

Compressão de Documentos Digitalizados Baseada no Casamento de Padrões Utilizando o H.264/AVC

¹Alexandre Zaghetto, ²Ricardo L. de Queiroz

Resumo—O presente artigo apresenta um codificador híbrido para documentos digitalizados baseado no casamento aproximado de padrões e na codificação por transformada. O objetivo é aplicar tal codificador na geração de documentos codificados e reconstruídos de alta qualidade objetiva e subjetiva. Tal condição torna-se especialmente necessária quando documentos de valor históricos são comprimidos. A ideia é utilizar a estimação e a compensação de movimento propostas pelo padrão H.264/AVC como um algoritmo de casamento aproximado de padrões. De uma maneira geral, o documento é transformado em uma sequência de vídeo, posteriormente codificada pelo H.264/AVC. Nessas condições, quadros anteriormente codificados podem ser utilizados como referência (dicionário) na codificação do quadro atual. O algoritmo de estimação e compensação de movimento (casamento de padrões) gera um quadro residual que pode ser mais eficientemente comprimido por técnicas baseadas em transformação, quantização e codificação de entropia. Os resultados mostram que o método proposto supera objetiva e subjetivamente os codificadores JPEG2000 e o AVC-I.

Palavras-Chave—Compressão de documentos digitalizados, H.264/AVC, casamento aproximado de padrões.

Abstract—This paper presents a hybrid approximate pattern matching/transform-based scanned document encoder. The main objective is to generate high quality encoded and reconstructed versions of scanned documents. This is specially important when rare books of historical value must be digitally stored. The idea is to use motion estimation and motion compensation algorithms, proposed by H.264/AVC, as an approximate pattern matching engine. In summary, the document is transformed into a video sequence, which is further H.264/AVC encoded. In this scenario, the current frame may use previously encoded frames as reference frames (dictionary). The motion estimation and compensation (pattern matching) algorithms generate a residual frame that can be more efficiently compressed by transform-based, quantization and entropy coding techniques. Results show that the proposed method outperforms JPEG2000 and AVC-I.

Keywords—Scanned document compression, H.264/AVC, approximate pattern matching.

I. INTRODUÇÃO

Recentemente, um novo algoritmo universal para codificação de imagens, denominado *Multidimensional Multiscale Parser* (ou MMP), foi proposto [1], [2], [3], [4], [5]. Tal algoritmo não é fundamentado no tradicional esquema de codificação composto por transformação, quantização e codificação de entropia, como o JPEG2000 [6], [7] e o AVC-I (H.264/AVC operando exclusivamente no modo intra-quadro). Ao invés disso, a imagem é recursivamente segmentada em blocos de comprimento variável, sendo cada bloco codificado por meio

de expansões e contrações de elementos em um dicionário, esse atualizado a partir também de expansões e contrações de concatenações de blocos previamente codificados. O efeito observado é o de grande adaptação em relação aos dados de entrada, o que justifica, em parte, a eficiência do codificador na compressão de um conjunto bastante heterogêneo de imagens.

Inspirado no princípio do casamento aproximado de padrões, proposto pelo MMP, o presente artigo apresenta um método para codificação, em alta qualidade objetiva e subjetiva, de documentos digitalizados. Sinais dessa natureza apresentam uma repetição muito grande de padrões de texto, o que sugere o uso da estimação de movimento como uma forma de segmentação da imagem original em blocos de tamanhos variados, com o objetivo de se realizar o casamento aproximado de padrões. Nesse cenário, macroblocos (blocos de 16×16 pixels) anteriormente codificados podem ser utilizados como elementos de um dicionário na codificação dos macroblocos que ainda estão por codificar. A atualização desse dicionário, por sua vez, ocorre à medida em que novos macroblocos tornam-se disponíveis como referências para a predição inter-quadros. Além disso, a estimação em metade e quarto de pixel pode ser interpretada como uma forma de super-atualização do dicionário, uma vez que deslocamentos fracionários também passam a ser possíveis durante o processo de casamento de padrões.

De uma maneira geral, o método aqui apresentado consiste em transformar o documento digitalizado em uma sequência de vídeo, de modo a permitir que a estimação de movimento do H.264/AVC possa encontrar em partes do documento previamente codificadas algum padrão que seja semelhante àquele contido no macrobloco atual. A partir dessa busca, o codificador prossegue com a geração do resíduo de predição, a transformação, a quantização e a codificação de entropia. Isso faz com que o método proposto seja um híbrido entre o esquema de codificação tradicional e aquele que utiliza o casamento aproximado de padrões, proposto pelo MMP.

