ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

Low Footprint Compression of Scanned Documents Using an Advanced Hybrid Video Coding Technique

Sudeshani V. Balwatkar, Prof. Savita S. Raut

ME II Year Student, Siddhant College of Engineering, Sudumbre, Pune, Maharashtra, India

Project Guide, Department of Electronics, Siddhant College of Engineering, Sudumbre, Pune, Maharashtra, India

ABSTRACT: A light-weight scanned document is aimed to be achieved using a promising mechanism & implementation of a method that is gives a low-memory-footprint compression of scanned documents. Storing of scanning documents, images, etc as good as the originally written documents preserving the quality of scanned document as well as storing it in minimal possible storage size is the need of time. High quality documents generally are memory heavy and thus take up lot of disc space making high quality compression of scanned documents critical.

The implementation is done using a hybrid pattern matching/ transform-based compression method for scanned documents. Document encoding is achieved with a regular video inter-frame prediction method like an algorithm for pattern matching. This implementation terms the efficient method as advanced document coding (ADC), which makes a sequential video frame from the fragments of the separate page(s) of the scanned document and then the video frame is further encoded with regular H.264/AVC. Performance of this method is considerably higher over other prevalent methods for scanned document compression giving us a superior subjective quality.

KEYWORDS: Advanced document coding (ADC), H.264/AVC, pattern matching, scanned document compression.

I. INTRODUCTION

T his world we live in today is digitizing itself- be it the Business, Education, Administration, etc. Each and every aspect of the world has a digital clone or is migrating itself to the digital world, documents in written form are no exception to it. Government offices, various businesses, historians, researchers, students, law agencies, etc. are trying to gain & grant a wider access to various forms of documents (older as well as the latest ones that are produced daily). Storing & duplicating the physical documents and maintaining these for long period is not easy, safe and mostly cannot be guaranteed. Thus importance of scanning documents, images, etc has gained lot importance and has also become a standard practice with many. To ensure that these stored documents are as good as the originally written documents, quality of scanning and storage is very important. High quality documents generally take up lot of disc space and thus high quality compression of scanned documents has become more than important. Further, to ensure archiving with backups for guaranteed retrieval of the scanned documents, a highly efficient method preserving the quality of the scanned

Given the nature of scanned documents which are scanned images of documents, compression is not a straight forward job. The compression may be done as image compression as a continuous tone picture or compression may be done after binarization. In the later case, the binarized document is compressed using the existing methods of two-level lossless compression through JBIG and JBIG2 or through character recognition. But the drawbacks of binarization which cause heavy degradation of object counters and textures, it is preferable to go for the former method of continuous-tone compression. In case of multiple page documents, continuous-tone algorithms such as JPEG or JPEG2000 can be used, which encode the individual scanned document images. Further, the soft edges of the scanned documents challenge the multi-layer approach followed by MRC (mixed raster content), thus calling for pre and post-processing**.**

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

Due consideration has to be given to the efficient dictionary based encoder relying on continuous tone pattern matching of the natural text having repetitive symbols as shown in Fig.1.

Fig. 1. Digitized books usually present recurrent patterns across different pages and across regions of the same page.

The method considered for compression of the documents scanned uses an encoder which finds such repetitions by pattern-matching predictors and efficient transform encoding of the residual data. The considered encoder was chosen to suit the need for single coder which is a standard coder [5] which can output high quality compound documents (includes text, images), refer Fig. 2.

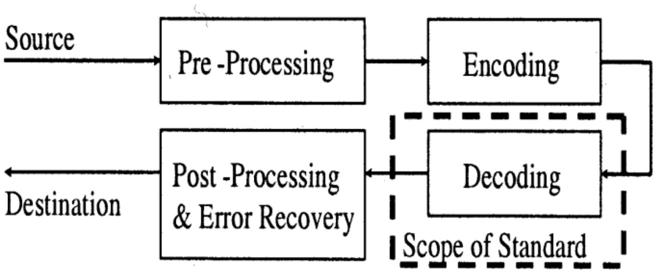


Fig. 2. General scope of coding standardization

The implementation method followed uses a hybrid pattern matching/ transform-based compression method for scanned documents [1]. Using a regular video inter-frame prediction method like an algorithm for pattern matching, it is implemented for document encoding. The considered coder uses the standard video coder H.264/AVC. This standard is well explained through [3] – [6]. The method followed, termed as advanced document coding (ADC), segmentizes (into blocks) the originally independent scanned pages of a document to create a video sequence, which is then encoded through regular H.264/AVC. The intra-frame macroblock prediction combined with the context-adaptive binary arithmetic coding (CABAC) makes the H.264/AVC a powerful image compression method [1]. The considered method provides superior results over the existing standards of compression.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

