Comparison of H.264/Avc-Intra Technique for Compound and Natural Image Compression

V. Radha, *Member, IAENG*

Abstract – **Currently, the notion of paperless office is being promoted as part of eco-projects in many industries, where paper documents are converted into electronic documents. These images are termed as 'Compound images' and are defined as images that contain a combination of text, natural (photo) images and graphic images. The number of documents stored in electronic format is increasing enormously and this in turn, is increasing the need for efficient storage and transmission techniques. Different compression techniques are proposed for natural and compound images. Algorithms that work for natural images need not work well for compound images and vice versa. A single algorithm that suits both compound and natural images is beneficial in many applications including the Internet. In this paper, two such algorithms based on H.264/AVC Intra are selected and its efficiency in compressing both natural and compound images are analyzed and compared.**

Index terms— **Compound Image Compression, Natural Image Compression, H.264/AVC.**

I. INTRODUCTION

The current era of information explosion is envisaging tremendous growth in both communication medium and hardware growth. This in turn, is generating huge volume of digital signals in form of images, videos, audio, textual, which proves to be challenging in terms of storage and transmission. In spite of several breakthroughs in the price and performance of digital hardware and firmware, the demand for high data storage capacity and data-transmission bandwidth continues to outstrip the capabilities of available technologies (Raviraj and Sanavullah, 2007). According to ESG White Paper of ESG Strategy Group, the database archiving and storage volumes is going to increase by more than 11 per cent each year (Turner, 2008), which stresses the need to rethink the approaches available to accommodate future storage and information management. Data compression is a technology that is proposed as a solution to such situations. When this technology is applied to digital images, they are termed as digital image compression techniques.

Manuscript received July 10, 2011; revised August 12, 2011.
V. Radha is with the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore, Tamilnadu, PIN 641043 India. (phone: +91422-2404983; e-mail:radhasrimail@gmail.com).

The main objective of digital image compression algorithm is to reduce redundancy of the image data in order to be able to store or transmit data in an efficient form without degrading the quality of the original image. The reduction in file size allows more images to be stored in a given amount of disk or memory space. It also reduces the time required for images to be sent over the Internet or downloaded from Web pages. It is considered as a mature branch of image processing research area, where several efficient algorithms that produce good compression rates with adequate image quality have already proposed. These techniques focus on the characteristics of image data to produce good compression results.

Currently, the notion of 'paperless office' is being promoted as part of eco-projects in many industries, where paper documents are converted into electronic documents using applications like scan-to-print, document archiving, internet, fax and internet browsing. These images are often termed as 'Compound images' and can be defined as images that contain a combination of text, natural (photo) images and graphic images (Figure 1). The number of documents stored in electronic format is increasing enormously and this in turn, is increasing the need for efficient storage and transmission techniques (Feng and Bouman, 2006).

There are several different ways in which image files can be compressed. The two most common used algorithms are JPEG and the GIF. Another method that has gained popularity in recent decades is the PNG format. Wavelets and fractals are also widely used. Both methods offer higher compression ratios than JPEG and GIF for some types of images. However, the performance of all these techniques degrades while used for compressing compound images. The reason is that all these algorithms focus on one type of image data. For example, the JPEG method is more often used for photographs or natural images, while the GIF method is commonly used for line art and other images in which geometric shapes are relatively simple. Run Length Encoder performs well with text files. Thus, how to deal with different data types in a compound image is a challenging task faced by the compound image compression algorithms. A single algorithm that meets the requirement of all parts of the compound image is still elusive. An efficient algorithm that compresses document images should take into account the various requirements of the text / image and graphical portions of the image and should also detect and use the layout and structural information from the image during compression.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The solutions provided to the problem of compound image compression can be grouped into two categories. They are, layer-based coding (Mixed Raster Content (MRC), 1998; Haffner *et al.*, 1998) and block-based coding

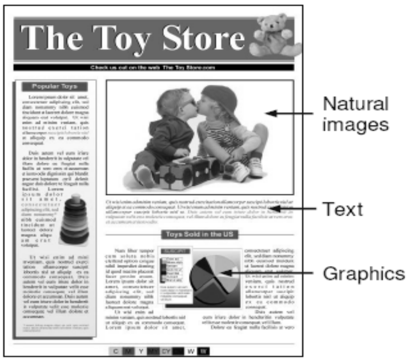

Figure 1 : Compound Image

(Said and Drukarev, 1999). Lin and Hao (2005). Layer based coding separates images into layers with binary masks and use separate techniques for each layer. Block-based approaches classify images into blocks of several types and use different coding methods for different types of blocks. These block-based approaches own very flexible coding structure. While these approaches show satisfactory coding performance on some compound images, their coding performances on purely natural image decrease, because the segmentation algorithms are not robust enough to adapt to both natural and compound images. Moreover, while applying these techniques to natural or photographic images, the compression performance degrades.

