ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2011.0041



# USING H.264/AVC-INTRA FOR DCT BASED SEGMENTATION DRIVEN COMPOUND IMAGE COMPRESSION

**S. Ebenezer Juliet[1], V. Sadasivam[2] and D. Jemi Florinabel[3]**
*Department of Computer Science and Engineering, Manomaniam Sundaranar University, Tamil Nadu, India*
E-mail: [1]juliet-sehar@yahoo.com, [2]vs_msu@yahoo.com, [3]jemidelta@yahoo.co.in

***Abstract***
***This paper presents a one pass block classification algorithm for efficient coding of compound images which consists of multimedia elements like text, graphics and natural images. The objective is to minimize the loss of visual quality of text during compression by separating text information which needs high special resolution than the pictures and background. It segments computer screen images into text/graphics and picture/background classes based on DCT energy in each 4x4 block, and then compresses both text/graphics pixels and picture/background blocks by H.264/AVC with variable quantization parameter. Experimental results show that the single H.264/AVC-INTRA coder with variable quantization outperforms single coders such as JPEG, JPEG-2000 for compound images. Also the proposed method improves the PSNR value significantly than standard JPEG, JPEG-2000 and while keeping competitive compression ratios.***

***Keywords:***
***Compound Image Compression, Block Classification, DCT Coefficients, H.264/AVC-Intra***

## 1. INTRODUCTION

Besides natural images, there are millions of artificial visual contents generated by computers every day, such as Web pages, PDF files, slides, online games, and captured screens. They are usually a combination of text, graphics, and natural image. This is why they are generally called as compound images. In particular, with cloud computing becoming more and more popular, compound images often need to be displayed on remote clients, wireless projectors, and thin clients. Some of the clients are unable to directly render them from files. Compressing and transmitting compound images provides a generic solution to these clients. However, in this solution, how to efficiently compress compound images has become a prevalent and critical problem.

There are several compression formats such as JPEG, JPEG-2000 available for compression of natural images. Similarly, there are several compression formats for binary texts and simple graphics. But no single algorithm is best across all types of images or applications. When compressing text, it is important to preserve the edges and shapes of characters accurately to facilitate reading.

### 1.1 COMPOUND IMAGE COMPRESSION

Layer-based and block-based algorithms are two main methods frequently used in compound image compression. Most layer-based approaches use the standard three-layer mixed raster content (MRC) representation [1]. Document image compression DjVu [2] uses a wavelet-based coder (IW44) for the background and foreground, and JB2 for the mask. The segmentation is based on hierarchical color clustering and a variety of filters. Block-based approaches are studied mainly due to its low complexity. Li and Lei [3] proposed a histogram analysis to classify each 16x16 block into one of the four types: smooth block, text block, graphics block and image block. Cheng and Bouman [4] investigated two segmentation algorithms (TSMAP and RDOS) to classify 8x8 blocks into four classes (picture blocks, two-color blocks, one-color blocks, and other blocks). Tony Lin and Pengwei Hao[5] proposed block based compound Image compression algorithm for real-time computer screen image transmission and classified image into picture and text/graphics blocks in two passes based on color count, shape and sizes. Shape primitives extracted from text/graphics blocks and picture block types are losslessly coded with shape-based and palette-based coding algorithm. Pictorial pixels are coded by lossy JPEG. So numerous coding algorithms are needed and basic shape primitives defined in this method is not flexible for fonts of different sizes.

Alexandre Zaghetto and Ricardo de Queiroz[6] proposed Compound document coding based on H.264/AVC-INTRA frame coding which classified the segmented text block into text interior, text border and background macro blocks and coded by H.264/AVC coder with variable quantization parameter called AVC-C. With AVC-C, for the same bit rate, there is not an overall objective gain over AVC-I, it is possible to improve significantly the quality of text regions, with a user controlled quality loss to pictorial regions. A. Zaghetto, R. L. de Queiroz,D. Mukherjee[7] presented an MRC compound document codec that uses H.264/AVC operating in INTRA used threshold-based layer segmentation algorithm and an iterative data-filling algorithm for redundant regions. Cuiling Lan Feng Wu Guangming Shi[8] proposed a coding scheme based on the H.264 intra-frame coding. Two new modes residual scalar quantization (RSQ) mode and base colors with index map (BCIM) mode as well as previous intra modes in H.264 are selected by the rate-distortion optimization (RDO) method in each block. This method improved the coding efficiency even more than 10dB for compound images with increase in the complexity of intra-frame coding.

