Journal of Engineering Sciences, Assiut University, Vol. 38, No. 2, pp. 533-546, March, 2010.

COMPOUND IMAGE SEGMENTATION

Ibrahim Ismail[1], Alaa Hamdy[2] and Seham Mostafa[3]

1- Professor Doctor Engineer, 2 - Doctor Engineer, 3 - Engineer.

1, 2 are with the faculty of engineering, Helwan University, Cairo, Egypt.

(Received October 13, 2009 Accepted January 9, 2010).

Compound document images contain a mixture of natural image and text/graphics. They are very common forms of documents found in magazines, websites, etc. Text and graphics components need special care in the use of compression because text and graphics cannot withstand the significant distortion that is acceptable for natural images. This paper represents a study of different algorithms for segmentation to identify the compound image components. This study focuses on discrete cosine transform (DCT) algorithm, fast Fourier transform (FFT) algorithm, and block based segmentation algorithm; mean, variance, and mean/variance. The segmentation process in general starts with dividing the whole image into non overlapping blocks. Then each block has to be classified either to text/graphics or image via a certain threshold according to the algorithm. the results show that after applying any algorithm the compound image is already classified and the percentage of correct classified text pixels CCPt is calculated and also the percentage of correct classified image pixels CCPi is calculated to present the efficiency of the algorithm, that FFT algorithm is the best one of them that the CCPi between 87.59% in CNN image and 100% in letter image also CCPi between 97.33% TOY image and 100% in leteer image.

KEYWORDS: *image segmentation, block based segmentation, discrete cosine transform, fast Fourier transform, mean based algorithm, variance algorithm.*

1. INTRODUCTION

Documents are now present in a wide spectrum of printing system. From offset printers to home desktop computers, documents in digital form are common place. Compound documents are assumed here as images which contain a mixture of textual, graphical, or pictorial contents [1]. Image segmentation algorithms partition the set of pixels of an image into a specific number of different, spatially homogeneous groups [2].

In general, there are three approaches for image segmentation: object based segmentation, layer based segmentation, and block based segmentation [3] [4]. This paper focuses on block based segmentation which is very simple and faster than layer based segmentation. The general function for block based segmentation is block based algorithm which will be discussed in the second section of this paper. DCT based algorithm will be discussed in the third section and Wavelet based algorithm will be explained in the fourth section. FFT based algorithm will be shown in the fifth section

The experimental results and conclusion will be presented in section six and seven respectively.

2. BLOCK BASED ALGORITHM

The following algorithms only rely on the n x n image blocks, without using any transform [5]. Each of our algorithms computes a certain 'activity' score for blocks in an image. Then a threshold is assigned to the score, under which the block is considered a natural image, and over which is considered text [6].

2.1 MEAN M

The mean of pixel intensities (μ) in an image block is a useful feature for segmentation. The compound images can be separated based on μ into two classes; natural image, and text [6].

2.1.1 BASIC GLOBAL THRESHOLD

Segmentation is accomplished by scanning the image pixels block by block and labeling each block as natural image or text/graphics depending on whether the gray level of those pixels is greater than the value of threshold T.
The following algorithm can be used to obtain T automatically:

1- Select an initial estimate value for T.
2- Segment an image using T. This will produce two groups of pixels G_1 and G_2. G_1 consists of all pixels with gray level values that are greater than T and G_2 consists of all pixels with gray level values that are less than or equal T.
3- Compute the mean of gray level values μ_1 and μ_2 for the pixels in the regions G_1 and G_2.
4- Compute the threshold value $T = (\mu_1 + \mu_2) / 2$.
5- Repeat steps from 2 to 4 until the difference in T in recursive iterations is smaller than a predefined T_0 [7].

The parameter T_0 is used to stop the algorithm after changes become small in terms of this parameter.

2.2 ARIANCE

This algorithm is based on pixel variance within a block. Text/graphics blocks are likely to have a higher variance than non-text block [8].

3. DISCRETE COSINE TRANSFORM (DCT) BASED ALGORITHM

Energy is distributed differently among DCT coefficients for text and non-text blocks. Thus segmentation is achieved by examining the appropriate set of DCT coefficients that capture this difference between text and non-text, and then compare the energy or absolute sum of these coefficients (the activity) to the threshold. Similarly, the computed absolute sum also shows an important difference between text and non-text. So, it is not necessary to use energy (absolute sum is faster to compute) [9], i.e. the sum of 64 coefficients is computed to get the activity of the block.