In the present scenarion, image processing applications are using different algorithms for compressing natural and compound images. As the most advanced applications like WWW use both natural and compound images in abundance, the current requirement is to have a single algorithm that can efficiently compress both natural and compound images efficiently. Recently, the idea of applying H.264/AVC video coding technique to compress both types of images was taken up several researchers recently (de Queiroz et al., 2005). 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 (Marpe *et al.*, 2004a and Marpe *et al.*, 2004b). One of the components of these advances is the intraframe macroblock prediction method, which, combined with the context-adaptive binary arithmetic coding (CABAC), turns the H.264/AVC into a powerful still image compression engine. There are basically two schemes used during compression process : (i) intra-only compression and (ii) inter compression.

As most of the advanced applications, like WWW, uses both natural and compound images extensively, the current need of the market is to have a single codec which will work efficiently for both natural and compound images. Zaghetto and de Queiroz (2007) and Ding *et al*. (2007) considered the problem of compressing compound images with a novel idea of using a single coder for compressing all the content types. The Ding model use a spatial domain representation for text blocks with base colors and color index map and a structure aware content-based arithmetic coder to compress color index map for compressing colour compound and natural images. The Zaghetto and Queiroz model, on the other hand, uses a segmentation driven adaptation method that modifies H.264 AVC model's quantization parameter on a macroblock by macroblock basis for compressing compound images. This paper is a study that analyzes and compares the performances of these systems while compressing natural and compound images. The remaining part of the paper is organized as below. Section II describes the working of the two models selected for analysis. Section III presents the experimental results while Section IV presents a brief summary along with future research direction.

II. COMPRESSION MODELS

This section discusses the two selected compression models. Both the models modify H.264/AVC video coder to compress natural and compound images. An H.264 video encoder carries out prediction, transform and encoding processes to produce a compressed H.264 bitstream. An H.264 video decoder carries out the complementary processes of decoding, inverse transform and reconstruction to produce a decoded video sequence. In this paper, two models based on H.264/AVC Intra coding are considered for compressing both natural and compound images. The general steps followed by both the coders are outlined below.

Step 1 : Color transformation from RGB image to YUV image

Step 2 : Separate text and picture block

Step 3 : Apply H.264/AVC coding algorithm to picture block

Step 4 : Apply appropriate text block encoding techniques (either Ding model or Queiroz model)

Step 5 : Display results

A.Color Space Transformation

The two most widely used colour spaces for storing digital images are RGB colour space and YUV colour space. YUV stores more relevant data at a lower accuracy than RGB. Moreover, it is well known that the RGB components of color images are highly correlated and if the transforms of each color component is obtained, the transformed components will also be highly correlated (Nobuhara and Hirota, 2004). The equation used for transforming the RGB components to YUV space is given in Equation 1.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.200 & 0.587 & 0.114 \\ -0.418 & -0.289 & 0.437 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \quad (1)$$

B..H.264/AVC algorithm

H.264 is an industry standard for video compression, the process of converting digital video into a format that takes up less capacity when it is stored or transmitted. This section explains the algorithm as outlined by Richardson (2009). The Figure 2 shows the encoding and decoding processes and highlights the parts that are covered by the H.264 standard.

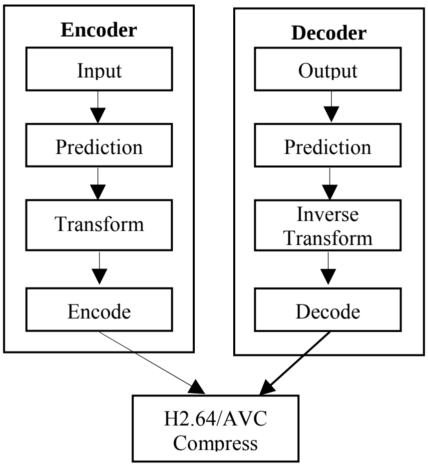

Figure 2 : Encoder and Decoder

An H.264 video encoder carries out prediction, transform and encoding processes to produce a compressed H.264 bitstream. An H.264 video decoder carries out the complementary processes of decoding, inverse transform and reconstruction to produce a decoded video sequence.