H.264/AVC based coders [6-8] outperformed JPEG-2000 with increase in the complexity of encoder. But for storage and retrieval of compound images application context, encoder complexity is not very significant. The compression algorithm should keep visual quality of text regions with high compression ratios. Hence the proposed method employs H.264/AVC single coder with variable quantization for compressing both picture and text regions.

### 1.2 H.264/AVC-INTRA

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 advances 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 [9]. One of the components of these advances is the intraframe macro block 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 JPEG-2000 [10]. 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 JPEG-2000 are typically in the order of 0.25 to 0.5 dB in PSNR for pictorial images. However, the AVC-I seem to have an extra capacity of adapting itself to heterodox content. For compound images (mixed pictures and text) the PSNR gains are more substantial, even surpassing the mark of 3-dB improvement in some cases.

Image features, such as histogram, gradient and the number of colors, are often used for classification [11], [12]. Then different type blocks are compressed by different coding schemes to better adapt to their statistical properties [11]–[13]. Considering the sparse histogram distribution of colors in text/graphics blocks, a novel method is proposed to represent a text/graphics block by several base colors and an index map. Furthermore, Ding *et al.* develop this method as an intracoding mode and incorporate it into H.264[14]. Thus, the coding performance of that scheme on compressing compound images is significantly improved.

### 1.3 MOTIVATION AND CONTEXT

It is important to place our coder within the proper application context. We are concerned with a distributed scenario such as the web, where a document/compound image is stored once, somewhere, and many users are to retrieve and decode the document/compound image at a later time. In this scenario, the encoder has no complexity restriction. Indeed, as we will see later, we may use a large number of passes at encoding time and we apply segmentation algorithms, as well. These high-complexity operations are to be done off-line and do not affect the decoding process at all. Thus, our encoder complexity was not considered here in comparison to other encoders such as JPEG-2000.

Compound images have typically been compressed as a single image. However, different compression algorithms may be applied to each of the regions of the images. That is the way multiple-coder-based algorithms work. Instead of a multiple-code approach, this paper proposes a single-coder algorithm based on a modified version of the AVC-I that adjusts itself as an effort to encode text and pictorial regions differently.

This paper is organized as follows. In section 2, a detailed description of proposed algorithm including System, DCT based segmentation, coding of blocks is provided. Experimental results are presented in section 3. Finally, the paper is concluded in section 4.

## 2. PROPOSED METHOD

### 2.1 SYSTEM

As shown in Fig.1, the proposed algorithm consists of two stages: segmentation and coding. The algorithm first segments 4x4 non overlapping blocks of pixels into text/graphics and picture/background classes and then compresses both text/graphics pixels and picture/background pixels H.264/AVC with variable quantization parameter. Finally, the losslessly compressed text/graphics data, the lossy compressed picture/background data and losslessly coded block segmentation map are put into one bit stream and compressed losslessly in the encoder.

### 2.2 DCT BASED SEGMENTATION

In 4x4 DCT transform coding, the 16 transformed co-efficient are zig-zag ordered such that coefficients are arranged approximately in the order of increasing frequency. The DCT transform coefficients can be classified into two groups namely, DC and AC coefficients. The DC coefficient determines the average brightness in a block. All other coefficients describe the variation around this DC value and these are referred to as AC coefficients. In a DCT block, most of the energy is concentrated in the DC and very few AC coefficients. After analyzing various compound images like wallpapers, webpages and power point screen images, it is found that most of the energy of text/graphics pixel is concentrated in the 5 AC coefficients, and their location in DCT transform is specified as binary in the binary mask H.

The proposed block based segmentation procedure classifies each 4x4 non overlapping blocks of pixels into text/graphics, picture/background blocks by thresholding DCT energy computed from selected AC coefficients of block. Discrete cosine transform is applied to each 4x4 block and then DCT energy of block is determined from selected AC coefficients. Also for various test compound images it is found that, DCT energy of 5 AC coefficients in text/graphics block exceed 90. Hence threshold T1 is set as 90. If the energy of block go above threshold T1, the block is classified as text/graphics block. Otherwise, it is classified as picture/background block. Fig.2(b) shows the plot of DCT energy from selected AC coefficients in each 4x4 block of Fig.2(a) which clearly shows that the DCT energy of text block is very high (i.e. above 90) and DCT energy of picture/background blocks drop below T1. Hence the proposed segmentation procedure uses DCT energy as feature vector for block classification and also the block classification based on DCT energy can be extremely fast. The discrete cosine transform is of a N x M image *f(x, y)* is defined by,