4. WAVELET BASED ALGORITHM

The wavelet transform is an effective means of mapping an image from the space domain to the frequency domain in terms of wavelets so it becomes wavelet spectrum instead of frequency spectrum. In a typical wavelet coding scheme, an image's frequency components are subdivided recursively, refining the lower sub-band (in the two band decomposition case) at each step [3], [4], [7]. When applying wavelet transform on an image, the four coefficients CA approximation matrix and details matrices CH, CV, CD must be calculated, where CH is the horizontal coefficient matrix, CV is the vertical coefficient matrix, and CD is the diagonal coefficient matrix [10].the effective coefficient matrix in segmentation process is the value of CA .

5. FAST FOURIER TRANSFORM (FFT) BASED ALGORITHM

The fast Fourier transform of an image f(x,y) of size M x N is given by the variables u, v which are the transform or frequency variables [11], [12]. According to the frequency components, the compound document will be classified to its two items by applying a suitable value for threshold that will be calculated automatically as shown before.

6. EXPERIMENTAL RESULTS

This section represents a comparison between different compound image segmentation algorithms, discrete cosine transform (DCT) algorithm, wavelet transform, fast Fourier transform, and block based segmentation algorithm, i.e. mean, variance, and mean/variance. The entire compound images are applied on MATLAB 6 software and the operating system is windows XP on a personal computer with processor 1.3 GHZ and 128 RAM. The results are given for three tested images; each of them is gray scale, and the extension of all tested images is JPEG with size 256x256. All the algorithms are applied on compound images after dividing them into 16x16 non overlapping blocks.

Several experiments are carried out in order to verify the possibility of achieving good results with the proposed algorithm, in case of different images.

FOR CNN IMAGE:

This is a test image in gray scale level the total number of text pixels in this test image are 47208 pixels also the number of image pixels in the same test image are 16552. Table 1 shows the results after applying each algorithm on this test image where CCPi means the correct classified image pixels, CCPt is the correct classified text pixels and the time measure in seconds.

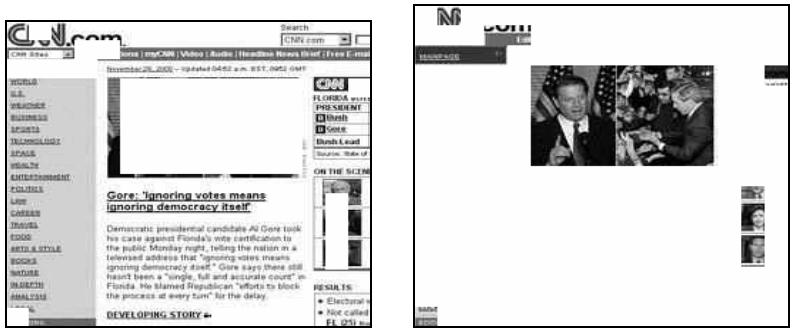
Algorithm	Threshold	CCPi	CCPi %	CCPt	CCPt%	Time sec
MEAN	161.66	14498	87.59	47208	99.06	14.46
VARIANCE	65	14498	87.59	46543	98.59	7.79
MEAN& VARIANCE	173 180	14498	87.59	46543	98.59	10.45
DCT	11.10	16280	98.356	38442	81.43	4.56
FFT	153	14498	87.59	46510	98.52	4.02
WAVELET	46	16280	98.36	38442	81.43	7.37
Algorithm	Threshold	CCPi	CCPi %	CCPt	CCPt%	Time sec
MEAN	161.66	14498	87.59	47208	99.06	14.46
VARIANCE	65	14498	87.59	46543	98.59	7.79
MEAN& VARIANCE	173 180	14498	87.59	46543	98.59	10.45
DCT	11.10	16280	98.356	38442	81.43	4.56
FFT	153	14498	87.59	46510	98.52	4.02
WAVELET	46	16280	98.36	38442	81.43	7.37
Algorithm	Threshold	CCPi	CCPi %	CCPt	CCPt%	Time sec
MEAN	161.66	14498	87.59	47208	99.06	14.46
VARIANCE	65	14498	87.59	46543	98.59	7.79
MEAN& VARIANCE	173 180	14498	87.59	46543	98.59	10.45
DCT	11.10	16280	98.356	38442	81.43	4.56
FFT	153	14498	87.59	46510	98.52	4.02
WAVELET	46	16280	98.36	38442	81.43	7.37

Table 1: The results of segmenting CNN image

And now you can see yourself the image after segmentation:

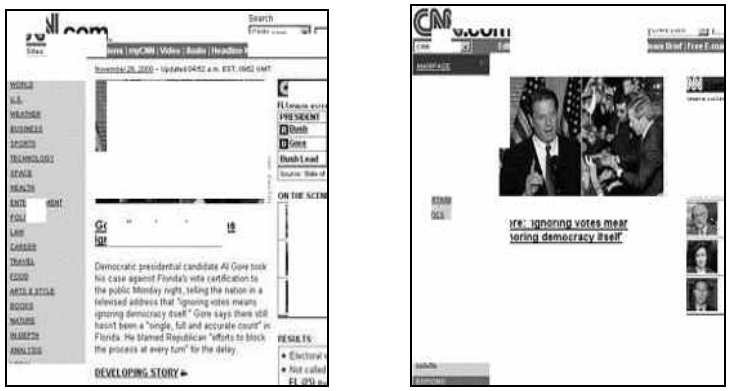
(a) (b) (c)

Fig. 1: (a) Original image CNN, (b) pixels classified as text, and (c) pixels classified as image after applying mean algorithm.