C.DING Model

Ding *et al*. (2007) presented an efficient compound image compression approach based on H.264/AVC intra coding. The input image in this model is divided into two regions, namely, text block and picture block. A new coding scheme to extract and compress the text block is proposed. At the base, the H.264/AVC intra mode is used to compress the compound image. The working of the algorithm is explained below.

- Coding of Text Blocks

As text region exhibits features that vastly differ from picture block and similar structures can be observed, this model represents text block in spatial domain. The text blocks are described using base colors and index map (structure information). A color quantization algorithm is used to decompose a text block, followed by a context-adpative coding algorithm. Both these processes are explained below.

- Color Quantization

Quantization methods like vector quantiation, K-means, have the disadvantage that the entropy of the color-quantized image is high which makes the compression process difficult. To solve this, the Ding Model also considers the number of bits for compression during color quantization. Since base colors cost much less bits than the index map, the entropy of a text block usually depends on the index map. To reduce the entropy, a two-step quantization method is proposed to achieve the rate and distortion trade-off.

Step 1 : First, a local quantization is performed to cluster the neighboring similar pixels into a group, which guarantees that neighboring pixels tend to be quantized to the same base color. As a result, the disorder of the index map will be reduced. In particular, given the allowed maximum distortion among neighboring pixels, all neighboring pixels with distance under a threshold are clustered to the same group, which is represented by their average value.

Step 2: Second, the block after local quantization is further quantized to several base colors. Note that the pixels in the same group are quantized to the same color. The number of base colors of a block depends on the content. Instead of quantizing each block to a fixed number of colors, the allowed maximum distortion is set to a constant value, e.g. $q^2/4$, where q is the quantization step used in H.264/AVC intra coding. Thus, the number of base colors of a 16x16 macroblock may vary from 1 to 8. A Tree Structure Vector Quantization (TSVQ) method is used, where each pixel is treated as a vector. The maximum distortion is the criterion to split a tree in TSVQ.

Both base colors and the index map are compressed with context-adaptive arithmetic coder. The YUV components of a base color are first quantized. The index map shows similar patterns among text blocks. The contexts and the remapping are used to exploit the similar patterns to enhance the compression. When coding the index map, the indices are arranged in scan line order. The context for an index to be coded is deduced from its neighboring index values. The current index is then mapped to a symbol according to its context and neighboring indices, and the symbol is coded with an arithmetic coder using the context.

In particular, 15 contexts are defined according to their four neighboring indices (i.e. left, left-top, top, right-top). As shown in Figure 3, each shape represents a neighboring index, while the different shapes indicate the different index values. The 15 contexts are classified into 5 categories. The contexts of the same category have the same number of different color indices. The contexts in Figures 3(a) and (b) indicate that the number of identical neighboring color indices is four and three, respectively. The contexts in Figure 3(c) indicate that there are two pairs of identical neighboring indices. The contexts in Figure 3(d) indicate that only two identical neighboring indices exist. When all neighboring indices are different, the context in Figure 3(e) is used.

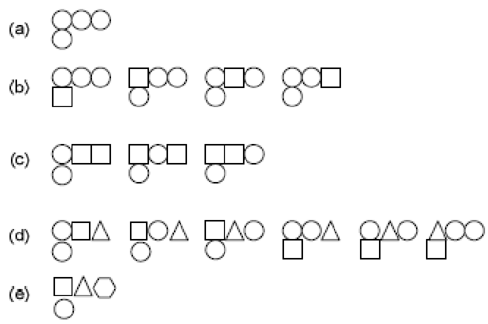

Figure 3: Contexts for the coding of index map in text block

The index values represented by the symbols (e.g., “O”) will be remapped. The reason is that, in the same context, the probability distributions of indices still depend on the actual values of neighboring indices. Generally, neighboring index values have higher probabilities than other values to occur in the current index. The indices are remapped with high probability to small values. Taking the context of type (b) as an example, the index value of “O” has the highest probability in occurrence. This index value is then remapped

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

to 0, which makes the compression of the current index more efficient. The detailed mapping methods for all cases are shown in Table I.