Por fim, assume-se que o usuário deseja comprimir o documento preservando-se as características que o aproximam ao máximo de sua versão impressa, condição necessária quando documentos antigos, de valor histórico, são comprimidos. Dessa forma, não serão considerados, para efeito de comparação, algoritmos de compressão de imagens binárias, mas sim algoritmos de compressão de imagens em tons contínuos. Além disso, deseja-se evitar o uso de múltiplos codificadores como no caso de abordagens baseadas no modelo de imageamento MRC (*Mixed Raster Content*) [8], [9], [10], [11], [12].

A seguir serão tratadas alguma das principais características do padrão H.264/AVC.

¹Departamento de Ciência da Computação, ²Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, Brasil, E-mails: alexandre@cic.unb.br, queiroz@ieee.org.

II. O PADRÃO H.264/AVC

O H.264/AVC [13], [14], [15] é o mais novo padrão para compressão de vídeo. Recentemente adotado, resulta em ganhos em taxa e distorção substanciais [16], [17] quando comparado a padrões como o MPEG-2 [18] e o H.263 [19]. Entre as melhorias apresentadas pelo H.264/AVC, pode-se citar [13], [20]:

- Predição inter-quadros utilizando blocos de tamanho variável;
- Quadros de referência arbitrários;
- Estimacão em quarto-de-pixel;
- Predição intra-quadro;
- Codificação aritmética binária adaptativa baseada em contexto; e
- Filtro redutor de efeito de bloco.

Em particular, a predição intra-quadro, aliada à codificação aritmética binária adaptativa baseada em contexto (do inglês, *context-adaptive binary arithmetic coding* ou, simplesmente, CABAC), transforma o padrão em um poderoso compressor de imagens estáticas. O objetivo da predição intra-quadro é explorar as similaridades entre macroblocos vizinhos, na tentativa de se minimizar a quantidade de informação a ser codificada. Existem nove modos de predição para blocos de luminância de 4×4 e 8×8 pixels, além de quatro modos de predição para blocos de luminância de 16×16 pixels. Se aplicamos nossa implementação do H.264/AVC à compressão de um único quadro de vídeo, seu comportamento será o de um compressor de imagens estáticas. Tal codificador será denominado AVC-I. Ganhos médios em PSNR (*Peak Signal-to-Noise Ratio*) do AVC-I sobre o JPEG2000 situam-se entre 0,25 dB e 0,5 dB, para imagens pictóricas [20], [21], [22]. Na compressão de documentos os ganhos em PSNR do AVC-I são ainda mais expressivos, até mesmo ultrapassando a marca de 3 dB sobre o JPEG2000, em alguns casos [20].

Quanto à predição inter-quadros, a ideia geral é tentar reduzir a quantidade de informação a ser codificada com base nas similaridades presentes entre quadros. Para se atingir esse objetivo, o quadro atual é inicialmente dividido em blocos menores. Em seguida, buscam-se em quadros de referência, dentro de uma janela, blocos que melhor predizem os blocos do quadro atual. Esse procedimento é chamado de estimacão de movimento e tem como resultado vetores de movimento, que indicam os deslocamentos entre os blocos atuais e seus respectivos preditores. A partir dos quadros de referência e dos vetores de movimento, gera-se um quadro predito que é subtraído do quadro atual. O resultado desse processamento é um quadro residual. No H.264/AVC, a componente de luminância de cada macrobloco é predita a partir de uma partição única de 16×16 pixels, duas 16×8 , duas 8×16 ou quatro de 8×8 . No caso em que as partições de 8×8 pixels são escolhidas, tais partições podem ainda ser subdivididas em duas de 8×4 , duas de 4×8 ou quatro de 4×4 pixels.

A próxima seção descreve o método proposto.

III. O MÉTODO PROPOSTO

A ideia principal aqui apresentada consiste em explorar a recorrência de padrões de texto em um documento, com o