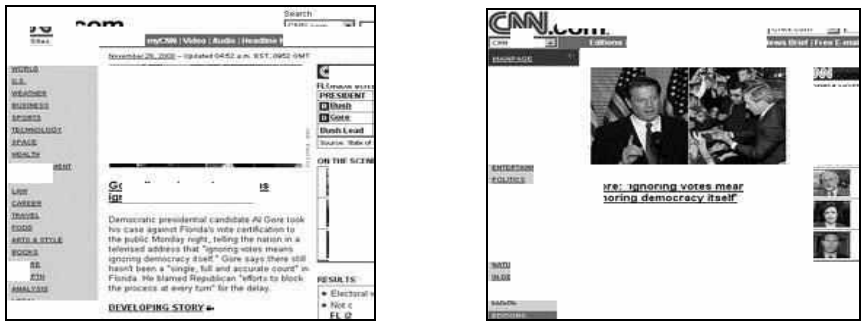
(a) (b)

Fig. 2: (a) Pixels classified as image and (b) pixels classified as text after applying variance algorithm.



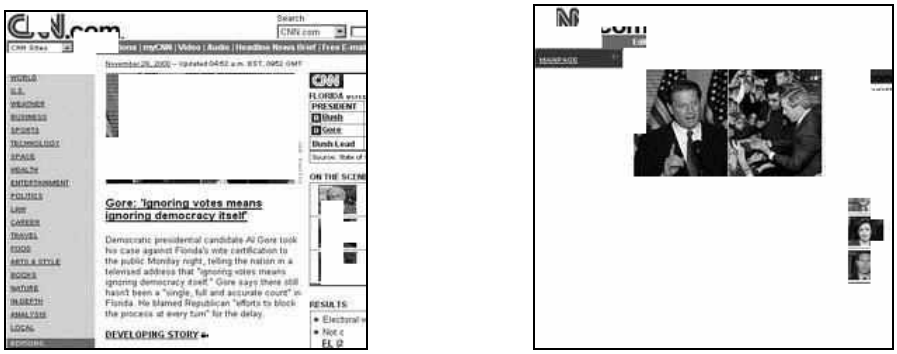
(a) (b)

Fig. 3: (a) Pixels classified as text and (b) pixels classified as image after applying mean / variance algorithm.



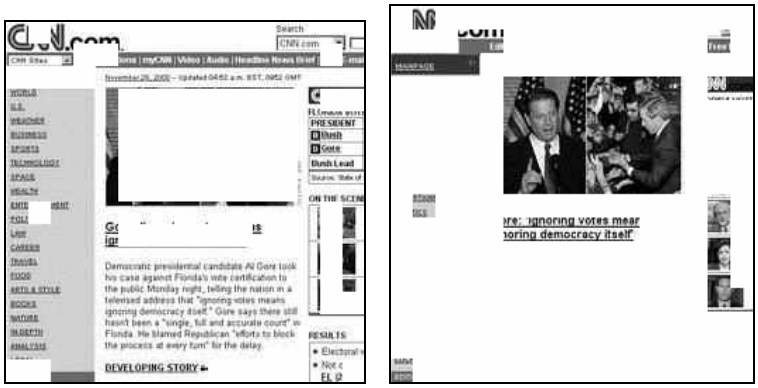
(a) (b)

Fig. 4: (a) Pixels classified as text and (b) pixels classified as image after applying DCT algorithm.



(a) (b)

Fig. 5: (a) Pixels classified as text and (b) pixels classified as image after applying FFT algorithm.



(a) (b)

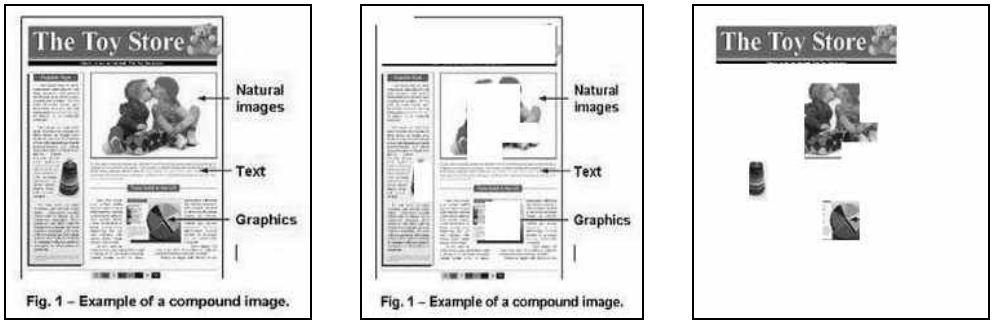
Fig. 6: (a) Pixels classified as text and (b) pixels classified as image after applying wavelet transform.