TABLE I : Mapping methods for all the contexts

Context Category	Mapping
(a)	○→0
(b)	○→0 □→1
(c)	○→0 □→1
(d)	○→0 □→1 △→2
(e)	○→0 □→1 △→2 ⬡→3

Moreover, other twenty-two contexts are developed for boundary indices, whose neighboring indices are partially unavailable. These boundary indices are compressed with the similar method.

- Compound Image Compression

The above explained text block coding procedure is incorporated into H.264 intra coding as a new mode to compress compound images with both picture content and text content efficiently. H.264/AVC intra coding is adopted in the existing scheme for its high coding efficiency on natural images. Further, the mode selection algorithm is adopted based on rate distortion optimization (RDO) in H.264/AVC reference software to distinguish the text blocks. The RDO-based mode selection can balance the quality of picture content and text content in terms of mean square errors. It can also select the mode adaptively according to the targeted bit-rate.

D.QUEIROZ MODEL

In the work proposed by Zaghetto and de Queiroz (2007) the H.264/AVC operating in intraframe mode is used to compress a compound image. Here, the text and picture block are separated by taking distortion into account. This approach used a segmentation-driven adaptation strategy to change the H.264/AVC quantization parameter on a macroblock by macroblock basis, i.e. bits are deviated from pictorial regions to text in order to keep text edges sharp.

- Segmentation-Driven Rate Allocation

In this model, the analysis on a macroblock by macroblock basis is adapted and the steps followed are given in this section. First, a region classification algorithm that identifies text and pictorial regions is applied. This classification algorithm is derived from an edge detector and identifies edges belonging to text as opposed to textures. Fan's Text Segmentation algorithm is used for this purpose (Fan, 2003). The next step is to classify each macroblock (MB) of 16x16 pixels block size. The binary image containing the segmented text is analyzed and each MB is classified as type 0 (picture region), 1 (text regions) or 2 (mixer of picture and text region) and finally, a coding mask is constructed. The coding mask is passed on to a modified version of AVC-I, which will adapt the value of Q_p for each MB, according to the class it belongs. The idea is to "transfer" quality of a MB class to another. Class 0 and 1 regions are encoded with a quantizer parameter Q_p, while class 2 regions are encoded with a quantizer parameter Q_{pText}, being $Q_{pText} < Q_p$. This means that more compression is applied where there is texture, and less compression is applied to the text letter borders. This algorithm is referred to as H.264/AVC-INTRA Compound, or simply AVC-C. Alternatively, it can be used to encode a region of interest (ROI) differently from the rest. The idea is to allow a single H.264 coder to compress compound images more efficiently.

The quality of the text edges is increased by lowering the quality of the picture and text interior regions so that they appear sufficiently. This decrease in quality has to be adjusted in such a way that it does not reduce the overall quality of the image. The Q_p and Q_{pText} selection algorithm used here is explained below.

- A bitrate R is chosen.
- A bitrate variation δr around R is set.
- Among all possible (Q_p, Q_{pText}) combinations, those which present bitrates inside the interval R ± δr are selected.
- Among all selected combinations, the maximum PSNR value, PSNRmax, is determined.
- A PSNR variation δq is set, and a minimum PSNR value,
  $PSNR_{min} = PSNR_{max} - \delta q$, is calculated.
- Among all selected (Q_p, Q) in step 3, those whose PSNR values are greater than $PSNR_{min}$ are chosen as candidates.

Select the candidate with the largest $d = Q_p - Q_{pText}$.

III. EXPERIMENTAL RESULTS

In order to verify the performance of the selected compression schemes, experimental evaluations were performed using a series of different benchmark images. For experimentation, 8 colour images (Figure 4) were selected (4 natural and 4 compound images). All the images were of size 256x 256. Three quality factors, namely, compression ratio, peak signal to noise ratio and compression and decompression time were used to assess the performance of the selected compression models. The selected systems were compared with the standard JPEG compressor.

A. Compression Ratio

The degree of data reduction obtained as a result of the compression process is known as the compression ratio. This ratio measures the quantity of the compressed data in comparison to the quantity of the original data. Table 4.1 and Figure 4 shows the consequence of compression on pure picture images, text images, hybrid images and full compound colour images. Both the models were also compared with the standard compressor JPEG 2000.