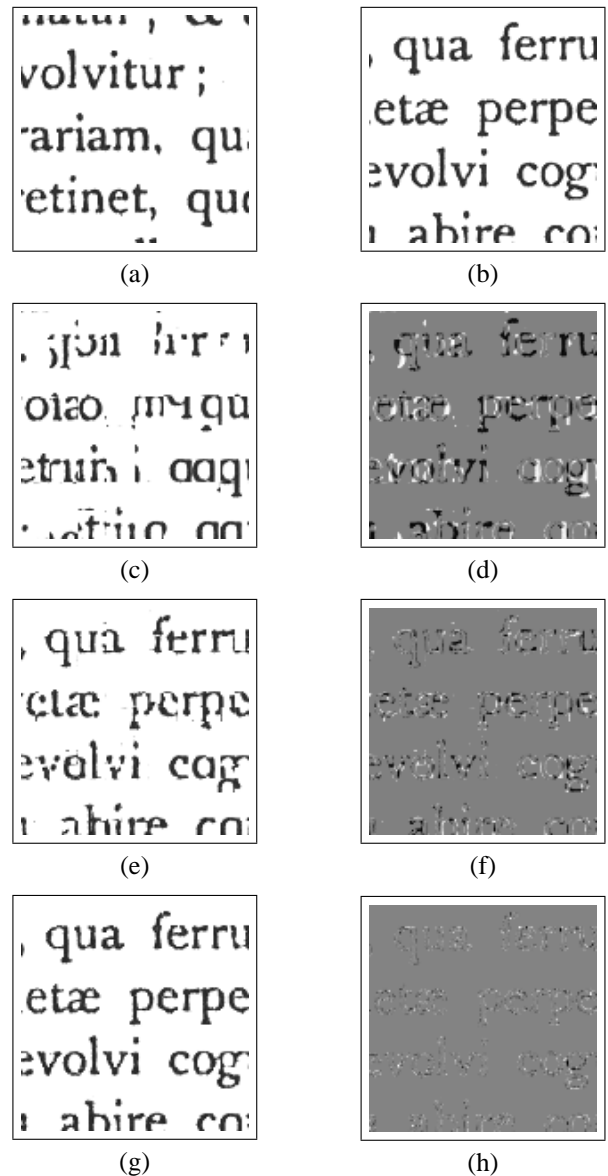


Fig. 1. Casamento de padrões: (a) texto de referência; (b) texto atual; (c) texto predito (tamanho da partição: 16×16 pixels); (d) residuo de predição (tamanho da partição: 16×16 pixels); (e) texto predito (tamanho da partição: 8×8 pixels); (f) residuo de predição (tamanho da partição: 8×8 pixels); (g) texto predito (tamanho da partição: 4×4 pixels); e (h) residuo de predição (tamanho da partição: 4×4 pixels).

objetivo de comprimi-lo mais eficientemente. Sabendo-se que a codificação será realizada pelo H.264/AVC, o documento é processado de forma a permitir que a estimacão de movimento atue como um algoritmo de casamento aproximado de padrões. As Figuras 1 (a) e (b) exemplificam uma área de texto de referência e uma área de texto atual, respectivamente. As Figuras 1 (c), (e) e (g) representam as predições do texto atual utilizando-se partições de 16×16 , 8×8 e 4×4 pixels. As Figuras 1 (d), (f) e (h) mostram os dados residuais correspondentes. Os exemplos da Figura 1 sugerem que áreas de texto previamente codificadas (quadros de referência) podem ser vistos como um dicionário utilizado pelo algoritmo de casamento de padrões (predição inter-quadros). Observe, ainda, como partições

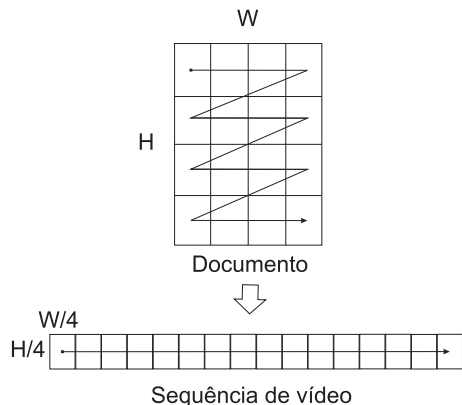


Fig. 2. Esquema de processamento sugerido. O documento é transformado em uma sequência de vídeo.

menores resultam em dados residuais de menor energia. É importante lembrar, porém, que menores partições requerem uma maior quantidade de vetores de movimento. Isso implica que a seleção das dimensões das partições tem um grande impacto na eficiência do codificador. A otimização taxadistorção implementada pelo H.264/AVC será responsável por indicar os modos de predição a serem utilizados.

Para se atingir o objetivo proposto, inicialmente o documento de dimensões $H \times W$ pixels é segmentado em 16 quadros de $H/4 \times W/4$ pixels. Em seguida, os quadros são dispostos de forma a comporem uma sequência de vídeo, conforme ilustrado na Figura 2. Por fim, o vídeo resultante é codificado pelo H.264/AVC que, além de escolher a combinação de predições intra/inter-quadros que resulta em um melhor desempenho em taxa e distorção, aplica sobre o resíduo uma transformada inteira com propriedades semelhantes à DCT (*Discrete Cosine Transform*), quantiza os coeficientes transformados e codifica entropicamente os coeficientes quantizados por meio do CABAC.

A seguir serão apresentados alguns resultados.