FOR TOY IMAGE:

This is a test image in gray scale level the total number of text pixels in this test image are 20975 pixels also the number of image pixels in the same test image are 16528. Table2 shows the results after applying each algorithm on this test image where CCPi means the correct classified image pixels, CCPt is the correct classified text pixels and the time measure in seconds.

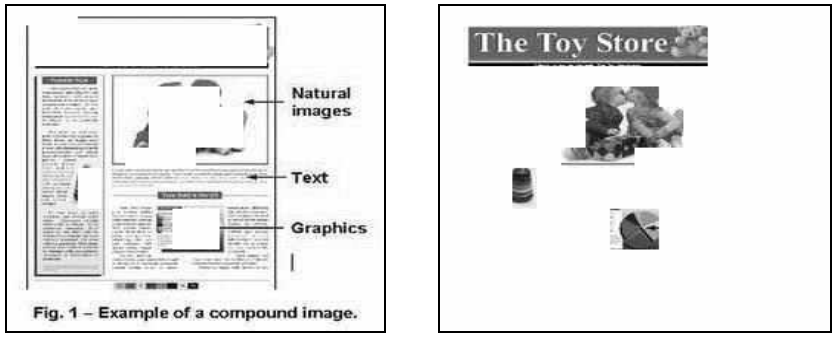
Table 2: The results of segmenting TOY image

Algorithm	Threshold	CCPi	CCPi %	CCPt	CCPt%	Time sec
MEAN	172.42	13318	78.82	20929	99.78	46.27
VARIANCE	70	13318	78.82	20929	99.78	10.20
MEAN& VARIANCE	182 100	1318	97.83	20417	97.33	6.93
DCT	11.2	16303	96.75	1994	95.08	3.20
FFT	208	16528	97.83	20417	97.33	3.54
WAVELET	50	16061	94.80	20417	97.33	7.31



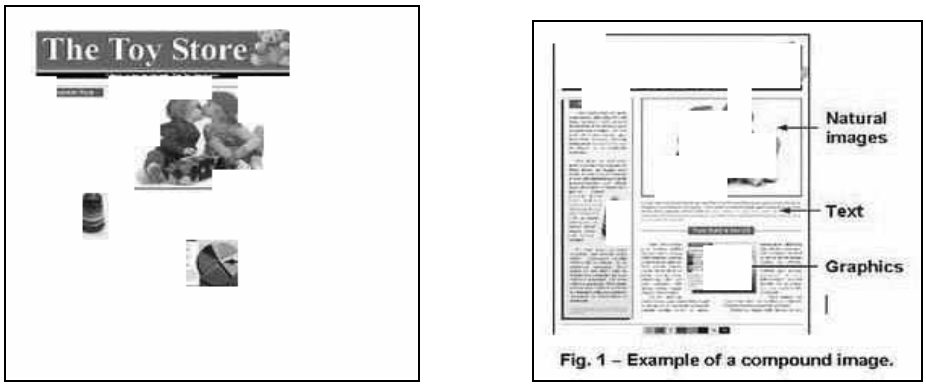
(a) (b) (c)

Fig. 7: (a) Original image TOY, (b) pixels classified as text, and (c) pixels classified as image after applying mean algorithm.



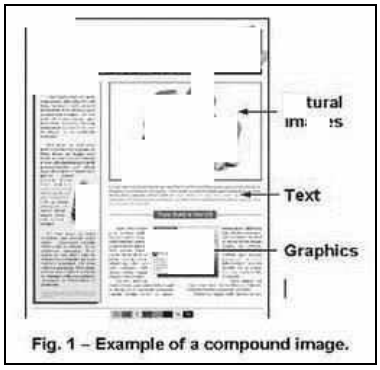
(a) (b)

Fig. 8: (a) Pixels classified as text and (b) pixels classified as image after applying variance transform.



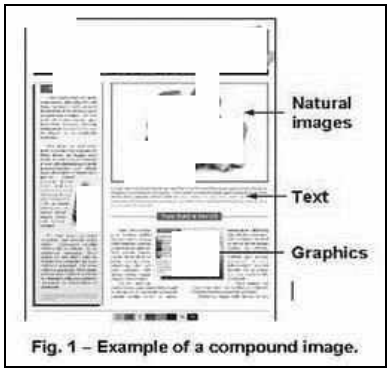
(a) (b)

Fig. 9: (a) Pixels classified as image and (b) pixels classified as text after applying mean / variance transform.



