HIGH QUALITY SCANNED BOOK COMPRESSION USING PATTERN MATCHING

Alexandre Zaghetto and Ricardo L. de Queiroz

Universidade de Brasilia, Brazil
alexandre@cic.unb.br, queiroz@ieee.org

ABSTRACT

This paper proposes a hybrid approximate pattern matching/ transform-based compression engine. The idea is to use regular video interframe prediction as a pattern matching algorithm that can be applied to document coding. We show that this interpretation may generate residual data that can be efficiently compressed by a transform-based encoder. The novelty of this approach is demonstrated by using H.264/AVC, the newest video compression standard, as a high quality book compressor. The proposed method uses segments of the originally independent scanned pages of a book to create a video sequence, which is encoded through regular H.264/AVC. Results show that the proposed method outperforms AVC-I (H.264/AVC operating in pure intra mode) and JPEG2000 by up to 4 dB and 7 dB, respectively. Superior subjective quality is also achieved.

1. INTRODUCTION

In multiple pages document compression, each page is typically individually encoded by some continuous-tone image compression algorithm, such as JPEG [1] or JPEG2000 [2, 3]. Considering the recurrence of text patterns across pages, or across different areas of the same page, the main idea here presented is to use the many improvements brought into H.264/AVC [4, 5, 6], the newest state-of-the-art video compression standard, to enable a hybrid approximate pattern matching/transform-based scanned book compressor.

It is important to place our proposal within the proper scenario. First, the use of one single coder is required, thus avoiding the inconvenience of handling multiple coders, as in the MRC (Mixed Raster Content) imaging model [7, 8, 9, 10, 11]. Second, the encoded document should be decoded by a codec that common users have access to. Third, the codec must output high quality reconstructed versions of scanned documents. This is specially important when rare books of historical value must be digitally stored. In this case, one must guarantee a reconstructed version of the document which is as close as possible to the original one.

This work was supported by a grant from HP Brazil.

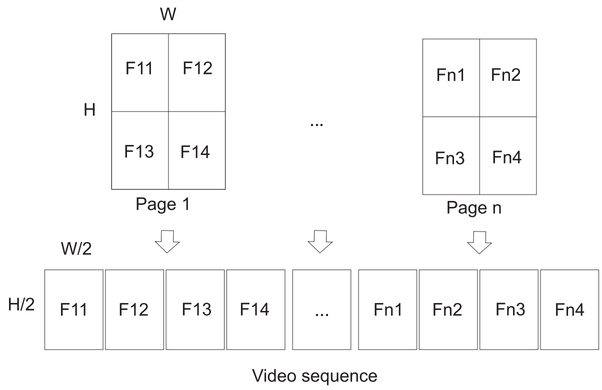


Fig. 1. Proposed page processing algorithm.

2. THE PROPOSED METHOD

Recently adopted, H.264/AVC leads to substantial performance improvement when compared to other existing standards [12, 13], such as MPEG-2 [14] and H.263 [15]. Among such improvements of we can mention [4, 16]: interframe variable block size prediction; arbitrary reference frames; quarter-pel motion estimation; intraframe macroblock prediction; context-adaptive binary arithmetic coding; and in-loop deblocking filter.

Giving that the book will be compressed using H.264/AVC, the proposed encoding method organizes the scanned pages in such a way the interframe prediction may find on previously encoded macroblocks (16×16 pixels blocks) text patterns that are similar to those on the macroblock currently being encoded. Figure 1 illustrates the proposed page processing algorithm.

First, each scanned $H \times W$ pixels page is segmented into four $H/2 \times W/2$ pixels sub-pages. Then, these sub-pages are used to build a video sequence. The only reason page segmentation should be used is that in some cases similar text patterns are more likely to be found on the same page rather than on different pages. If the text style is constant throughout the whole book, each page may be converted into one single frame and segmentation may be skipped. The final step is to compress the resulting video using H.264/AVC.