IV. RESULTADOS

É importante ressaltar que dois parâmetros terão grande influência na eficiência do codificador. O primeiro é a quantidade de quadros de referência utilizados na predição inter-quadros e o segundo é a janela dentro da qual será realizada a busca por padrões, conforme ilustrado na Figura 3. Na geração dos resultados, os documentos do conjunto de teste foram comprimidos pelos codificadores JPEG2000, AVC-I e pelo método proposto. No caso do JPEG2000 e do AVC-I, o documento foi codificado como uma imagem única. Quanto ao método proposto, o primeiro quadro foi codificado como um quadro *I* (que utiliza somente os modos de predição intra-quadro). Os restantes foram codificados como quadros *P* (que também utilizam os modos de predição inter-quadros). Inicialmente avaliou-se o efeito da escolha do tamanho da janela de busca ($JB = 08, 16$ e 32 pixels) e da quantidade de quadros de referência ($QR = 3, 5$ e 10 quadros). Depois avaliou-se a eficiência do codificador para todo o conjunto de teste, fixando-se JB e QR .

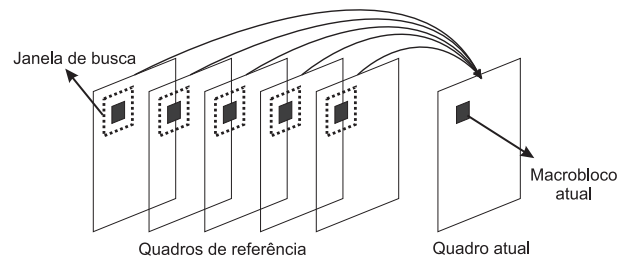


Fig. 3. Ilustração do processo de predição do macrobloco atual a partir da busca dentro de uma janela em quadros de referência anteriormente codificados.

As Figuras 5 (a) e (b) mostram as curvas PSNR resultantes da codificação do documento “cerrado”, que pode ser visualizado na Figura 4 (a). A PSNR foi calculada a partir do MSE (*Mean Squared Error*) global. Na Figura 5 (a) o desempenho do codificador foi avaliado para $JB = 08, 16$ e 32 pixels, mantendo-se $QR = 10$ quadros. Já na Figura 5 (b) a janela de busca JB foi mantida constante em 32 pixels e ajustou-se QR para $3, 5$ e 10 quadros. Observe como o aumento de JB ou de QR determina um melhor desempenho do codificador em termos de taxa e distorção. Em qualquer caso, a eficiência do método proposto é superior à abordagem clássica, aqui representada pelos codificadores JPEG2000 e AVC-I. Em especial, quando $JB = 32$ pixels e $QR = 10$, a proposta supera o AVC-I em mais de 2 dB e o JPEG2000 em mais de 5 dB, a uma taxa de $0,5$ bit/pixel.

As Figuras 4 (b) a (e) mostram os outros documentos que compõem o conjunto de teste. Os documentos (b) e (c) possuem a característica de serem puramente textuais. Os documentos (d) e (e) são documentos compostos (mistura de conteúdos pictóricos e textuais). Suas respectivas curvas PSNR são mostradas nas Figuras 6 (a) a (d). Na compressão desses documentos, JB foi mantido constante em 32 pixels e QR foi fixado em 10 quadros. Observe a superioridade objetiva do método proposto sobre o JPEG2000 e o AVC-I, principalmente para documentos de conteúdo puramente textuais.

Por fim, as Figuras 7 (a) a (d) mostram um trecho original do documento “cerrado”, bem como suas versões codificadas e reconstruídas por meio do método proposto e dos codificadores JPEG2000 e AVC-I, a aproximadamente $0,25$ bit/pixel. Observe como o método proposto fornece imagens com qualidade subjetiva superior.

V. CONCLUSÕES

Foi proposto um método de codificação híbrido, que utiliza o paradigma fundamentado na recorrência de padrões, representado pelo MMP, e o esquema de codificação tradicional, baseado nas etapas de transformação, quantização e codificação de entropia. O H.264/AVC foi utilizado como arcabouço na avaliação do método proposto. Os resultados mostram que é possível alcançar qualidade objetiva e subjetiva superior aos codificadores JPEG2000 e AVC-I.