$$F(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2M}\right] \quad (1)$$

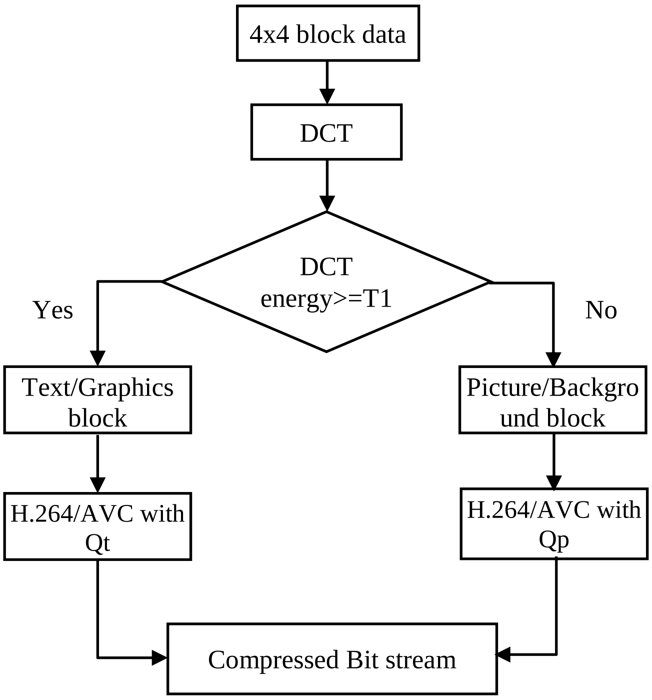


Fig.1 Flow chart of the proposed system


Fig.2(a). Part of Compound Image

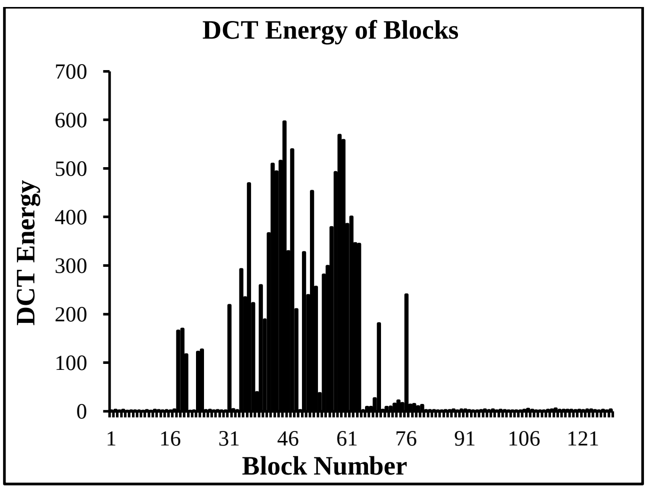


Fig.2(b). Plot of Block No Vs DCT Energy in each block

where,

$$\alpha(u)=\begin{cases}\frac{1}{\sqrt{N}}, & u=0\\ \sqrt{\frac{2}{N}}, & 1\le u\le N-1\end{cases}\qquad \alpha(v)=\begin{cases}\frac{1}{\sqrt{M}}, & v=0\\ \sqrt{\frac{2}{M}}, & 1\le v\le M-1\end{cases}$$

The inverse transform is defined by,

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{M-1}\alpha(u)\alpha(v)F(u,v)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2M}\right] \quad (2)$$

where, f(x, y) is the pixel value at the (x, y) coordinate position in the image, N, M be the height and width of window, F(u,v) is DCT domain representation of f(x,y) image. u, v represents vertical and horizontal frequencies. x,y,u,v have values from 0 to 3, N=4 and M=4. The classification is performed on 4x4 blocks of input images, so each block is represented by 16 coefficients (features) in the DCT domain. Out of 16 DCT coefficients, 5 AC coefficients are selected to form best feature vector using 4x4 binary masks H. DCT energy of block is determined using this feature vector and compared against a threshold T1 to segment text/graphics and picture/background regions.