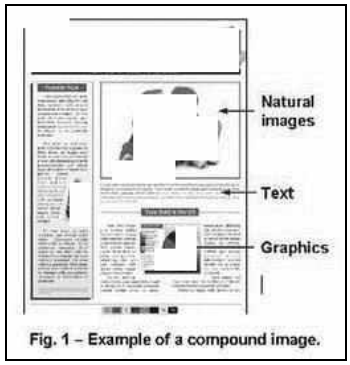
(a) (b)

Fig. 10: (a) Pixels classified as text and (b) pixels classified as image after applying DCT algorithm.



(a) (b)

Fig. 11: (a) Pixels classified as text and (b) pixels classified as image after applying FFT algorithm.



(a) (b)

Fig. 12: (a) Pixels classified as text and (b) pixels classified as image after applying wavelet transform algorithm.

FOR LETTER IMAGE:

This is a test image in gray scale level the total number of text pixels in this test image are 26458 pixels also the number of image pixels in the same test image are 16512. Table3 shows the results after applying each algorithm on this test image where CCPi means the correct classified image pixels, CCPt is the correct classified text pixels and the time measure in seconds. But here all algorithms performe the segmentation processs the comparison only between the threshold and the segmentation time.

Table 3 The results of segmenting LETTER image

Algorithm	MEAN	VARIANCE	MEAN& VARIANCE	DCT	FFT	WAVELET
Threshold	189.36	110	198/100	12	165	73
Time sec	41.26	7.79	10.70	3.46	2.95	7.46
Algorithm	MEAN	VARIANCE	MEAN& VARIANCE	DCT	FFT	WAVELET
Threshold	189.36	110	198/100	12	165	73
Time sec	41.26	7.79	10.70	3.46	2.95	7.46

(a) (b) (c)

Fig. 13: (a) Original image LETTER, (b) pixels classified as text, and (c) pixels classified as image after applying any algorithm.

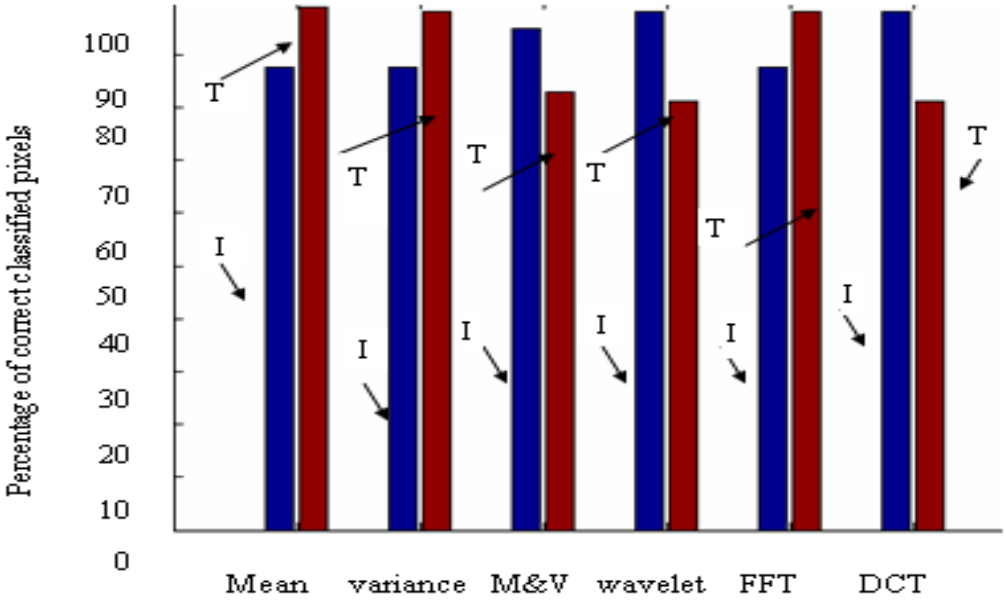


Fig. 14: Comparing the efficiency of different algorithms applied on CNN image where I is the correct classified image pixels and T is the correct classified text pixels.