Trabalhos futuros podem incluir a pré-determinação do tamanho do quadro a ser utilizado na construção da sequência de vídeo, bem como do tamanho da janela de busca e o número de quadros de referência a serem considerados na predição

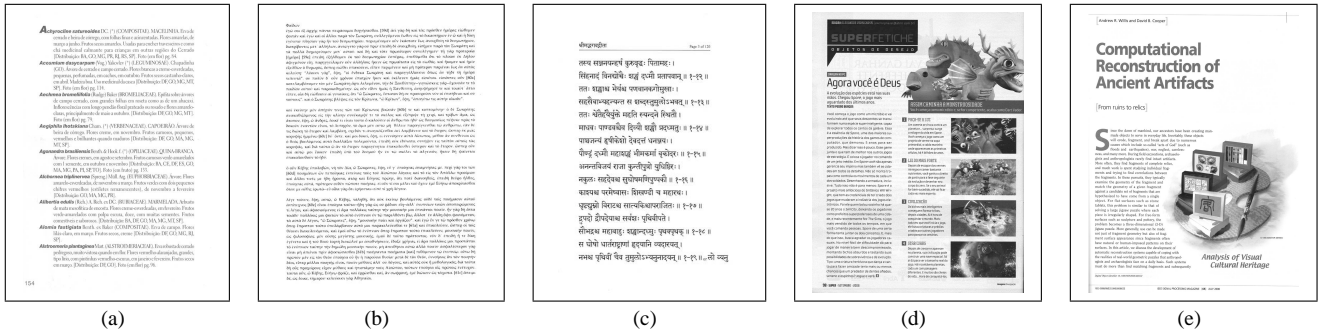


Fig. 4. Conjunto de teste: (a) documento “cerrado”, conteúdo puramente textual, 1568 × 1056 pixels, 200 dpi); (b) documento “fedon” (em grego, conteúdo puramente textual, 1408 × 1056 pixels, 300 dpi); (c) documento “sanskrit” (em sânscrito, conteúdo puramente textual, 1088 × 768 pixels, 200 dpi); (d) documento “spore” (em português, conteúdo misto, 1360 × 1024 pixels, 300 dpi); (e) documento “ancient” (em inglês, conteúdo misto, 1360 × 1024 pixels, 300 dpi).

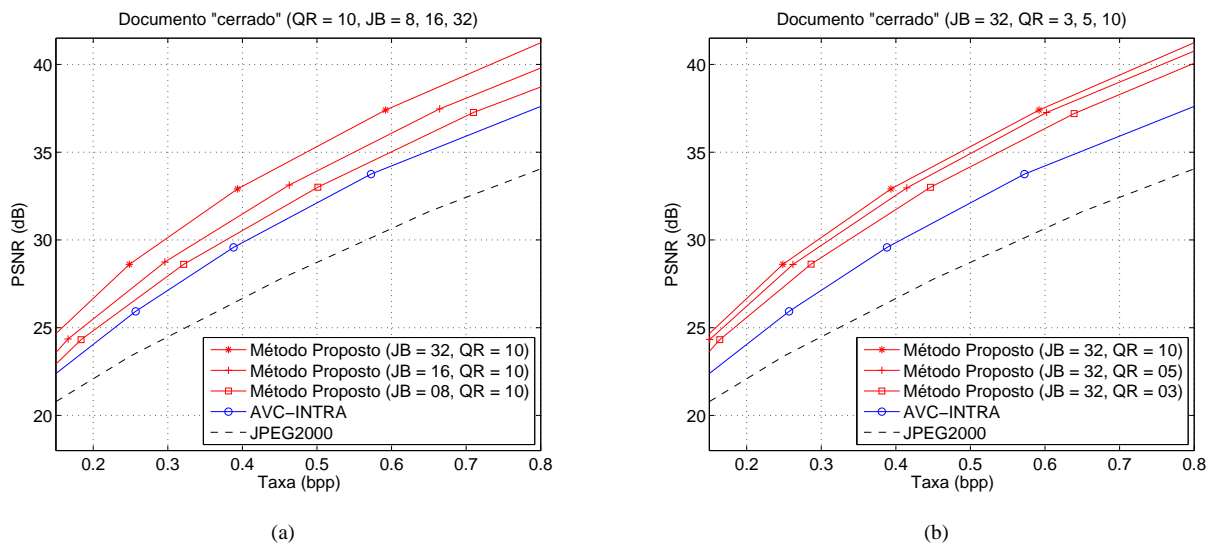


Fig. 5. Desempenho do método proposto na compressão do documento “cerrado”, em comparação com o JPEG2000 e o AVC-I: (a) desempenho para QR = 10 quadros e JB = 08, 16 e 32 pixels; e (b) desempenho para JB = 32 pixels e QR = 3, 5 e 10 quadros. Para JB = 32 e QR = 10, por exemplo, o método proposto supera o AVC-I em mais de 2 dB e o JPEG2000 em mais de 5 dB, a uma taxa de 0,5 bit/pixel.

inter-quadros. O método poderia, ainda, ser empregado na codificação de livros compostos. O pré-processamento, nesse caso, incluiria a classificação das páginas em três tipos: *pictóricas*, *textuais* e *compostas*. As páginas pictóricas e as páginas textuais seriam codificadas com o AVC-I e com o método proposto, respectivamente. Já as páginas compostas seriam submetidas a um processo de segmentação em *regiões pictóricas* e *regiões textuais*. As regiões pictóricas seriam codificadas com o AVC-I e as regiões textuais passariam por uma reorganização antes de serem codificadas pelo H.264/AVC. Além disso, pretende-se, para publicações futuras, comparar o método proposto com o MMP.