$$H=\begin{bmatrix}0&0&0&1\\0&0&0&1\\0&1&0&1\\1&0&0&0\end{bmatrix}$$

$$F_s(u,v) = F(u,v).*H \quad (3)$$

$$Energy(B_i)=\sum_{u=1}^{N}\sum_{v=1}^{N}\left|F_s(u,v)\right|^2 \quad (4)$$

$$\text{type}(B_i)=\begin{cases}1, \text{text/ } graphics \text{ block} & \text{if Energy}(B_i) >= T1\\ 0, Picture/background \text{ block} & otherwise\end{cases}$$

where, $F(u,v)$ represent the DCT transform of each 4x4 block, $F_s(u,v)$ represent selected DCT coefficients by the mask $H$, Array multiplication $F(u,v).*H$ is the element-by-element product of the arrays $F(u,v)$ and $H$. $B_i$ denote $i^{th}$ block.

### 2.3 H.264/AVC-INTRA FOR COMPOUND IMAGES

H.264 /AVC allows for the change of quantizer parameter $Q$ at each macro block. Hence, there is a possibility to use less aggressive quantizer steps for text regions in order to keep edges sharp, while permitting high frequency losses to pictorial regions. In H.264/AVC, there are many decisions per block or macro block that the encoder has to make. Examples are the division of the macro block for motion estimation, the transform to be used (4x4 or 8x8), the intraframe block prediction mode, etc. These decisions are all made by rate-distortion considerations. Since the proposed segmentation classifies the compound image into 4x4 image blocks, 4x4 transform mode is set in H.264/AVC.

First, a region classification algorithm is applied to identify text and pictorial regions. It is assumed that, in these text regions, the viewer would pay greater attention to edges. In DCT based segmentation-driven image compression techniques, the image to be compressed is first segmented as shown in Fig.3(a), 3(b). As an example, Fig.2(a) shows a compound document which is input to the segmenter. The result of block classification is shown in Fig.3(a). Each block is classified as type 0, 1 and a coding mask is constructed. Each 4x4 block is classified either as class 0(picture regions) or class 1(text regions). Blocks of class 0 (pictorial regions) are composed exclusively by pixels marked as background. Class 1 block (text regions) are those composed exclusively by pixels marked as text. Fig.3(b) shows the coding mask of Fig.3(a). To make it easier to visualize, block classes 0, 1were represented as white, black respectively. The coding mask is passed on to a modified version of AVC-I, which will adapt the value of $Q$ for each 4x4 block, according to the class it belongs. The idea is to "transfer" quality of a text/graphics class to picture class. Class 0 regions are encoded with a quantizer parameter $Q_p$ while class 1 regions are encoded with a quantizer parameter $Q_t$, being $Q_t < Q_p$. This means that more compression is applied where there is picture/background, and less compression is applied to the text regions. The idea is to allow a single H.264 coder to compress more efficiently compound image.

## 3. EXPERIMENTAL RESULTS

The proposed algorithm is implemented on an Intel P-IV 2.66 GHz using MATLAB 7.0, and ten 800×600 true color screen images are tested, including four webpages and two power point screen images, and four wallpaper images. The test images are tested with three algorithms.


Fig.3(a). Result of Block classification

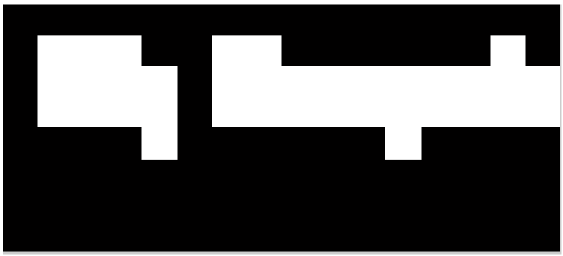

Fig.3(b). Binary coding mask

(JPEG[15], JPEG-2000[15]). DjVu[16] is the benchmark scheme of compound image compression and Any2DjVu web service is used to create DjVu files. The segmentation and reconstruction result of wallpaper image *wall1* is shown in Fig.4(a)-(d).