II. METHOD CONSIDERED FOR IMPLEMENAION OF ADC

The considered method proposes to use a hybrid pattern matching/ transform-based compression method for scanned documents. The hypothesis [1] referred is that 'A scanned document encoder that employs state-of-the-art video coding techniques and generates a bit-stream that can be decoded by H.264/AVC so as to get excellent rate-distortion performance

in comparison with other compressors for continuous tone still-images. The implementation can be understood as first the document(s) is split into many pages which are further segmented to form frames which are fed to the Coder. This method uses segments of the separate page(s) of the scanned document to make sequential a video frame, and then encoded with regular H.264/AVC. H.264/AVC leads to substantial performance improvement when compared to other available standards, such as MPEG-2 and H.263. The many coding advances brought through H.264/AVC which are benchmark for video compression, also making this standard an excellent compressor for still images [1]

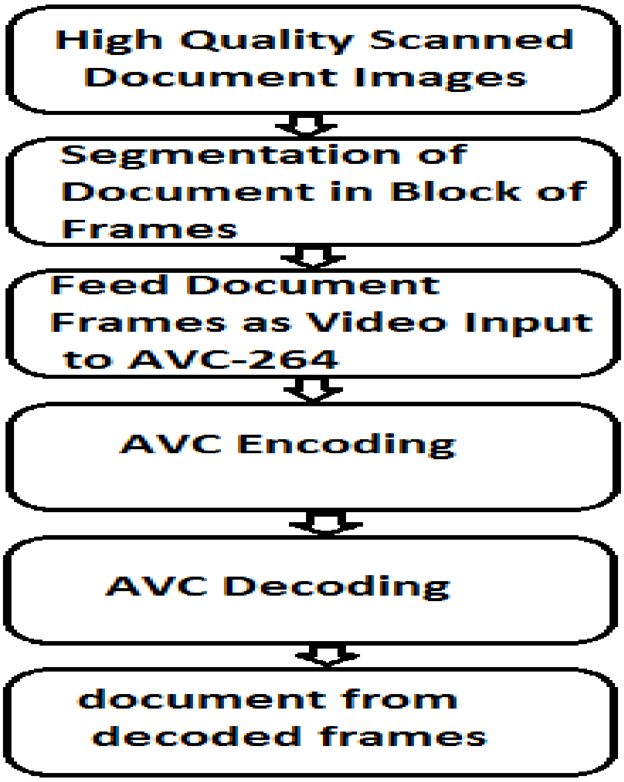


Fig. 3. Flow diagram of proposed method.

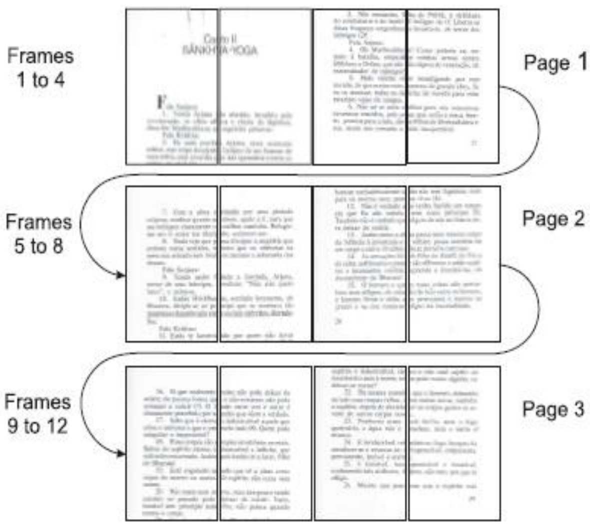


Fig. 4. Partitioned scanned document page in sequential segments. Each one is considered a frame and can be sequentially encoded.