Considerando que bibliotecas digitais estão se tornando cada vez mais comuns, o método aqui exposto apresenta-se como uma alternativa bastante interessante, por ser capaz de gerar documentos de alta qualidade, a taxas de *bits* razoavelmente baixas.

AGRADECIMENTOS

Os autores gostariam de agradecer aos alunos Oseas Alves Cardoso e Gilberto Wilson Barbosa da Silva, dos cursos de

Tecnologia da Faculdade Jesus Maria José, pela ajuda na realização dos testes.

REFERÊNCIAS

- [1] M. de Carvalho, E. da Silva, and W. Finamore, “Multidimensional Signal Compression using Multiscale Recurrent Patterns,” *Elsevier Signal Processing*, no. 82, pp. 1559–1580, Novembro 2002.
- [2] N. M. M. Rodrigues, E. A. B. da Silva, and M. B. de Carvalho, “Universal Image Coding using Multiscale Recurrent Patterns and Prediction,” *Proceedings of IEEE International Conference on Image Processing*, pp. 245–248, Setembro 2005.
- [3] N. M. M. Rodrigues, E. A. B. da Silva, M. B. de Carvalho, S. M. M. de Faria, and V. M. M. da Silva, “Improving Multiscale Recurrent Pattern Image Coding with Enhanced Dictionary Updating Strategies,” *Proceedings of International Telecommunications Symposium*, pp. 257–262, Setembro 2006.
- [4] E. B. de Lima Filho, E. A. B. da Silva, M. B. de Carvalho, and F. S. Pinagé, “Universal Image Compression Using Multiscale Recurrent Patterns With Adaptive Probability Model,” *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 512–527, Abril 2008.
- [5] N. C. Francisco, N. M. M. Rodrigues, E. A. B. da Silva, M. B. de Carvalho, S. M. M. de Faria, V. M. M. da Silva, and M. J. C. S. Reis, “Multiscale Recurrent Pattern Image Coding with a Flexible Partition Scheme,” *Proceedings of IEEE International Conference on Image Processing*, pp. 141–144, Outubro 2008.