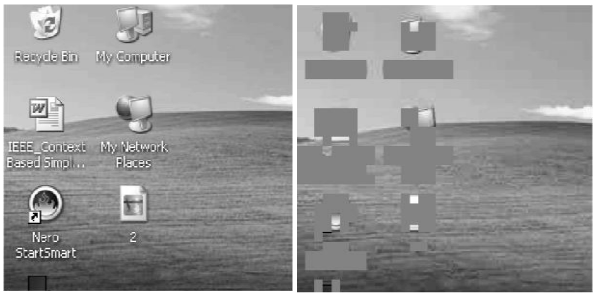


Fig.4(a) Fig.4(b)


Fig.4(c) Fig.4(d)

Fig.4(a)-(d). Segmentation and reconstruction result of the Wallpaper image *wall1*. (a) A portion of wallpaper image (b) Background layer of pictorial pixels, with holes filled with average colors of pictorial pixels in the block (c) Foreground layer of text/graphics pixels (d) Reconstructed image (27.16:1 compression)

Compression ratios and PSNR ratio for ten compound images are listed in Table.1, indicating that our algorithm attained the highest PSNR value for all images at similar compression ratio. PSNR plots for JPEG, JPEG-2000 and the proposed method are shown in Fig5. For the image *wall1*, proposed method clearly outperforms traditional algorithms such as JPEG, JPEG-2000 and AVC-I[17].

Table.1. Compression Ratio Vs PSNR of TEN compound images. The original file size of all images is 1403 kb

| **Image (800x600)** | **Proposed** | | **JPEG[15]** | | **JPEG2000[15]** | |
|---|---|---|---|---|---|---|
| | **Comp Ratio** | **PSNR (dB)** | **Comp Ratio** | **PSNR (dB)** | **Comp Ratio** | **PSNR (dB)** |
| web1 | 18.00 | 41.19 | 18.00 | 31.94 | 18.12 | 34.83 |
| web2 | 21.04 | 37.29 | 21.37 | 31.02 | 21.05 | 33.93 |
| web3 | 18.30 | 42.90 | 18.04 | 31.58 | 18.51 | 35.63 |
| web4 | 19.75 | 39.43 | 19.81 | 30.45 | 19.71 | 34.15 |
| ppt1 | 60.60 | 41.57 | 60.13 | 30.01 | 60.5 | 37.41 |
| ppt2 | 60.60 | 36.39 | 60.55 | 32.37 | 60.03 | 35.25 |
| wall1 | 37.45 | 42.54 | 37.24 | 37.18 | 37.02 | 37.55 |
| wall2 | 42.24 | 41.20 | 42.51 | 34.20 | 42.45 | 37.04 |
| wall3 | 31.53 | 41.03 | 31.78 | 33.95 | 31.29 | 36.89 |
| wall4 | 15.88 | 39.14 | 15.88 | 32.46 | 15.97 | 33.92 |

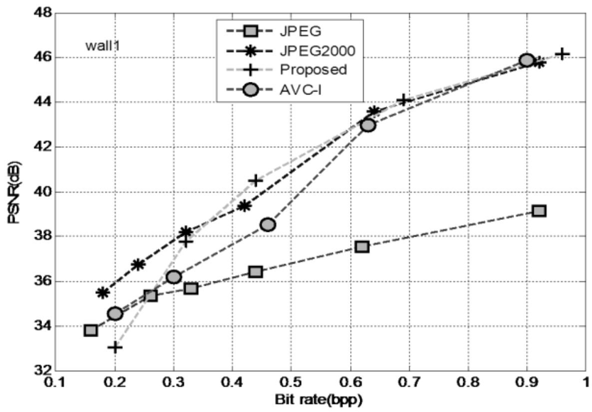


Fig.5. PSNR plots for JPEG, JPEG2000, and AVC-I and Proposed method for wall1

Fig.6 compares the visual quality of a test image compressed at 0.2 bpp using JPEG2000, H.264/AVC intra coding, DjVu and the proposed scheme. It can be easily observed that the overall quality of our scheme is better than that of the H.264/AVC intra coding. In particular, the text part coded with our scheme is free from ringing artifact, which can always be observed in traditional image coding methods. The proposed scheme also outperforms DjVu by better preserving the chrome components.