A. Pattern matching

The implementation is captured below through a simplified workflow as depicted in Fig. 3. The single/multi-page documents when scanned are fed to do segmentation of these scanned documents. The segmentation divides the scanned documents into blocks of pixels. These blocks are sequentially clubbed to form frames. Single/ Multiple pages of the scanned document segmented into multiple blocks of H x W pixels, into NP sub-pages or frames ($H/\sqrt{Np} \times W/\sqrt{Np}$). These partitioned blocks are matched to an existing pattern in a dictionary which is derived by previous contents of same document. The segmentation of a scanned document is shown in Fig. 4 which also depicts a frame sequence derived from a set of three pages. A digitized video signal consists of a periodical sequence of images forming frames consisting of a two dimensional array of pixels made of three color components, R, G and B.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

Generally, the pixel data is converted from RGB to another color space called YUV in which U and V components to be further sub-sampled [2]. The encoding technique that is block-based breaks a frame into macroblocks of 16 x 16 pixels. In a 4:2:0 format, each macroblock (MB) consists of 16 x 16 = 256Y components and 8 x 8 = 64U and 64V components. Each of three components of a macroblock is processed separately. To compress an MB, we use a hybrid of three techniques: prediction, transformation & quantization, and entropy coding.

In a 16 x 16 pixel block, each pixel is matched with patterns found in the previous frames. This creates a dynamic dictionary of patterns for look-up when a current frame is being encoded and this dictionary is continuously updated as more frames encoded. After matching, the matching pattern is used for prediction of the block. The residue i.e., the prediction error is encoded with the frame number and the reference vector (position where the match was found). As shown in Fig. 5, the effect of the pattern matching using inter-frame prediction in AVC can be understood. Fig 5(a) is the reference text and Fig 5(b), 5(c) and 5(d) illustrates examples of predicted text against the reference text with respective block size of 16 x 6, 8 x 8, 4 x 4 pixels. Further in Fig. 6, the effect of the pattern matching using inter-frame prediction in ADC can be understood. Fig 6(a) is the reference text and Fig 6(b), 6(c) and 6(d) illustrates examples of predicted text against the reference text with respective block size of 16 x 6, 8 x 8, 4 x 4 pixels. As seen, the smaller block sizes like the 4 x 4 pixels requires encoding more reference vectors but generates lower-energy residual during the predictions when compared with predictions for 16 x 16, 8 x 8 blocks.

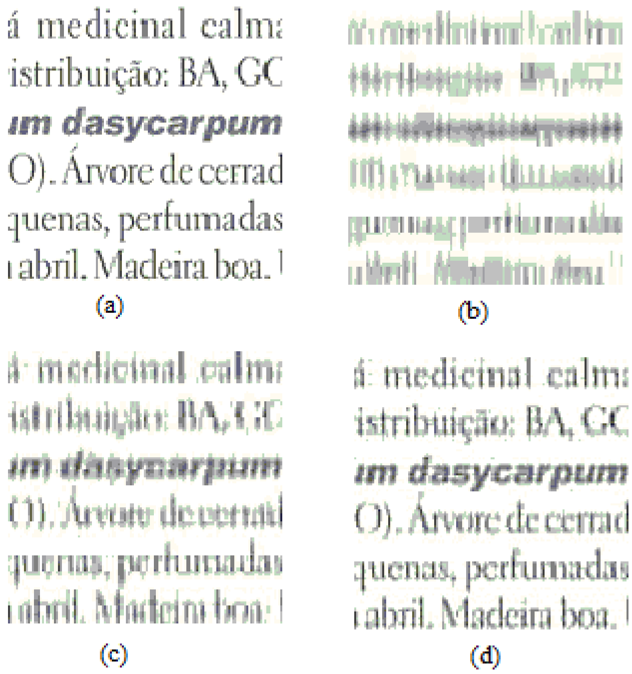


Fig. 5. Pattern Matching using Inter-Frame Prediction using AVC, (a) Reference text. And (b), (c), (d) Predicted text (block size: 16 x16, 8 × 8, and 4 × 4 pixels, respectively).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