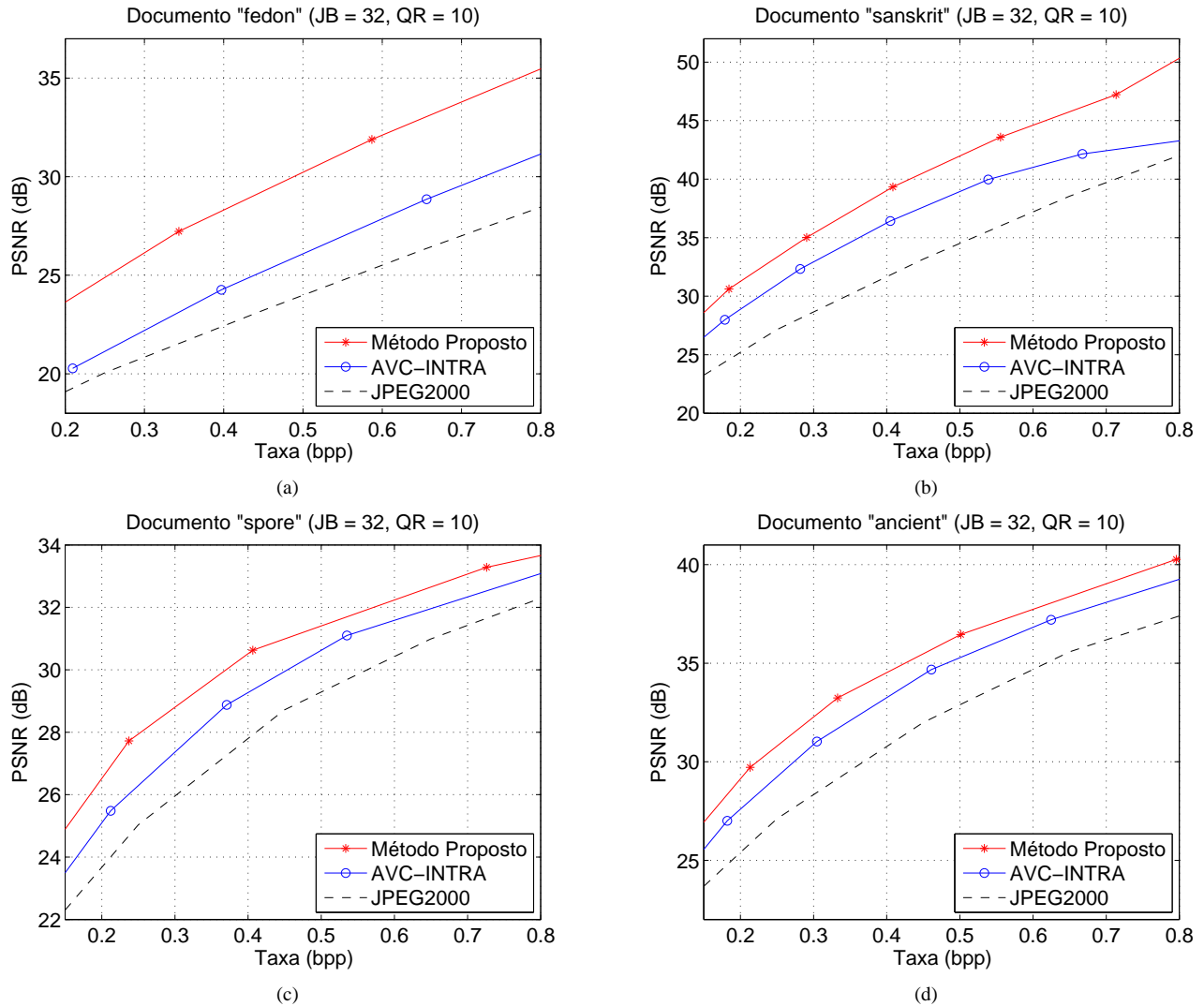


Fig. 6. Comparação entre o método proposto e os codificadores JPEG2000 e AVC-I, para os documentos mostrados no Figura 4: (a) "fedon"; (b) "sanskrit"; (c) "spore"; e (d) "ancient". Tanto em documentos puramente textuais como em documentos compostos o método proposto supera objetivamente a abordagem clássica, aqui representada pelos codificadores JPEG2000 e AVC-I.

[6] JPEG, "Information Technology - JPEG2000 Image Coding System - Part 1: Core Coding System. ISO/IEC 15444-1." 2000.

[7] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. EUA: Kluwer Academic, 2002.

[8] MRC, "Mixed Raster Content (MRC). ITU-T Recommendation T.44." 1999.

[9] R. L. de Queiroz, *Compressing Compound Documents, in The Document and Image Compression Handbook*. EUA: by M. Barni, Marcel-Dekker, 2005.

[10] R. L. de Queiroz, R. Buckley, and M. Xu, "Mixed Raster Content (MRC) Model for Compound Image Compression," *Proceedings of SPIE Visual Communications and Image Processing*, vol. 3653, pp. 1106-1117, Janeiro 1999.

[11] A. Zaghetto, R. L. de Queiroz, and D. Mukherjee, "MRC Compression of Compound Documents using Threshold Segmentation, Iterative Data-filling and H.264/AVC-INTRA," *Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing*, Dezembro 2008.

[12] A. Zaghetto and R. L. de Queiroz, "Iterative Pre- and Post-processing for MRC Layers of Scanned Documents," *Proceedings of IEEE International Conference on Image Processing*, pp. 1009-1012, Outubro 2008.

[13] JVT, "Advanced Video Coding for Generic Audiovisual Services. ITU-T Recommendation H.264," Novembro 2007.

[14] I. E. G. Richardson, *H.264 and MPEG-4 video Compression*. EUA: Wiley, 2003.

[15] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, Julho 2003.

[16] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained Coder Control and Comparison of Video Coding Standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688-703, Julho 2003.

[17] N. Kamaci and Y. Altunbasak, "Performance Comparison of the Emerging H.264 Video Coding Standard with the Existing Standards," *Proceedings of 2003 International Conference on Multimedia and Expo*, vol. 1, pp. 345-348, July 2003.

[18] B. G. Haskell, A. Puri, and A. N. Netravalli, *Digital Video: An Introduction to MPEG-2*. EUA: Chapman and Hall, 1997.

[19] ITU-T, "Video Coding for Low Bit Rate Communication. ITU-T Recommendation H.263," Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

[20] R. L. de Queiroz, R. S. Ortis, A. Zaghetto, and T. A. Fonseca, "Fringe Benefits of the H.264/AVC," *Proceedings of International Telecommunications Symposium*, pp. 208-212, Setembro 2006.