The basic idea of the interframe prediction is to exploit

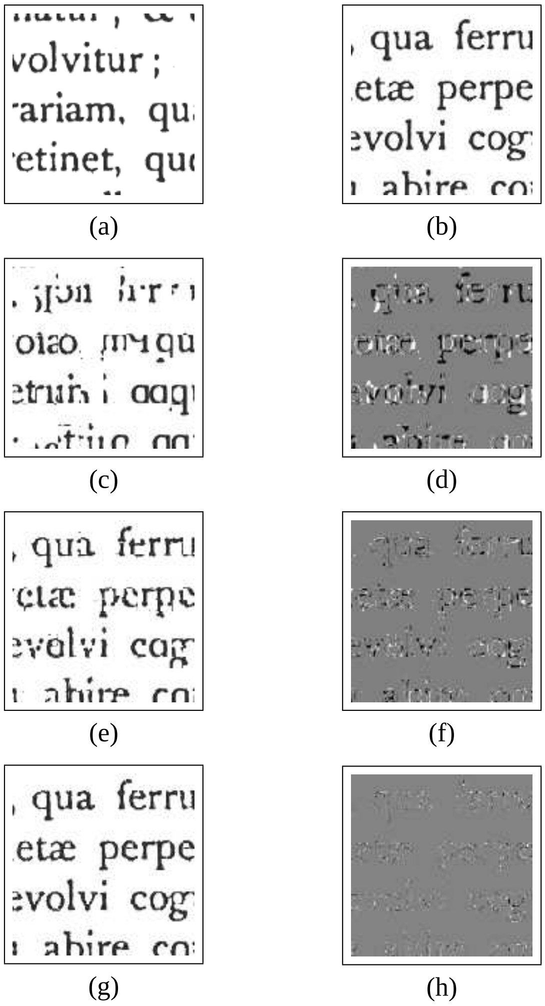


Fig. 2. Approximate pattern matching using interframe prediction: (a) reference text; (b) current text; (c), (e) and (g) predicted text (block size: 16×16, 8×8 and 4×4 pixels, respectively); and (d), (f) and (h) prediction residue (block size: 16×16, 8×8 and 4×4 pixels, respectively).

similarities between video frames in order to reduce the amount of information to be encoded. Based on previously encoded blocks, it first constructs a prediction of the current frame and then creates a residual frame by subtracting the prediction from the current frame. In H.264/AVC, the luma component of each current macroblock is predicted as one 16×16 partition, two 16×8, two 8×16 or four 8×8 macroblock partitions. In case partitions with 8×8 pixels are chosen, the 8×8 sub-macroblocks may be further partitioned in one 8×8 partition, two 8×4, two 4×8 or four 4×4 sub-macroblock partitions. The prediction of each luma block is constructed by displacing an area of the reference frame, determined by a motion vector and a reference frame index.

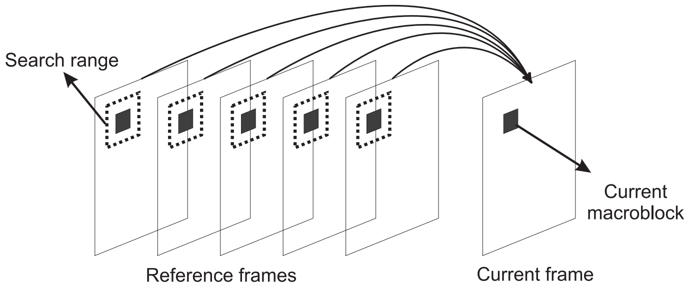


Fig. 3. Configuration parameters that have greater influence on the encoder performance: R_f (number of reference frames) and S_r (search range).

Figure 2 illustrates the effect of using interframe prediction as an approximate pattern matching algorithm. Figures 2 (a) and (b) show examples of a reference and a current text area, respectively. Figures 2 (c), (e) and (g) represent the predictions of the current text using 16×16, 8×8 and 4×4 block partitions. Figures 2 (d), (f) and (h) are the corresponding residual data. Notice that the 4×4 prediction generates a lower-energy residual, when compared with the 16×16 and 8×8 prediction. However, smaller partitions require a larger number of bits to encode the motion vectors. This implies that partition size selection has a major impact on compression performance.

Examples shown in Fig. 2 suggest that previously encoded text areas (reference frames) can be seen as a dictionary used by the pattern matching (interframe prediction) algorithm. The dictionary is updated in parallel with the encoding process, since new reference frames become constantly available. Furthermore, a rate-distortion optimization algorithm is used to estimate which intra/inter modes combination should be applied.

Once the residual data is available, H.264/AVC uses an integer transform with similar properties as the DCT (Discrete Cosine Transform) and the resulting transformed coefficients are quantized and entropically encoded using CABAC.

In the next section we show results that demonstrate the efficiency of the proposed method.