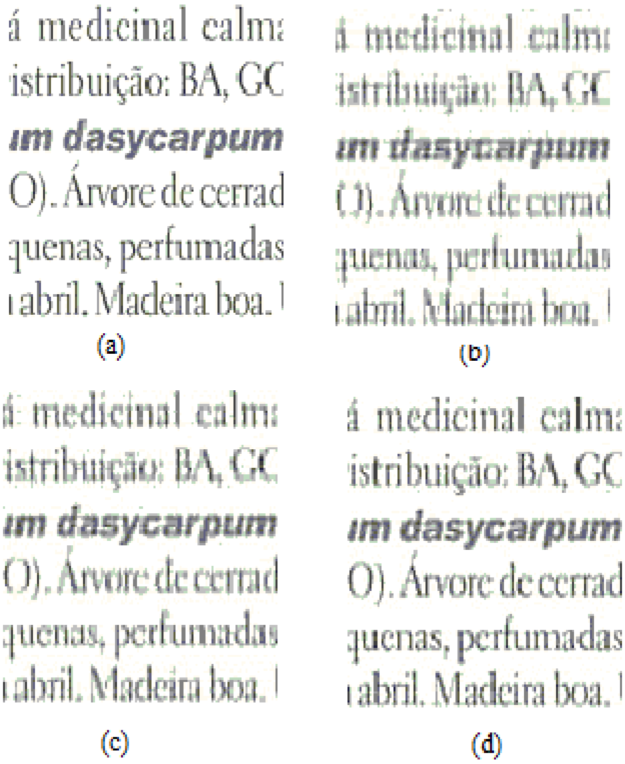


Fig. 6. Pattern Matching using Inter-Frame Prediction using ADC, (a) Reference text. (b) ,(c), and (d) Predicted text (block size: 16 x16, 8 × 8, and 4 × 4 pixels, respectively).

The pattern matching is possible to be achieved through the motion estimation technique of H.264/AVC. The input to motion estimation algorithm are document frames in the form of sequential video frames formed from various blocks of 16 x 8, 8 x 16, 8 x 8, 4 x 8, 8 x 4, and 4 x 4 pixels derived from the macroblocks of 16 x 16. Care has to be taken to counter the motion estimation techniques reliance on the method which considers a close correlation between the neighboring frames at the same position. This can be compensated by making the search window to cover more reference frames or the whole frame so that the dynamic dictionary can be enriched further along with the elimination of the spatial dependencies.

B. Frame Prediction

Based on the reference MB, prediction is categorized as inter-frame prediction and intra-frame prediction. In the inter-frame prediction (P or B) mode, the reference macroblock is within the frame before or after the current frame. It may also be some weighted function of MBs from multiple frames. In an intra-frame prediction (I) mode, the reference MB is usually calculated with mathematical functions of neighboring pixels of the current MB. In Intra-frame prediction, the information from already transmitted macroblocks of the same image is used for predicting the samples of a macroblock.

In H.264/AVC, an I-type 16 x 16, 4:2:0 MB has its luminance component (one block of 16 x 16); chrominance components (two blocks of 8 x 8) separately predicted [2].

There are many ways to predict a macroblock as illustrated in Fig. 7.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

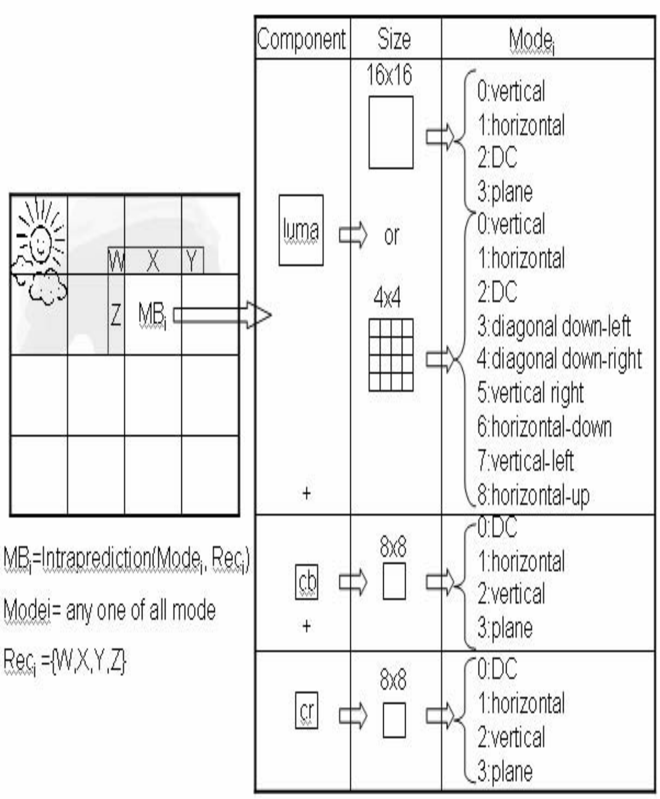


Fig. 7. Overview of H.264/AVC intra- prediction mode.

