

# IMPROVED BLOCK BASED SEGMENTATION FOR JPEG COMPRESSED DOCUMENT IMAGES

P.Mahesh<sup>1</sup>, P.Rajesh<sup>2</sup>, I.Suneetha<sup>3</sup>

<sup>1</sup>M.Tech Student, <sup>2</sup>Asst Professor, <sup>3</sup>Assoc Professor, Annamacharya Institute of Technology and Sciences, AP, India,  
p.maheshreddy169@gmail.com, rajshjoyful.smily@gmail.com, iralasuneetha.ait@gmail.com

## Abstract

Image Compression is to minimize the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. The compound image compression normally based on three classification methods that is object based, layer based and block based. This paper presents a block-based segmentation. for visually lossless compression of scanned documents that contain not only photographic images but also text and graphic images. In low bit rate applications they suffer with undesirable compression artifacts, especially for document images. Existing methods can reduce these artifacts by using post processing methods without changing the encoding process. Some of these post processing methods requires classification of the encoded blocks into different categories.

**Keywords-** AC energy, Discrete Cosine Transform (DCT), JPEG, K-means clustering, Threshold value

\*\*\*

## 1. INTRODUCTION

JPEG is one of the most widely used block based image compression technique. Although JPEG was first developed for natural images, in practice, it is also used for document images due to its simple structure. However, document images encoded by the JPEG algorithm exhibit undesirable compression artifacts. In particular, ringing artifacts significantly reduce the sharpness and clarity of the text and graphics in the decoded image. Many decoding schemes have been proposed to improve the quality of JPEG encoded images. Such methods neither require modification to current encoders, nor an increase in bit rate. Compression of a compound image is more critical than natural image. Basically a compound image is a combination of text, graphics and pictures. Most methods use different algorithms for various functionalities. Compound image compressions normally based on the following categories: Object Based, Layer Based and Block Based

Most of the recent researches in this field mainly based on either layer based or block based. In object Based method a page is divided into regions, where each region follows exact object boundaries. An object maybe a photograph, a graphical object, a letter, etc. The main drawback of this method is its complexity. In layer-based method, a page is divided into rectangular layers. Most of the Layer based methods follow the 3-layer MRC model. The basic three layers MRC model represents a color image as two color-image layers [Foreground (FG), Background (BG), and a binary Mask layer (mask)]. The mask layer describes how to reconstruct the final image from the FG/BG layers. All these methods first divides the JPEG encoded image into different classes of

regions and each region is decoded by using algorithm specially designed for that class of region. This region selection is done on the block basis, i.e. each encoded block is first classified as one of the required region. This paper focuses on jpeg coded document images. In a discriminative function based on DCT coefficients is calculated for every block. This function value is compared with a threshold parameter to classify the block. Discriminative function is defined as follows:

$$D_f(s) = \sum_{u=0}^7 \sum_{v=0}^7 g[y_s(u,v)] \quad (1)$$

$$\text{Where } g_i(x) = \begin{cases} \log_2(|x|) & \text{if } |x| \neq 0 \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

$y_s(u,v)$  is DCT coefficients of blocks 's'

$D_f(s) \approx 0 \Rightarrow$  's' is a background block

if  $D_f(s) < T \Rightarrow$  's' is a picture block

if  $D_f(s) \geq T \Rightarrow$  's' is a Non picture block

Threshold value proposed in above segmentation algorithm is not robust to different quantization tables used in baseline JPEG. Threshold value proposed gives good results, when default quantization matrices mentioned by JPEG standard are used. But when user defined Quality factor (QF) changes, obviously quantization matrices will change. It leads to

misclassification of the blocks. In for detection of background blocks AC energy thresholding is used. Next two feature vectors are developed for remaining each block. Authors characterized the feature vectors of the text blocks and those of the picture blocks by two Gaussian mixture models. Authors used these two Gaussian mixture models with the SMAP segmentation algorithm to segment the feature vector image. The result is combined with the background blocks detected by AC thresholding to produce the final segmentation map. This segmentation algorithm is robust to different quantization tables used in JPEG, but computational complexity is high.

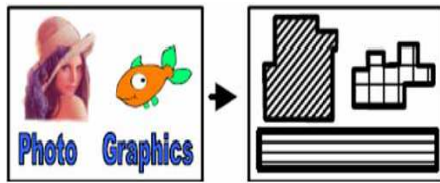
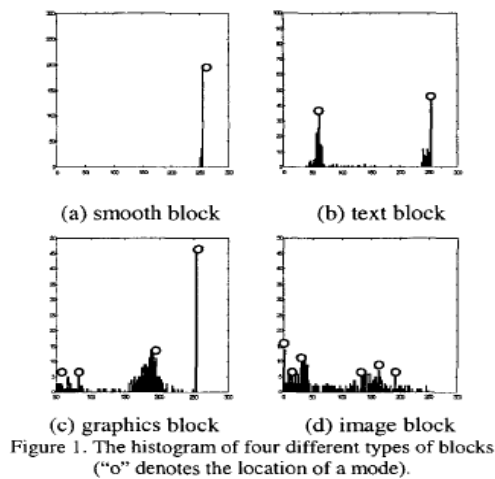


Fig. 1: Block based segmentation

In this paper propose a block based segmentation algorithm for document images encoded by conventional JPEG encoder. Our algorithm works by using AC energy thresholding for background blocks segmentation and for remaining blocks classification K-means clustering of feature components generated for each block.



