/FG e o JBIG2 para codificar a máscara binária. Também propomos a utilização de um algoritmo de segmentação baseado em limiarização de macrobloco e um algoritmo de preenchimento de regiões “irrelevantes”. Os resultados mostram que com o modelo MRC é possível alcançar melhor desempenho do que as abordagens baseadas em um codificador único, como o JPEG-2000 e o AVC-I. Além disso, usar o AVC-I para codificar o BG e o FG em um codificador MRC resulta em melhor desempenho que esquemas que utilizam o JPEG-2000.

Sem dúvida, a proposta do modelo MRC baseado em AVC-I, aliado às técnicas de segmentação de página e *data-filling* apresentadas, estabelece um novo patamar para compressão de documentos compostos.

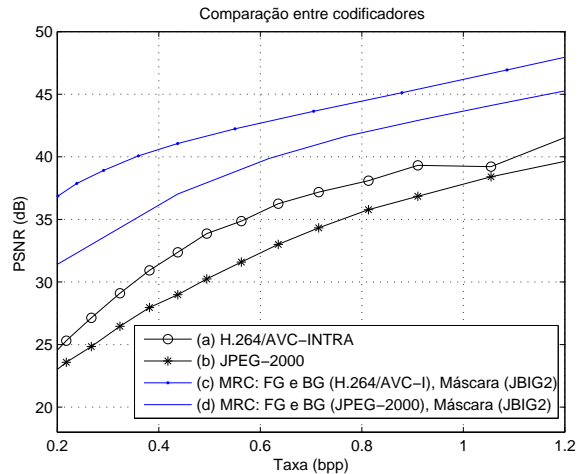
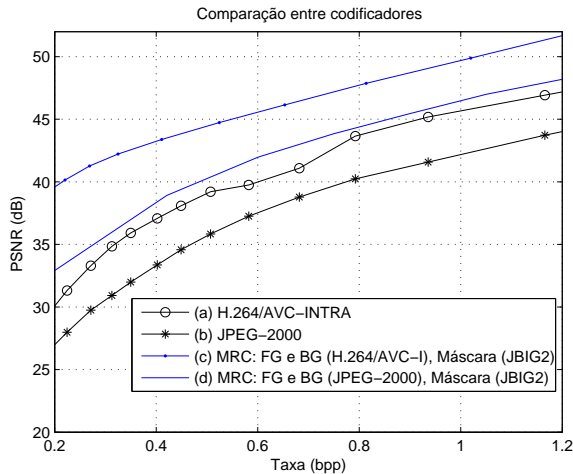


Fig. 8. Curvas PSNR para o documento "Pietà", comparando: (a) AVC-I; (b) JPEG-2000; (c) MRC: FG e BG codificados com AVC-I e Máscara codificada com JBIG2; (d) MRC: FG e BG codificados com JPEG-2000 e Máscara codificada com JBIG2. A curvas mostram que o modelo MRC baseado em H.264/AVC-I supera o modelo MRC baseado em JPEG-2000 em mais de 4 dB a 1 bit/pixel. Pesos considerados:  $\alpha_1 = 500$ ,  $\alpha_2 = 1$ ,  $\alpha_3 = 40$ .

Fig. 10. Curvas PSNR para o documento "Compound Lena", comparando: (a) AVC-I; (b) JPEG-2000; (c) MRC: FG e BG codificados com AVC-I e Máscara codificada com JBIG2; (d) MRC: FG e BG codificados com JPEG-2000 e Máscara codificada com JBIG2. A curvas mostram que o modelo MRC baseado em H.264/AVC-I supera o modelo MRC baseado em JPEG-2000 em mais de 3 dB a 1 bit/pixel. Pesos considerados:  $\alpha_1 = 300$ ,  $\alpha_2 = 1$ ,  $\alpha_3 = 40$ .



Fig. 9. "Compound Lena": exemplo de documento composto.

Apenas realizamos testes para imagens geradas por computador. Experimentos com imagens escaneadas estão sendo desenvolvidos.

REFERÊNCIAS

- [1] "Mixed Raster Content (MRC)", *ITU-T Recommendation T.44*, 1999.
- [2] A. Zaghetto and R. L. de Queiroz, "MRC compression of compound documents using H.264/AVC-I", *Simpósio Brasileiro de Telecomunicações*, Recife, PE, Brasil, Set. 2007.
- [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] R. L. de Queiroz, "Compressing Compound Documents", in *The Document and Image Compression Handbook*, edited by M. Barni, Marcel Dekker, 2005.
- [5] D. Mukherjee, N. Memon, 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.
- [6] D. Mukherjee, C. Chrysafis, A. Said, "JPEG2000-matched MRC compression of compound documents", *Proc. IEEE Intl. Conf. on Image Processing*, ICIP, Vol. 3, pp. 73-76, 2002.
- [7] 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.
- [8] "Final Draft International Standard of Joint Video Specification", *ITU-T Recommendation H.264, ISO/IEC 14496-10 AVC*, March 2003.
- [9] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, Wiley, 2003.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*, Kluwer Academic, 2002.
- [18] 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.
- [19] "Information technology - Coded representation of picture and audio information - Lossy/lossless coding of bi-level images", *ITU-T Recommendation T.88*, March 2000.
- [20] "Information technology - Coded representation of picture and audio

- information - Progressive bi-level image compression”, *ITU-T Recommendation T.82*, March 1995.
- [21] “Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus”, *ITU-T Recommendation T.6*, Nov. 1988.
- [22] “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.
- [23] D. Marr, *Vision*, San Francisco, CA: Freeman, 1982.
- [24] R. L. de Queiroz, Zhigang Fan, 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.
- [25] 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.
- [26] G. Lakhani, 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.
- [27] 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.