In H.264/AVC intra prediction for the prediction of the luminance component Y, is possible in two distinct types. The first type is called INTRA_4 x 4 and the second one INTRA_16 x 16, each 4 x 4 block utilizes one of nine prediction modes (one DC prediction mode and eight directional prediction modes) [2]. When using the mentioned INTRA_16 x 16 case that's suited for smooth image area, a uniform prediction is executed for the whole luminance component of a macroblock. Four prediction modes are defined which used to predict each chrominance component as a single 8x8 block.

Usage of intra-frame prediction depends a lot on rate-distortion optimization (RDO) which is available in the AVC whereas the in-frame motion vectors are not allowed in the AVC. In the considered encoder, using AVC structures each block portioning combination and prediction mode is tested to pick up best mode based on the RDO.

C. Residual Error

During prediction, there is a residual error data also termed as residual macroblock. This residual data is transformed using four 8 x 8-pixel discrete cosine transform (DCT) or an integer approximation of it. The transformed blocks are quantized and encoded using arithmetic coding. H.264/AVC uses an integer transform with similar properties as the DCT and the resulting transformed coefficients are quantized and entropy encoded using CABAC [1].

D. Compound Documents and Region Classification

Compound document which is a combination of image and text are compressed after segmentation of the image into regions which are further classified as containing text and graphics or images (or halftones, for instance). After classification of the regions, the encoding with suitable algorithm is initiated. Unlike this approach that is driven by objects such as text characters tagged based on the estimation of contents of the regions of the image, the considered method is based on the compression only. Instead of following prediction based on pattern matching of every block

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

partition, followed method makes testing of prediction by extrapolating neighboring blocks, as in "intra-prediction" in H.264/AVC. The RD-optimized selection of the best prediction assures that the best option is picked. Recurrence of patterns for text and graphics often makes the encoding using patterns from previous regions, while pictorial regions may resort to intra-prediction. Thus encoding integrates segmentation into its process itself. The block prediction and RDO may have the same effect of a segmentation map, benefiting the compression process [1].

E. Codec Summary

As mentioned earlier, we feed the sequential frames into AVC just like in a regular video coder. As these sequential frames are nothing but frames derived from document and then these sequential frames are being fed to the AVC, the terminology of ADC (Advanced Document Coding) is being referred for this method. We can summarize this process as an operation which breaks page (s) of a document into sequential frames, then feed all these frames to H.264/AVC to get an AVC compatible stream, then decode the bit stream and reassemble the frames to reflect the page-wise document. The H.246/AVC is operated in "High" profile mode that follows an IPPP... framework. If random access is required, no-reference I-frames have to be used in the encoder at regular intervals. RDO should be turned on. Motion estimation should be set to full search over a window that is as large as possible.

III. EXPERIMENTAL RESULTS

The system is based on relative improvement in ADC (H.264/AVC) for document compression of scanned documents over other existing methods like AVC-1, JPEG2000. The reference comparison tests tries to compress different page sets using JPEG2000, AVC-I (H.264/AVC operating in pure intra mode) and the proposed ADC (as per the above section-E of Codec Summary). JPEG2000 and AVC-I are used for comparison as these are the most suitable standards meeting the assumptions presented in the introduction section. For compression using JPEG2000 and AVC-I, the pages are individually encoded. For compression through ADC, initial frame of the sequence is encoded as an I-frame (by using intra-frame prediction mode) and balance frames are encoded as P-frames (in addition to intra-frame prediction, only past frames are used as reference by the inter-frame prediction). In these tests, segmentation of the page(s) is done as follows: N_P = 4 frames, N_P = 16 frames, or not segmented at all (N_P = 1 frame, for multi-page documents only). In Fig. 8, R_f – reference range and S_r – search range, influencing the encoder performances are two configuration parameters.

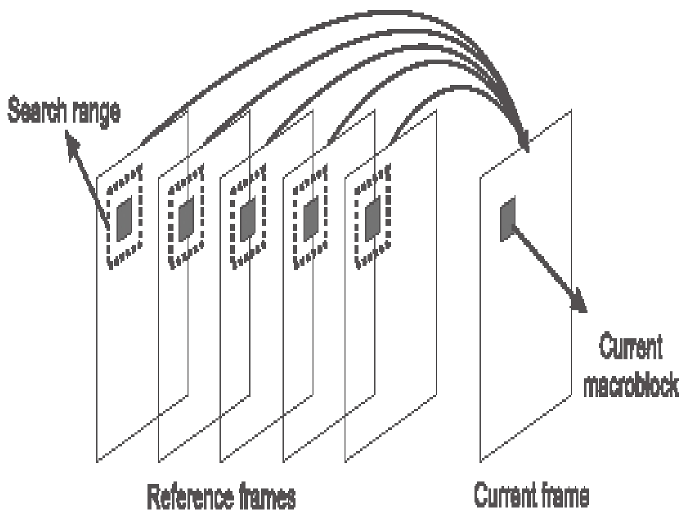


Fig. 8. Configuration parameters influencing the encoder performance: reference frames - R_f and search range - S_r