## 2. PROPOSED MODEL

The first model uses the AC coefficients introduced during DCT to segment the image into three blocks, background, text /graphic and image blocks. The background block has smooth regions of the image. While the text / graphics block has high density of sharp edges regions and image block has the non-smooth part of the compound image.

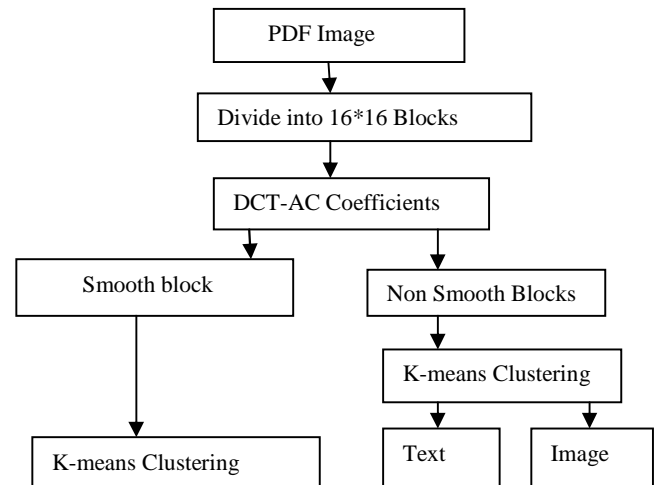


Fig 2: AC Coefficient Based Model

AC energy is calculated from AC coefficients and is combined with a user-defined threshold to identify the background block initially. It will consider  $8 \times 8$  luminance blocks as input and classifies them as background or picture or text block. AC energy ( $E_s$ ) of each block's is calculated as:

$$E_s = \sum_{i=1}^{63} y_{s,i}^2 \quad (2)$$

Where,  $y_{s,i}$  is the  $i^{\text{th}}$  quantized DCT coefficient of the  $8 \times 8$  block  $s$ . If  $E_s$  is smaller than the threshold value, then that block is classified as a background block. Next, we compute two-dimensional feature vectors for each block in order to classify the remaining blocks into the text and picture classes. The first feature component is based on the encoding length as proposed in . The encoding length of a block is defined as the number of bits in the JPEG stream used to encode that block. Typically, text blocks consists of high contrast edges as compared to other blocks, so encoding lengths for text blocks are longer than for the non text blocks. However, the encoding length also depends on quantization step sizes used. Larger the quantization steps, smaller the encoding length and vice versa. As from codec of Baseline-JPEG quantization step sizes are depends up on the user defined Quality Factor (QF). Quantization step sizes are inversely proportional to QF. So smaller the QF, smaller the encoding length and vice versa. To make the segmentation more robust to different quantization step sizes, we multiply the encoding length by a factor ( $\lambda$ ), determined from the quantization steps used in encoding of JPEG. Suppose  $Q_i^*$  are the default luminance quantization step sizes as defined in Table K.1 in, and  $Q_i$  are the quantization step sizes used to encode the luminance component. Quantity  $\lambda$  is used as a measure of the coarseness of the quantization step sizes  $Q_i$  as compared to the default

$Q_i^*$ . Larger quantization step sizes ( $Q_i$ ) corresponds to larger values of  $\lambda$  and vice versa.

$$\lambda = \frac{\sum_i (Q_i Q_i^*)}{\sum_i (Q_i^* Q_i^*)} \quad (3)$$

First feature component of the blocks is defined as:

$$D_{s,1} = \lambda^\gamma \times (\text{Encoding length of blocks}) \quad (4)$$

Threshold value used for AC thresholding and  $\gamma$  used in  $D_{s,1}$  are referred from [1]. This value are determined from training process and then uses a set of training images consisting of fifty four digital and scanned images. Each image is manually segmented and JPEG encoded with nine different quantization matrices, corresponding to  $\lambda_j$  with  $j=1$  to 9. For the  $i^{\text{th}}$  image encoded by the  $j^{\text{th}}$  quantization matrix, they first computed the average encoding lengths of the text blocks and the picture blocks, denoted by  $u_{i,j}$  and  $v_{i,j}$  respectively. The parameter  $\gamma$  is then determined from the following optimization problem:

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \min_{u,v} \sum_{i=1}^{54} \sum_{j=1}^9 [(\lambda_j^\gamma u_{i,j} - u)^2 + (\lambda_j^\gamma v_{i,j} - v)^2] \quad (5)$$

where  $u, v$  are average encoding lengths of the text blocks and the picture blocks by using default quantization step size as defined in Table K.1 in [1]. It is determined as  $\gamma=0.5$  and threshold value as 200 ( $\text{ac } \epsilon=200$ ).