3. RESULTS

Two configuration parameters have greater influence on the encoder performance. One is the number of reference frames, R_f, the other is the search range, S_r, as illustrated by Fig. 3. In our tests, different page sets were compressed using JPEG2000, AVC-I (H.264/AVC operating in pure intra mode) and regular H.264/AVC. In JPEG2000 and AVC-I compression, the pages are encoded separately. As for H.264/AVC, the first frame of the sequence is encoded as an I-frame (only intraframe prediction modes are used) and all the remaining frames are encoded as P-frames (in addition to intraframe prediction, only past frames are used as reference by the interframe prediction). Initially, we evaluated the effect of choosing different values for S_r and R_f. Figure 4 (a) and Figures 5 (a) and (b) show the first page of test sequence "guita" and PSNR plots comparing JPEG2000, AVC-I and H.264/AVC, for different combinations of S_r and R_f, respectively. The PSNR was calculated using the global mean square error (MSE). The higher S_r and R_f values, the better rate-distortion performance. In particular, for S_r = 32 pixels and R_f = 5 frames, H.264/AVC outperforms AVC-I by more than 2 dB and JPEG2000 by more than 5 dB, at 0.5 bit/pixel (bpp).

Four other page sets, whose first pages are shown in Figs. 4 (b)-(e), were also compressed using the proposed method (with $S_r = 32$ and $R_f = 5$). PSNR plots are shown in Figs, 5 (c)-(f). In all cases, JPEG2000 and AVC-I were objectively outperformed by H.264/AVC. Figure 6 (a) shows a zoomed part of the original "principia" sequence. Its encoded and reconstructed versions using AVC-I, JPEG2000 and H.264/AVC, at approximately 0.25 bits/pixel, are shown in Figs. 6 (b), (c) and (d), respectively. H.264/AVC also yields superior subjective quality.

4. CONCLUSION

In this paper we presented a hybrid pattern matching/ transform-based encoder for scanned books. The reason we decided to use H.264/AVC to evaluate the proposed method is that its interframe prediction scheme together with rate-distortion optimization indirectly implement an efficient pattern matching algorithm. In addition, the intraframe prediction, the DCT-based transformation and CABAC, proposed by the standard, also contribute to improve the encoding efficiency.

Results show that the proposed method objectively outperforms AVC-I and JPEG2000 by up to 4 dB and 7 dB, respectively. Furthermore, the encoder outputs documents with superior subjective quality. Future works may include single-page compound document and multi-page compound book compression. Comparison with the recently proposed Multidimensional Multiscale Parser (MMP) [17] may also be carried out.

5. REFERENCES

[1] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. Chapman and Hall, 1993.

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

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

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

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

[6] 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.

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

[8] 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.

[9] P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. Lecun, "High Quality Document Image Compression with DjVu," *Journal of Electronic Imaging*, vol. 7, pp. 410–425, 1998.

[10] G. Feng and C. A. Bouman, "High-quality mrc document coding," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3152–3169, Outubro 2006.

[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] 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.

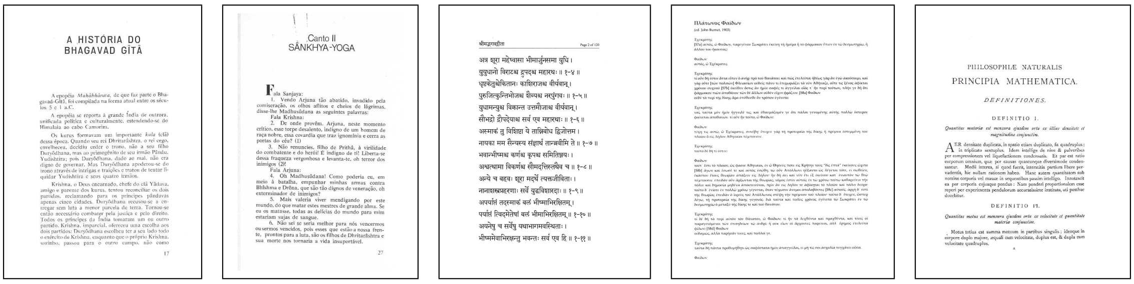
[13] N. Kamaci and Y. Altunbasak, "Performance Comparison of the Emerging H.264 Video Coding Standard with the Existing Standards," *Proceedings of 2003 International Conference on Multimedia and Expo*, vol. 1, pp. 345–348, July 2003.

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

[15] 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.

[16] 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.

[17] 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.



(a) (b) (c) (d) (e)

Fig. 4. First pages of test sets used in our experiments: (a) “guita” (total number of pages: 2, page size: 1568×1024 pixels); (b) “samkhya” (total number of pages: 3, page size: 1568×1008 pixels); (c) “sanskrit” (total number of pages: 4, page size: 1088×768 pixels); (d) “fedon” (total number of pages: 5, page size: 1408×1056 pixels); and (e) “principia” (total number of pages: 8, page size: 1600×1152 pixels).