From Table II, it is clear that both the models outperform the standard model with respect to compression ratio. Both models compress both natural and compounds images efficiently. However, the Ding model is better than Queiroz model. On average, the Ding model is has 10% compression efficiency gain than Queiroz model.

NA-1 NA-2 NA-3 NA-4

Natural Images

CI-1 CI-2 CI-3 CI-4

Compound Images

Figure 4: Test Images

TABLE II : Compression Ratio

IMAGE NAME	DING MODEL	QUEIROZ MODEL	JPEG 2000
NA-1	47.28	43.23	41.71
NA-2	48.61	42.34	40.89
NA-3	48.25	43.01	41.45
NA-4	49.62	42.16	39.45
CI-1	51.23	46.52	42.85
CI-2	59.12	49.19	44.63
CI-3	47.66	45.18	39.99
CI-4	49.91	47.76	42.22

B. Compression and Decompression Time

Compression and decompression time are the basic measurements used to evaluate an image compression algorithm. Compression and decompression time denotes the time taken for the algorithm to perform the encoding and decoding processes respectively. All the experiments were conducted using a Pentium IV machine with 512 MB RAM. Table III shows the compression and decompression time taken by the two selected models.

TABLE III : Compression and Decompression Time (Seconds)

Image Name	DING MODEL		QUEIROZ MODEL		JPEG 2000	
	CT	DT	CT	DT	CT	DT
NA-1	0.69	1.04	0.64	0.77	0.78	0.91
NA-2	0.85	1.32	0.41	0.5	0.66	0.71
NA-3	0.74	1.12	0.56	0.76	0.81	0.82
NA-4	0.46	0.63	0.78	0.79	0.78	0.87
CI-1	0.89	0.91	0.46	0.68	0.54	0.06
CI-2	0.19	0.98	0.42	0.94	0.68	0.31
CI-3	0.85	0.88	0.63	1.04	0.77	1.05
CI-4	0.54	0.98	0.46	1.03	0.49	0.72

CT – Compression Time; DT-Decompression Time

The results show that Queiroz Model is the fastest among both type of images. The Queiroz Model took less than 0.91 seconds for natural images and 1.14 seconds for compressing compound images. The Ding model took less than 1.33 seconds, 1.38 seconds for compressing natural and compound images respectively. Both the models outperformed JPEG model with more than 18.9% and 31.4% time efficiency. These results prove that the models are highly suitable for applications like Internet, where time is considered very important.

C. Peak Signal to Noise Ratio (PSNR)

Peak Signal to Noise Ratio ratio is often used as a quality measurement between the original and a compressed image. A high PSNR indicates a better quality of the compressed or reconstructed image. Figure 5 shows the PSNR values obtained during experimentation while testing the models with 8 test images.

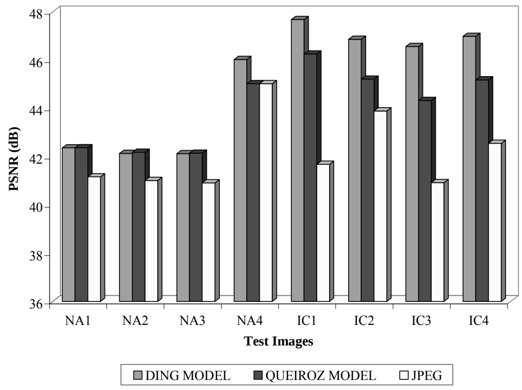

Figure 5 : Evaluation based on PSNR (dB)

While comparing the results obtained, it was noticed the quality of images after decompression was more or less same with natural images. However, for compound images, the quality of the image produced by Ding Model, was higher when compared with the other two models. However, both models outdistanced JPEG200 for picture and compound images. The average PSNR value obtained for Ding Model, Queiroz Model and JPEG are 45.08dB, 44.08dB and 42.14dB respectively.

The high PSNR obtained during experimentation clearly indicates that the proposed algorithm reduces the lossiness introduced by compression. According to the report of Schneier and Abdel-Mottaleb (1996), a PSNR value in the range 30-40 indicates that the resultant image is a very good match to the original image. In accordance with this report, the results of the proposed compression algorithm with both the models produce PSNR values in the range 42-47dB proving that it is an enhanced version for compression compound images.