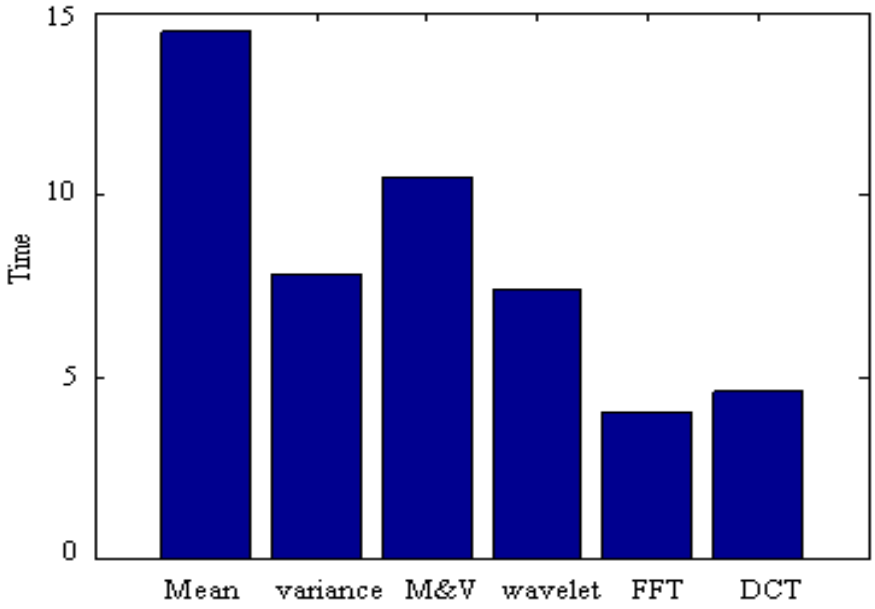


Fig. 15: The required time in seconds to segment compound image CNN by various techniques.

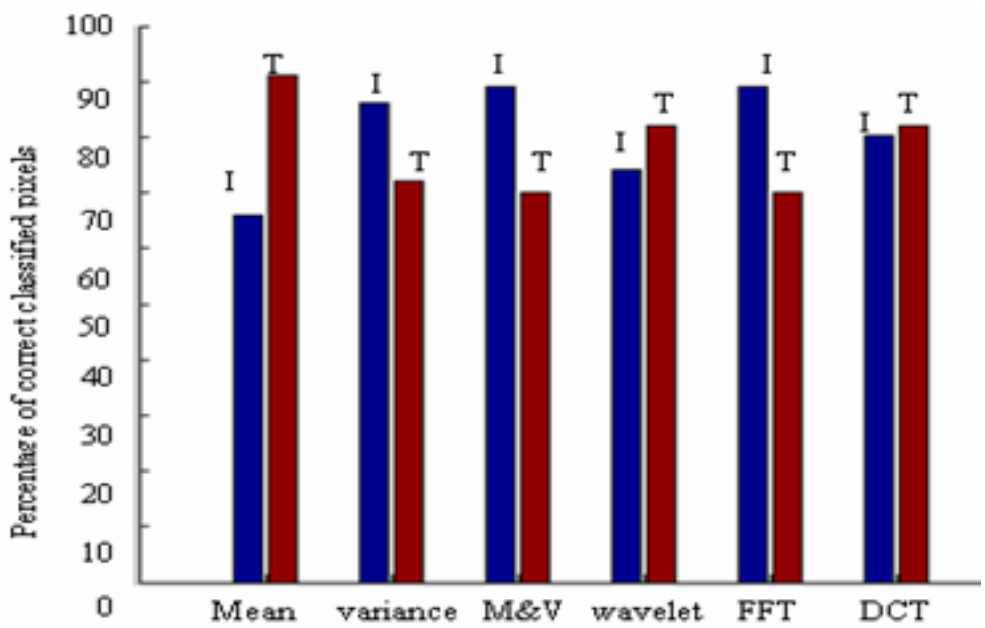


Fig. 16: Comparing the efficiency of different algorithms applied on TOY image where I is the correct classified image pixels and T is the correct classified text pixels.

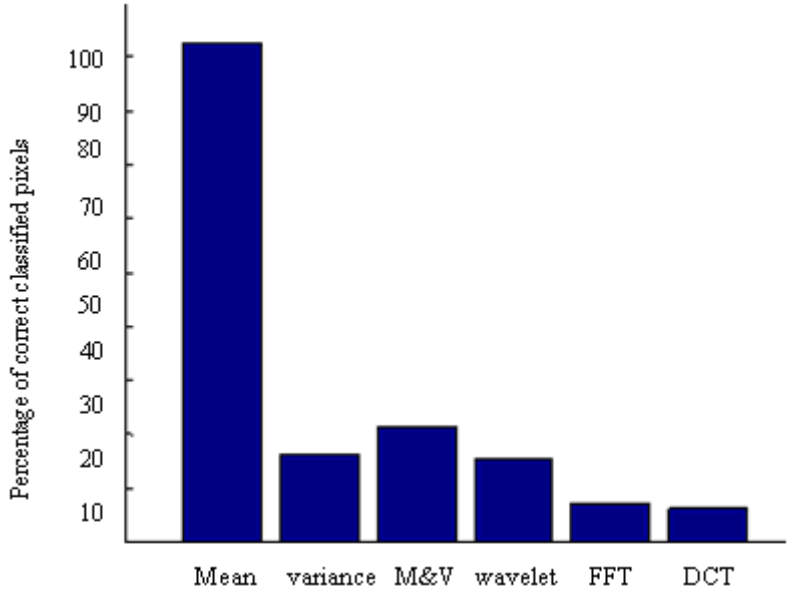


Fig. 17: The required time in seconds to segment compound image TOY by various techniques.