The expected reference results that our system should try to achieve are captured below through Fig. 9 to Fig. 11.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

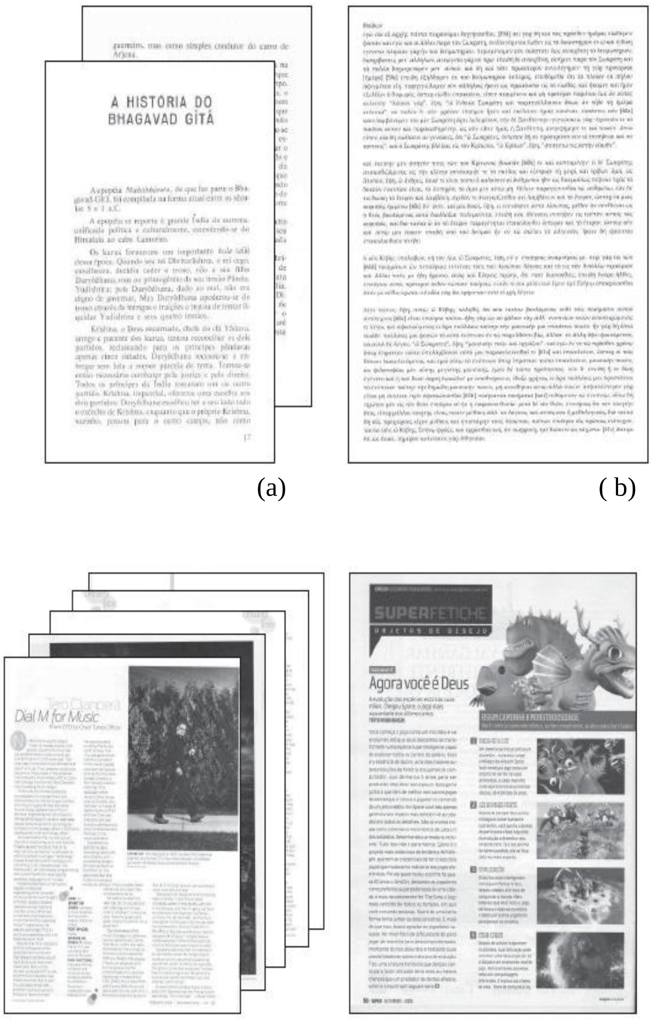
International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

The reference test set is composed by various documents divided into the following 4 classes.
1) Class 0: 6 multi-page text-only documents.
2) Class 1: 6 single-page text-only documents.
3) Class 2: 3 multi-page compound documents.
4) Class 3: 3 single-page compound documents.

Fig.9 shows examples for test performed for each class.



(a) (b)

(c) (d)

Fig. 9. Examples of documents used in our experiments for the test performed for each class.

The evaluation is through the effect of choosing different values for S_r and R_f. The PSNR plots (calculated using global mean square error - MSE) are compared between JPEG2000, AVC-I and ADC (Np = 4 frames/page), for different combinations of S_r and R_f which can be seen in Fig. 10.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