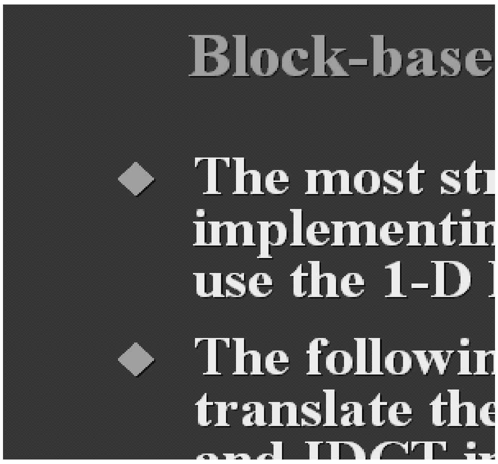


Fig.6(a). Input Image

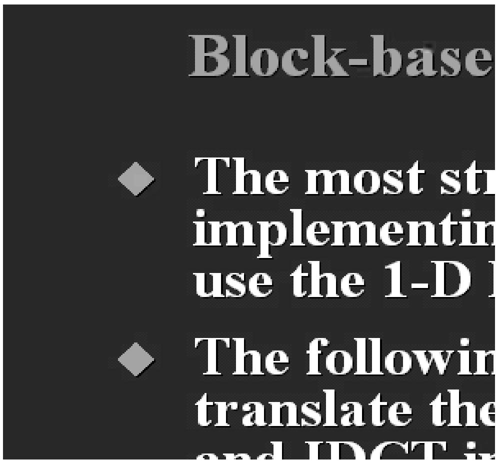


Fig.6(b). Proposed Scheme

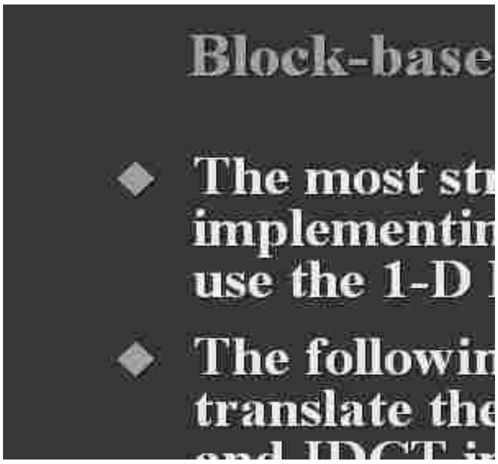


Fig.6(c). JPEG

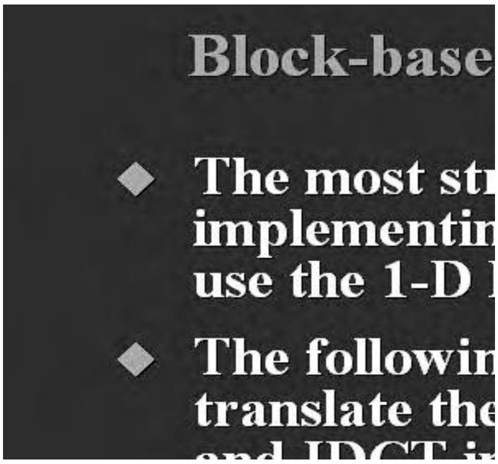


Fig.6(d). JPEG 2000

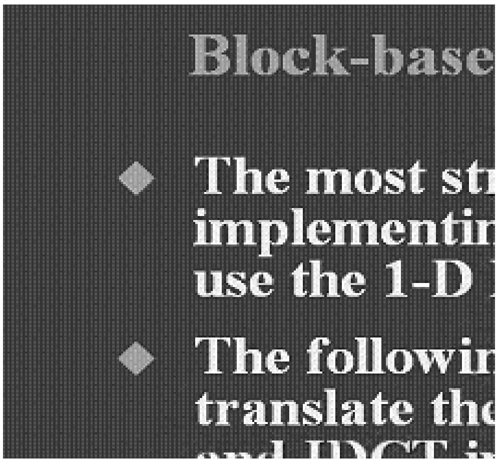


Fig.6(e). AVC-I

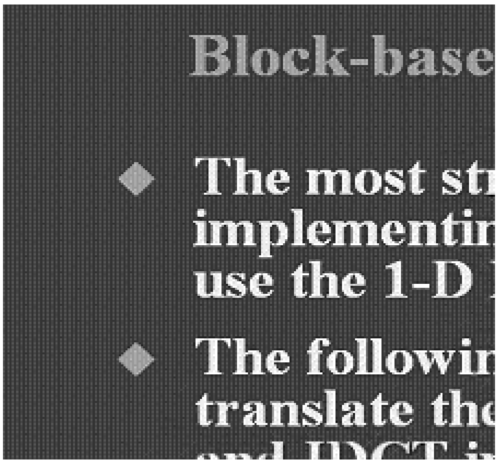


Fig.6(f). DjVu

Fig.6(a)-(f). Subjective performance i.e. visual quality comparison of proposed method, JPEG, JPEG 2000, DjVu and AVC-I for part of test image web1. (a) A portion of the original image, (b) Proposed method, 73 KB (19.3:1 compression), (c) JPEG, 76 KB (18.4:1 compression), (e) JPEG 2000,76.4 KB (18.3:1 compression), (e) DjVu, 84 KB (16.8:1 compression), (f) AVC-I, 84 KB (16.8:1 compression

The proposed H.264/AVC-I coder with variable quantization parameter outperforms JPEG and JPEG 2000 with an increase in encoder complexity. With this limitation, it is applicable for offline applications rather than online compound image compression.