IV.CONCLUSION

Compression concerns itself with the reduction of redundant or less perceivable information. In the present information explosion, a single algorithm that is suited for both natural and compound image is desired. This paper investigates two innovative techniques for compressing color still images using H.264 AVC/Intra that will work for both natural / photographic images and compound images. The general methodology followed is the same in both the models. Initially, a color transformation is applied to convert the RGB color space to YUV color space. In the next step, the text and picture portion in the image is separated into two blocks. The picture block is encoded using the standard H.264/AVC algorithm. In order to verify the performance of the selected compression schemes, experimental evaluations were performed using a series of different benchmark images, which was a combination of natural and compound images. The results from the various experiments conducted showed that the both the models outperform the standard model with respect to compression ratio, time and decompressed image quality. From the results it could be concluded that applying H.264/AVC Intra to compound image compression can be considered as a success. The main advantage obtained is the avoidance of segmentation of the picture into various blocks and using different coding algorithms for each block. Both the models considered in this paper, considers background as part of a picture. Efficient segmentation technique that separates background and picture can be incorporated and a separate compression technique that suits each region can be applied and analyzed.

REFERENCES

[1] de Queiroz, R.L. (2005) Compressing Compound Documents, Document and Image Compression Handbook, edited by M. Barni, Marcel-Dekker.

[2] Ding, W., Lu, Y. and Wu,F. (2007) Enable Efficient Compound Image Compression in H.264/AVC Intra Coding, IEEE International Conference on Image Processing, Vol. 2, Pp. II - 337 - II – 340.

[3] Fan, J. (2003) Text extraction via an edge-bounded averaging and a parametric character model", Proc. of the SPIE Document Recognition and Retrieval X, Santa Clara, CA, USA , Vol. 5010, Pp. 8-19.

[4] Feng, G. and Bouman, C.A. (2006) High-quality MRC document coding, IEEE Trans. on Image Processing, Vol. 15, No. 10, Pp. 3152-3169.

[5] Haffner, P., Bottou, L., Howard, P.G., Simard, P., Bengio, Y. and LeCun, Y. (1998) High Quality document image compression with DjVu, Journal of Electronic Imaging, Pp. 410-425.

[6] Lin, T. and Hao, P. (2005) Compound image compression for real-time computer screen image transmission, IEEE Transactions on Image Processing, Vol. 14, No. 8, Pp. 993-1005.

[7] Marpe, D., George, V. and Wiegand, T. (2004a) Performance comparison of intra-only H.264/AVC HP and JPEG2000 for a set of monochrome ISO/IEC test images, Contribution JVT ISO/IEC MPEG and ITU-T VCEG, JVT M-014.

[8] Marpe, D., George, V., Cycon, H.L. and Barthel, K.U. (2004b) Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode, Wavelet Applications in Industrial Processing, Proceedings of the SPIE, Vol. 5266, Pp. 129-137.

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

[10] Nobuhara, H. and Hirota, K. (2004) Color image compression/reconstruction by YUV fuzzy wavelets**,** IEEE Annual Meeting of the Fuzzy Information, Vol.2, Pp. 774 – 779.

[11] Raviraj, P. and Sanavullah, M.Y. (2007) The Modified 2D-Haar Wavelet Transformation in Image Compression, Middle-East Journal of Scientific Research, Vol. 2, No. 2, Pp. 73-78.

[12] Richardson, I.E.G. (2009) Video compression design, analysis, consulting and research, Vcodex White Paper**,** http://www.vcodex.com/h264overview.html**,** Last Access Date : 09-07-2009**.**

[13] Said, A. and Drukarev, A. (1999) Simplified segmentation for compound image compression", Proceeding of International Conference on Image Processing, Pp. 229-233.

[14] Schneier, M. and Abdel-Mottaleb, M. (1996) Exploiting the JPEG compression scheme for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell., Vol.18, No. 8, Pp. 849–853.

[15] Turner, M.J. (2008) IBM Information Infrastructure Initiative Aims to Tame the Information Explosion, White Paper, Enterprise Strategy Group, Pp.1-10.

[16] Zaghetto, A. and de Queiroz, (2007) Using H.264/AVC-Intra for Segmentation-Driven Compound Document Coding, IEEE ICIP 2007, Pp. II-333 – II-336.