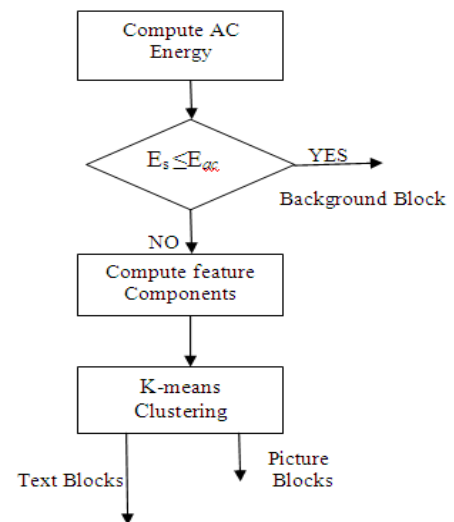
Due to multiplication of  $\lambda$  with the encoding length, effect of quantization step sizes on  $D_{s,1}$  will be reduced. Because, larger quantization step sizes leads to small values of encoding length and large values of  $\lambda$ , but the product  $D_{s,1}$  remains same. Similarly, for smaller quantization step sizes the same will occur. So the range of  $D_{s,1}$  for a particular class of block does not change with the quantization step sizes or QF. It depends only on the class of blocks and value is high for text class blocks as compared to remaining blocks. Generally text block is a combination of two colors foreground intensity and background intensity. So to take this advantage second feature component ( $D_{s,2}$ ) is developed to find how close a block is to being a two color block. We take the luminance component decoded by the convolutional JPEG decoder and use k-means clustering [6] to separate the pixels in a  $16 \times 16$  window centered at the block  $s$  into two groups. Let  $C_{1,s}$  and  $C_{2,s}$  denote the two cluster means. Then second feature component is computed by the difference of two cluster means, given in Eq.(6)

$$D_{s,2} = |C_{1,s} - C_{2,s}| \quad (6)$$

If difference of two cluster means is high, value of  $D_{s,2}$  is large, means block is a two color block. Otherwise block is a single color block. As explained above, text block is combination of two colors, so obviously  $D_{s,2}$  value is high for text class blocks as compared to remaining blocks. Compute vector  $D_s$ , for every non background block is defined sum of two feature components as:

$$D_s = D_{s,1} + D_{s,2} \quad (7)$$

As explained above for text blocks  $D_{s,1}$  and  $D_{s,2}$  are high when compared to picture blocks. So their sum  $D_s$  is also high value for text blocks compared to picture blocks. So we divide this vector  $D_s$  into two clusters by using k-means clustering and the resulting clusters with high mean value are segmented as text blocks and remaining blocks are named as picture blocks.



**Fig. 3: K-means Clustering Analysis**

K-means clustering is a method of cluster analysis, which partitions the given  $n$  observations into required  $k$  clusters such that each observation belongs to the cluster with the nearest mean. So for a given a set of observations ( $x_1, x_2, \dots, x_n$ ), k-means clustering aims to minimize the sum of squares within each cluster  $S = \{S_1, S_2, \dots, S_k\}$ :

$$\arg \min_s \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (8)$$

Where  $\mu_i$  is the mean of all points in  $S_i$ .

### 3. SIMULATION RESULTS

Proposed algorithm performance can be evaluated by taking seven different images. Each of these images contains some text and/or graphics. Segmentation algorithm results for all test images encoded by baseline JPEG at different compression ratios. Consider a test image-1 as in Fig. 1. Segmentation results of JPEG encoded test image-1 at QF values of 10, 25, and 60 are shown below in Fig. 4, Fig. 5 and Fig. 6 respectively. In resultant images, red color blocks represent the text region, yellow color blocks represent picture region and white color blocks represent background blocks. For all QF values, proposed segmentation gives almost same classification of regions.



Fig. 4: Original Test Image-1



Fig. 5: Segmentation Results (QF=10)



Fig. 6: Segmentation Results (QF=25)



Fig. 7: Segmentation Results (QF=60)

### CONCLUSIONS

Block based segmentation algorithm for JPEG encoded document images has been successfully implemented by using MATLAB software. Visual inspection of the simulation results indicate that proposed algorithm is robust to different quantization step sizes used in JPEG encoding. This paper considers true color images only. Block segmentation techniques can also be extended to gray scale images by small changes in the algorithm and take it as other image formats. The future work is the development of segmentation algorithm for documented images input take it as other image formats exclude JPEG.