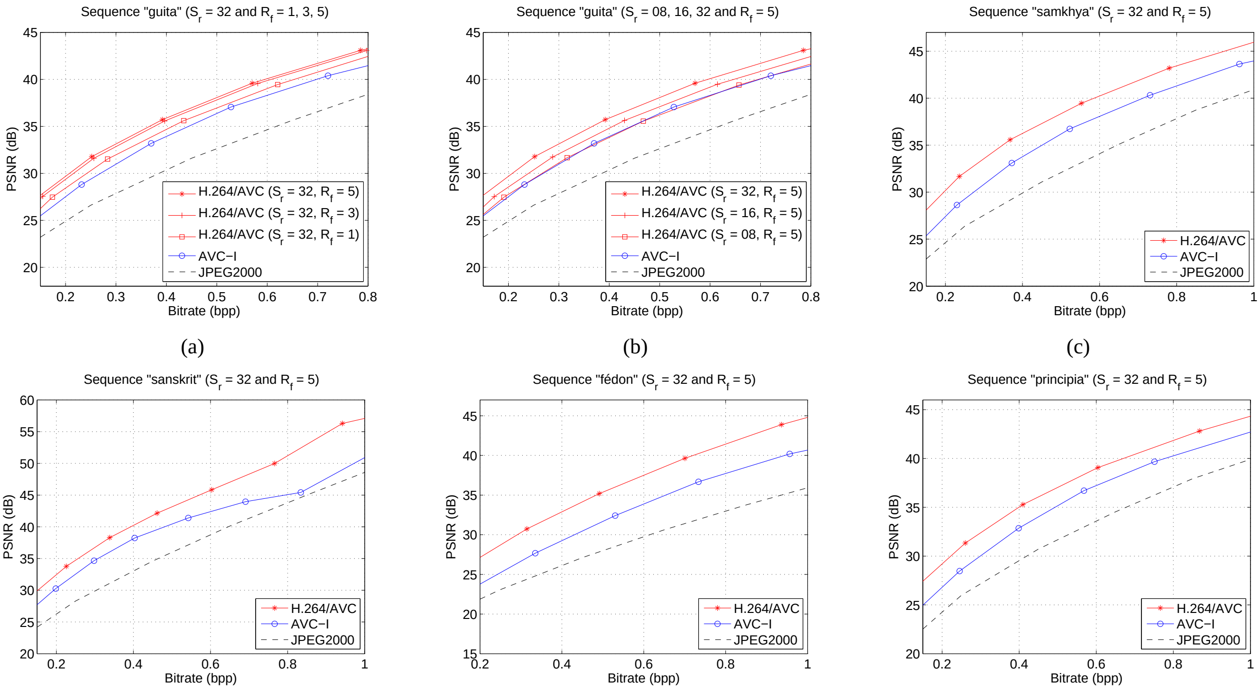
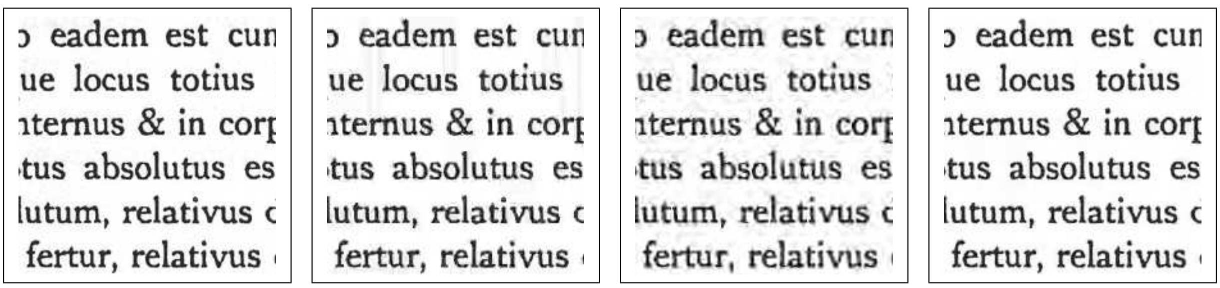


Fig. 5. PSNR plots for test sets shown in Fig. 4. (a) and (b) “guita”: comparison between JPEG2000, AVC-I and H.264/AVC, for different combinations of search ranges (S_r) and number of reference frames (R_f); (c) “samkhya” ($S_r = 32$ and $R_f = 5$); (d) “sanskrit” ($S_r = 32$ and $R_f = 5$); (e) “fedon” ($S_r = 32$ and $R_f = 5$); and (f) “principia” ($S_r = 32$ and $R_f = 5$).



(a) (b) (c) (d)

Fig. 6. Subjective comparison between coders: (a) zoomed part of “principia” sequence; encoded and reconstructed versions using (b) AVC-I, (c) JPEG2000 and (d) H.264/AVC, at approximately 0.25 bits/pixels. H.264/AVC yields superior subjective quality.