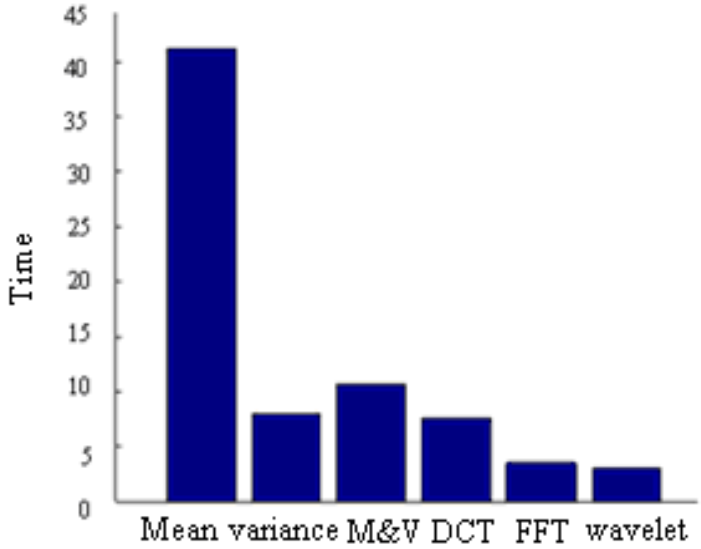


Fig. 18: The required time in seconds to segment compound image LETTER by various techniques.

7. CONCLUSION

This paper presents compound image segmentation using Mean, Variance, Mean / Variance, DCT, FFT, and Wavelet transform. The proposed algorithm is obtained according to some factors. The first one what is the interested pixels due to the applications. The second factor is the time and also the nature of the compound image. So, if the application interested in text pixels or if the majority pixels in the compound image are text pixels, then the recommended algorithm is the mean. However, it takes very long time. If the time is a necessary factor, one can use FFT algorithm, although the percentage of correct classified pixels is not very high as in the case of using the mean.

8. FUTURE WORK

This paper presents six different algorithms for compound image segmentation. The results show that there are some misclassified pixels, like the text written in bold font which is always classified as image. Also, the text written in normal font, but on a dark background like the toolbar, it classified as image as in the CNN compound image. In future work, it is possible to combine these methods to overcome these problems.

REFERENCES

[1] R. Buckley, D. Venable, and L. McIntyre, “New Developments in Color Facsimile and Internet Fax”, in Proc. IS&T 5th Color Imaging Conference, Scottsdale, AZ, pp. 296–300, Nov. 1997.

[2] P. Orbanz, J.oachim, M. Buhmann "Nonparametric Bayesian Image Segmentation" Int J Comput Vis (2008) 77: 25–45 DOI 10.1007/s11263-007-0061-0.
[3] T. Pavlidis and J. Zhou, "Page Segmentation and Classification", CVGIP: Graph. Models Image Proc., Vol. 54, pp. 484-496, Nov. 1992.
[4] International Telecommunication Union (ITU-T), "Mixed Raster Content (MRC)", Draft recommendation T.44.
[5] Kush R. Varshney, "Block-Segmentation and Classification of Grayscale Postal Images", School of Electrical and Computer Engineering Cornell University, Ithaca, NY 1485, May 2003.
[6] A. Said, "Compression of Compound Images and Video for Enabling Rich Media in Embedded Systems", SPIE Visual Commun. Image Processing Conf., SPIE Proc. Vol. 5308, pp. 69-82, Jan. 2004.
[7] J. Li and R. M. Gray, "Context-Based Multiscale Classification of Images", in Proceedings of the IEEE International Conference on Image Processing, 1998.
[8] R.L. de Queiroz, "Compression of Compound Documents", Proc. IEEE Int. Conf. Image Processing, Vol. 1, pp. 24–28, Oct. 1999.
[9] William B. Pennebaker, and Joan L.Mitchell. "JPEG: Still Image Data Compression Standard", Van Nostrand Reinhold, 1993.
[10] Mayer, Y Ondelettes et opėrateurs, Tome 1, Herrmann Ed. (English translation: wavelets and operators, ambridge Univ. Press. 1993).
[11] FFTW (http:/www.fftw.org).
[12] M. Kalman, I. Keslassy, D. Wang, and B. Girod, "Classifications of compound image based on transform coefficient likelihood," Information Systems Laboratory, Department of Electrical Engineering Stanford University, Stanford, CA94305 July 2002.