### REFERENCES

- [1] E. Y. Lam, "Compound document compression with model-based biased reconstruction," J. Electron. Imag., vol. 13, pp. 191–197, Jan. 2004.
- [2] Tak-Shing Wong, Charles A. Bouman, Ilya Pollak, and Zhigang Fan, "A Document Image Model and Estimation Algorithm for Optimized JPEG Decompression," IEEE Transactions On Image Processing, vol. 18, no. 11, November. 2009.
- [3] B. Oztan, A. Malik, Z. Fan, and R. Eschbach, "Removal of artifacts from JPEG compressed document images," presented at the SPIE Color Imaging XII: Processing, Hardcopy, and Applications, Jan. 2007.
- [4] L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. Lecun, "High quality document image compression with DjVu," J. Electron. Imag., vol. 7, pp. 410–425, 1998.
- [5] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," IEEE Trans. Image Process., vol. 3, no. 2, pp. 162–177, Mar. 1994.
- [6] J. McQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Mathematical Statistics and Probability, pp. 281–297, 1967.
- [7] ISO/IEC 10918-1: Digital Compression and Coding of Continuous Tone Still Images, Part 1, Requirements

- and Guidelines, International Organization for Standardization 1994.
- [8] G. K. Wallace, "The JPEG still picture compression standard," IEEE Transactions on Consumer Electronics, vol. 38, no. 1, pp. 18-34, Feb. 1992.
  - [9] K. Konstantinides and D. Tretter, "A JPEG variable quantization method for compound documents," IEEE Trans. Image Process., vol. 9, no. 7, pp. 1282-1287, Jul. 2000.
  - [10] Averbuch, A. Schclar, and D. Donoho, "Deblocking of block-transform compressed images using weighted sums of symmetrically aligned pixels," IEEE Trans. Image Process., vol. 14, no. 2, pp. 200-212, Feb. 2005.
  - [11] H. Siddiqui and C. A. Bouman, "Training-based descreening," IEEE Trans. Image Process., vol. 16, no. 3, pp. 789-802, Mar. 2007.
  - [12] Averbuch, A. Schclar, and D. Donoho, "Deblocking of block-transform compressed images using weighted sums of symmetrically aligned pixels," IEEE Trans. Image Process., vol. 14, no. 2, pp. 200-212, Feb. 2005.
  - [13] C.A.Bouman, I. Pollak, "A Document Image Model and Estimation Algorithm for Optimized JPEG ecompression," vol. 18, pages 2518-2535, Nov. 2009.
  - [14] Xiwen Owen Zhao, Zhihai Henry He, "Lossless Image Compression Using Super-Spatial Structure Prediction," IEEE Signal Processing Letters, vol. 17, no. 4, April 2010.
  - [15] Jaemoon Kim, Jungsoo Kim and Chong-Min Kyung, "A Lossless Embedded Compression Algorithm for High Definition Video Coding," 978-1-4244-4291 / 09 2009 IEEE. ICME 2009.
  - [16] Yu Liu, Student Member, and King Ngi Ngan, "Weighted Adaptive Lifting-Based Wavelet Transform for Image Coding," IEEE Transactions on Image Processing, vol. 17, NO. 4, Apr 2008.
  - [17] R. de Queiroz and R. Eschbach, "Fast segmentation of JPEG compressed documents," *J. Electron. Imag.*, vol. 7, Apr. 1998, pp. 367-377.
  - [18] H. T. Fung, K. J. Parker, "Color, complex document segmentation and compression," *Proc. of the SPIE*, vol. 3027, San Jose, CA, 1997.
  - [19] Zhe-Ming Lu, Hui Pei, "Hybrid Image Compression Scheme Based on PVQ and DCTVQ," *IEICE - Transactions on Information and Systems* archive, Vol. E88-D, Issue 10, October 2006.



**MS. I. Suneetha** received the B.Tech and M.Tech Degrees in E.C.E from Sri Venkateswara University College of Engineering (SVUCE), Tirupati, India in 2000 and 2003 respectively.

She is pursuing her Ph.D Degree at SVUCE, Tirupati and working as Associate Professor & Head, E.C.E department, Annamacharya Institute of Technology and Sciences (AITS), Tirupati. Her teaching and research area of interest includes 1D & 2D signal processing.

## BIOGRAPHIES



**Mr. P. Mahesh** received the B.Tech Degree in E.I.C.E from N.B.K.R.IST, Vidyanagar, Nellore under Affiliated Sri Venkateswara University (SVU). He is pursuing M.tech Degree with the Specialization of ECE (DSCE) in Annamacharya Institute of Technology and Sciences (AITS), Tirupati, AP, India.