[21] 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, Doc. JVT M-014*, Outubro 2004.

[22] A. Al, B. P. Rao, S. S. Kudva, S. Babu, D. Sumam, and A. V. Rao, "Quality and Complexity Comparison of H.264 Intra Mode with JPEG2000 and JPEG," *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 24-27, Outubro 2004.

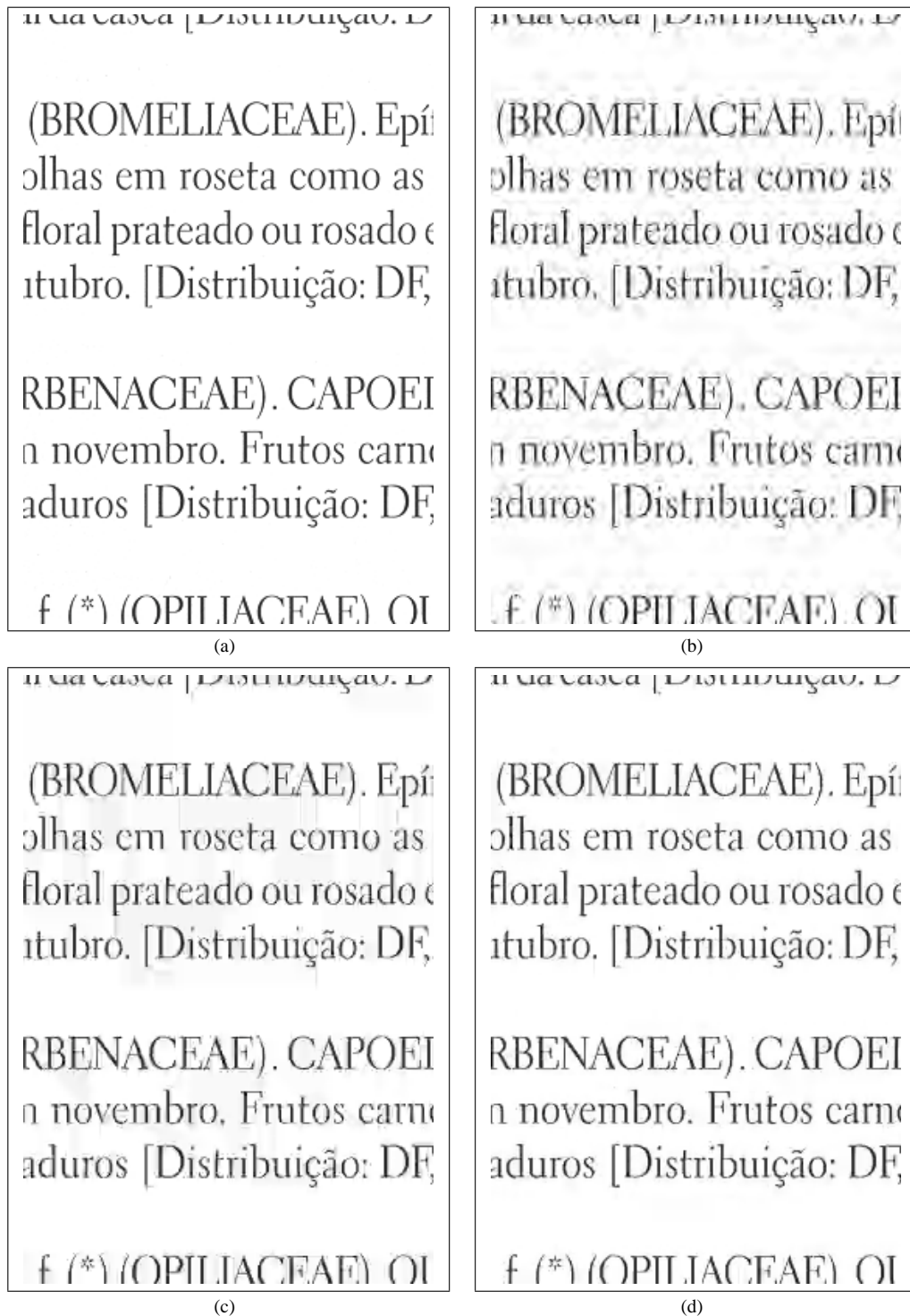


Fig. 7. Comparação subjetiva entre codificadores: (a) parte ampliada do documento “cerrado”; (b) JPEG2000; (c) AVC-I; e (d) método proposto. Taxa de *bits*: 0,25 bit/pixel. O método proposto supera também subjetivamente os codificadores JPEG2000 e AVC-I. Note que em (d) a distorção ao redor dos caracteres é menos perceptível do que em (b) e (c).