تقسيم الصورة المركبة

الصورة المركبة هى تلك اللتى تتكون من جزئين أساسيين الأول هو الذى يحتوى على كتابه نصية اما الجزء الثانى فهو يحتوى على الصورمثل الصور الفوتوغرافية والهدف من هذا البحث هو محاولة فصل جزئي الصورة المركبة عن بعضهما و تعتبر هذه العملية هى عملية تجهيز واعداد للصوره لإجراء أى عملية لكل جزء على حده.

تتم عملية الفصل عن طريق تقسيم الصوره المحولة إلي التدريج الرمادى إلي مربعات غير متداخلة بحجم 16x16 بداية من الجانب العلوى الأيسر للصورة وتطبيق بعض النظريات علي كل مربع وبواسطة حد فاصل معين متغير بتغير الصورة وكذلك نستطيع فصل أجزاء الصورة، وفي النهاية نقوم بحساب النسبة المئوية للنقاط الممثلة للكتابه النصية التي صُنفت بنجاح وكذلك بحساب النسبة المئوية للنقاط الممثلة للصوروايضا حساب الوقت اللازم لإتمام عملية فصل جزئي الصورة المركبة والأن سيتم عرض النظريات المختلفة التى استخدمت.

أولاً: القيمة الوسطي للنقاط الممثلة للصورة: نقوم بحساب القيمة الوسطي للنقاط الممثلة للصورة لكل مربع وعن طريق الحد الفاصل الشامل نستطيع فصل المربعات التى تحتوى على الصور وكذلك المربعات التى تحتوى على كتابه نصية.

ثانيا: أستخدام الخلاف الحسابى: نطبق معادلة الخلاف الحسابى علي كل مربع و تعين قيمة الحد الفاصل النى يمكن من خلالها فصل المربعات التى تحتوى على الصور عن المربعات التى تحتوى على كتابه نصية حيث أن قيمة الخلاف الحسابى للمربعات التى تحتوى على الصور أقل من قيمة الخلاف الحسابى للمربعات التى تحتوى على كتابه نصية.

ثالثا: أستخدام القيمة الوسطي للنقاط الممثلة للصورة و الخلاف الحسابى معا: حيث نقوم بتطبيق النظامين السابقين معا على كل مربع من مربعات الصورة وتحديد قيمتين للحد الفاصل يمكن من خلالهما فصل جزئي الصورة المركبة عن بعضهما.

رابعا: تحويل جيب التمام المنفصل (DCT) : عن طريق تطبيق هذا التحويل على كل مربع من مربعات الصورة المركبة تتحول قيم النقاط الممثلة للصورة إلي عوامل لها قيم أخرى و تعتبر قيمة أول عامل فى كل مربع من مربعات الصورة هو المحدد لنوع هذا المربع إذا كان يحتوى على صور أو على كتابه نصية لذلك يتم استخدام هذه القيمة لتصبح هي الحد الفاصل.

خامسا: تحويل فورير السريع : بتطبيق هذا التحويل على كل المربعات المكونة للصورة تتحول قيم النقاط الممثلة للصورة إلي عوامل لها قيم أخرى وتؤخذ قيمة أول عامل من كل مربع كحد فاصل يمكن من خلاله تصنيف هذا المربع لأحد أجزاء الصورة.

سادسا: التحويل المويجى : هذا التحويل يحول النقاط الممثلة للصورة من مجال الفضاء إلي مجال التردد حيث يحسب لكل نقطه من النقاط الممثلة للصورة أربعة عوامل هم العامل التقديرى (CA) والعامل الأفقى (CH) والعامل الرأسي (CV) و أخيرا العامل القطرى (CD) وفقط يم أخذ قيمة العامل التقديرى (CV) كقيمة للحد الفاصل الذي يقوم بفصل أجزاء الصورة.

من كل الأنظمة السابقة فقط نستخدم نظام واحد لفصل جزئي الصورة المركبة ويتوقف هذا النظام علي التطبيق والوقت اللازم لإتمام عملية فصل جزئي الصورة المركبة حيث هناك تطبيقات تهتم بكفاءة الصورة فعليها استخدام النظم التي تحقق أعلى نسب فصل للنقاط الممثلة للصورة ولكن هذا يتطلب وقت طويل ويصبح علي المستخدم المفاضلة بين عاملي الكفاءة والزمن. وكذلك أيضا نوع الغالبية العظمى للنقاط الممثلة للصورة المركبة، علي سبيل المثال إذا كانت الصورة المركبة تحتوى على كتابه نصية أكثر من الصور يقوم المستخدم بتطبيق النظم التي تحقق أعلى نسب فصل للنقاط الممثلة للكتابه نصية .