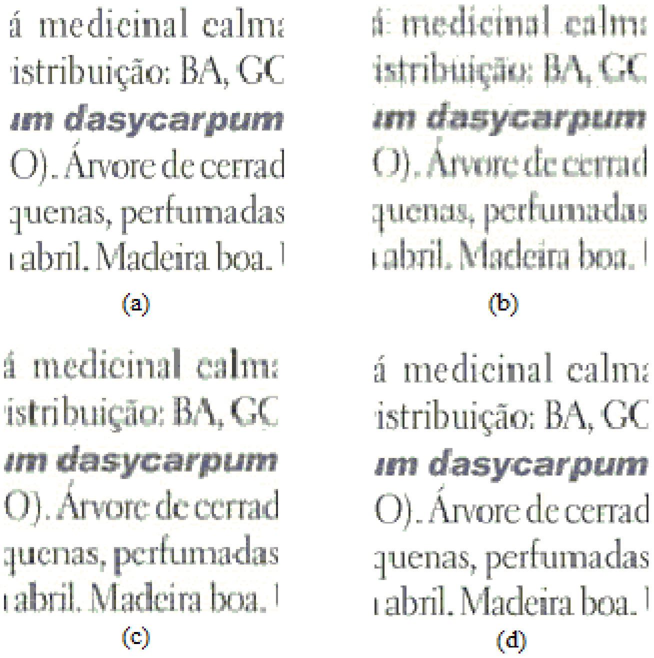


Fig.10. Result of block frame of cerrado page of class 0 dataset (a) Original Image, (b) Reconstructed image from JPEG 2000, (c) Reconstructed image from AVC-264, (d) Reconstructed image from ADC

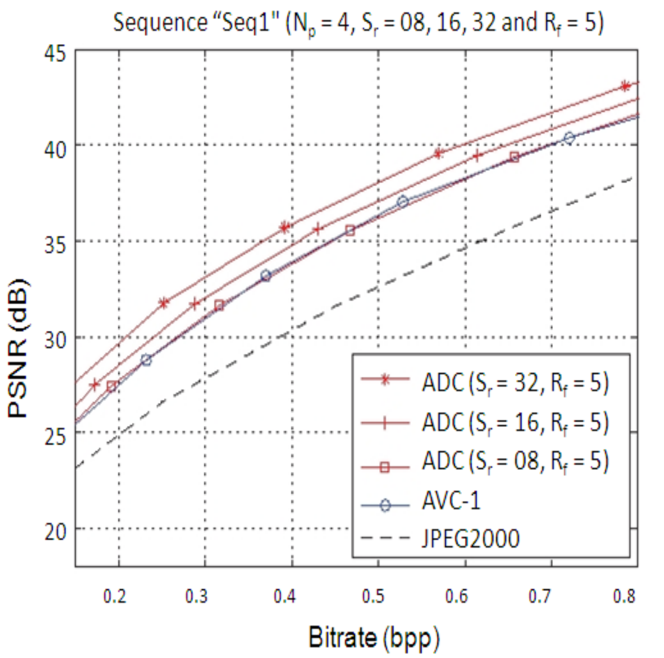


Fig. 11. Comparison between JPEG2000, AVC-I and proposed ADC for a test sequence 'Seq1' (a) Fixed search range (S_r =32), different number of reference frames (R_f = 1,3,5)

In Fig. 11, it can be observed that for better rate-distortion performance, higher the S_r and R_f values are required. For the ADC to outperform AVC-I and JPEG2000, the values of S_r and R_f are entered as S_r = 32 pixels and R_f = 5 frames.

IV. CONCLUSION

The work being referred and being represented through this paper helps us move one more step towards achieving better efficiency for storage of scanned documents while maintaining the quality of these compound documents. The simplified process of achieving this result is possible through the use of the ADC which is pattern matching/transform based encoder applied on scanned documents. The method followed, termed as advanced document coding (ADC), uses block segments of the separate page(s) of the scanned document to make sequential a video frame, and then encoded with regular H.264/AVC. The efficiency of encoding is improved due to the DCT-based transform and the CABAC, in addition to the intra-frame prediction.

The performance of this ADC method is superior to the other methods for scanned document compression giving a superior subjective quality. Targeted results for ADC are expected to outperform AVC-I and JPEG2000.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Volume 4, Issue 3, March 2015

REFERENCES

[1] A. Zaghetto, and R. L. de Queiroz , "Scanned Document Compression Using Block-Based Hybrid Video codec", *IEEE Transactions On Image Processing*, Vol. 22, No. 6, Jun. 2013

[2] J. Chen, C. Kao, Y. Lin, "Introduction to H.264 Advanced Video Coding", *ASP-DAC 2006 - 11th Asia and South Pacific Design Automation Conference*.

[3] 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, Mar. 2004.

[4] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, 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, Jul. 2003.

[5] 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, Jul. 2003.

[6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*. New York, USA: Wiley, 2003

[7] A. Puri, X. Chen, and A. Luthra, "Video coding using the H.264/ MPEG-4 AVC compression standard" *ScienceDirect - Signal Processing: Image Compression Journal,* Vol. 19, No. 9, Oct. 2004.