## 4. CONCLUSION

Simplified DCT based segmentation driven H.264/AVC-I single coder for compound image compression algorithm is proposed. In this paper, simplified one pass fast segmentation algorithm is developed to separate image into text/graphics, picture/ background regions. DCT based segmentation algorithm minimizes misclassification error and provides significant PSNR value at similar compression ratio. Experimental results demonstrate that it segments the image accurately irrespective of text appearances and background complexity. It also provides excellent visual quality and competitive compression ratio. For the same bit rate, the proposed method significantly improves the quality of text regions, with little to negligible losses to the pictorial regions. In future work can be done to reduce the encoder complexity by minimizing the number of modes in H.264/AVC-I.

## REFERENCES


[1] "*Mixed raster content (MRC)*", ITU-T Recommendation T.44, Study Group-8 Contribution, 1998.

[2] P. Haffner, L. Bottou, P.G. Howard, P. Simard, Y. Bengio and Y. Le Cun, "High Quality document image compression with "DjVu"", *Journal of Electronic Imaging*, Vol. 7, No. 3, pp. 410-425, 1998.

[3] X. Li and S. Lei, "Block-based segmentation and adaptive coding for visually lossless compression of scanned documents", *Proceedings of International Conference on Image Processing*, Vol. III, pp. 450–453, 2001.

[4] H. Cheng and C. A. Bouman, "Document compression using rate-distortion optimized segmentation", *Journal on Electronic Imaging*, Vol. 10, No. 2, pp. 460–474, 2001.

[5] Tony Lin and Pengwei Hao," Compound Image compression for Real-Time Computer Screen Image Transmission", *IEEE Transactions on Image Processing*, Vol. 14, No. 8, pp.993-1055, 2005.

[6] Alexandre Zaghetto and Ricardo L. de Queiroz, "Segmentation-Driven Compound Document Coding Based on H.264/AVC-INTRA", *IEEE Transactions on Image Processing*, Vol .16, No.7, pp. 1755-1761, 2007.

[7] Zaghetto A, 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 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, pp.679-686, 2008.

[8] Cuiling Lan, Feng Wu and Guangming Shi, "Compress Compound Images in H.264/MPGE-4 AVC by Exploiting Spatial Correlation", *IEEE Transaction on Image Processing*, Vol. 19, No. 4, pp. 946-957, 2010.

[9] 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*, 2004.

[10] D. S. Taubman and M. W. Marcellin, "*JPEG2000: Image Compression Fundamentals, Standards, and Practice*", Norwell, MA: Kluwer, 2002

[11] Li. X and S. Lei, "Block-based segmentation and adaptive coding for visually lossless compression of scanned documents", *Proceedings of International Conference on Image Processing*, Vol. 1, pp. 219–223, 2009.

[12] W. Ding, D. Liu, Y. He, and F. Wu, "Block-based fast compression for compound images", *Proceedings of International Conference Multimedia and Expo (ICME)*, pp. 809–912, 2006.

[13] D. Mukherjee, C. Chrysafis, and A. Said, "Low complexity guaranteed fit compound document compression", *Proceedings of International Conference on Image Processing*, Vol. 1, pp. 225–228, 2002.

[14] W. Ding, Y. Lu, and F. Wu, "Enable efficient compound image compression in H.264/AVC intra coding", *Proceedings of International Conference on Image Processing*, Vol. 2, pp. 337–340, 2007.

[15] Gonzalez, R. C., Woods, R. E., and Eddins, S. L, "*Digital Image Processing Using MATLAB*", Prentice Hall, Upper Saddle River, NJ, 2004.

[16] Available at: http://Any2DjVu.djvuzone.org/

[17] Reference software of H.264/AVC, version jm14.0: downloadable at

http://iphome.hhi.de/suehring/tml/download/old_jm/