Text Segmentation in Mixed-Mode Images

Navin Chaddha, Rosen Sharma, Avneesh Agrawal and Anoop Gupta
Computer Systems Laboratory
Stanford University, Stanford, CA 94305.

Abstract

Block based algorithms have found widespread use in image and video compression. However, popular algorithms such as JPEG, which are very effective in compressing continuous tone images, do not perform well with mixed-mode images which have a substantial text component. With a growing number of applications where such images occur, e.g., color facsimile, digital libraries and educational videos, there are advantages in being able to classify each block as being text or continuous tone. With such a classification, different compression parameters or even algorithms may be employed for the two kinds of data to obtain high compression with minimal loss in visual quality. In this paper we analyze and compare four methods for block classification in mixed mode images, namely variance, absolute-deviation, edge, and DCT based methods. Our evaluation of each scheme is based on the accuracy of segmentation, robustness across different types of images and sensitivity to the threshold used for segmentation. Our results show that DCT based segmentation offers the best accuracy and robustness. Another advantage of DCT is that it is compatible with standards like JPEG, MPEG and H.261.

1. Introduction

With the increasing popularity of multimedia applications, there is a growing need for a robust and efficient text segmentation algorithm. Applications like educational videos, color fax and scanned documents in digital libraries, are rich in both continuous tone and textual data. For efficient storage or transmission over a band-limited channel, a viable compression algorithm should attempt to maximize compression with minimal loss in visual quality. Since text and image data have different properties, proper segmentation of the data would help in the process of compression.

In this paper we compare several algorithms for block-based text segmentation. Our evaluation is based on two criteria: the accuracy of segmentation and the robustness across different types of images. To evaluate accuracy, we compare the different segmentation algorithms based on the number of text and non-text blocks wrongly classified. To evaluate robustness, we present the results for two sets of experiments: one where the parameters of the algorithm are adjusted to give the best possible segmentation for a particular image and the other where the parameters are kept constant across all images. The latter is important, because in many real situations the application may not allow the parameters to be adjusted (due to time constraints, computational constraints, or lack of guiding information).

There has been a lot of work in segmentation [1], [2], [3], [4]. This paper differs from the others in its unique perspective of looking for a segmentation algorithm which offers high robustness and accuracy across different images and it also presents a new DCT based segmentation algorithm.

This paper is organized as follows. Section 2 describes the different algorithms we evaluate for segmentation. Section 3 describes the test image database. Section 4 gives a comparison of the different algorithms and we conclude in Section 5.

2. Algorithms for Text Segmentation

In this section we briefly describe the four different segmentation algorithms we analyze in this paper.

2.1 Variance Based Approach

One popular approach for segmentation [1] uses the block *variance* for classifying blocks as text or non-text. This algorithm is based on the observation that text blocks are likely to have a higher variance than non- text blocks. A key parameter for this algorithm is the threshold used for segmentation. We will study how good the algorithm works and how robust the threshold is across images.

2.2 Absolute Deviation Based Approach

A variant of the above is the *absolute deviation (AD)* [1] approach. In this approach, the mean absolute deviation of each block is used for classifying the block as text or non-text. The key parameter again is the threshold used for segmentation. The main advantage of this approach over the variance based approach is its reduced computational complexity.

2.3 Edge Based Segmentation

The edge-based approach relies on the intuition that text blocks are likely to have more edges than non-text blocks. The segmentation can then be done on the basis whether or not the fraction of pixels belonging to edges is above or below a certain threshold. In this study we use an eight-neighbor Sobel filter as it has the advantage of smoothing and reducing the sensitivity of the derivative operations to noise. The Sobel filter uses two masks, one for horizontal edges and the other for vertical edges.

For each pixel we compute the sum of squares of the gradients in X and Y direction and compare this value to a threshold to classify that pixel as edge pixel or non-edge pixel. The *edge-threshold* is determined for each image individually. For block based segmentation if the number of edge pixels is more than a *number-threshold* then the block is classified as a text block otherwise it is classified as a non-text block.

2.4 DCT Based Approach

DCT based algorithms have found use in compression algorithms based on transform coding. The energy distribution of text blocks amongst some DCT coefficients is significantly higher compared to the non-text blocks. Thus segmentation can be obtained by choosing an appropriate set of DCT coefficients which captures the difference between the text and the non-text blocks and comparing their absolute sum to a threshold. The key parameters are which DCT coefficients to choose and the value for the threshold. Although DCT is much more expensive than the absolute deviation approach, we note that it is the most popular transform for block based compression algorithms and has been adopted as the algorithm of choice in JPEG, MPEG and H.261 standards. For such algorithms, the incremental computational cost of summing some of the DCT coefficients is negligible.

1058-6393/95 $4.00 © 1995 IEEE

3. Evaluation Methodology

This section is divided into two parts. The first section describes the test image database on which the segmentation algorithms are tested. The second section discusses the evaluation criterion for comparing different segmentation algorithms.

3.1 Test Image Database

An important feature of any segmentation or compression study is a test image data base on which different algorithms can be tested and compared. There is no mixed mode image data base available which spans the different types of images which are found in educational videos, color-fax documents and images scanned in digital libraries. We have developed our own database containing several images corresponding to the above applications.

Table 1 gives a description of the five mixed mode images we use in this study. Images T1, T2, T3, and T4 are typical of educational videos and have been obtained from the Stanford Instructional Television Network (SITN). These images represent the most frequently used instructional media over SITN. Image T5 is characteristic of scanned documents.

3.2 Evaluation Criteria

As stated earlier, we are interested in two main criteria: accuracy and robustness. We will measure accuracy in terms of *false-negatives*, i.e., the number of actual text blocks segmented as non-text blocks, and*false-positives*, i.e.,the number of non-text blocks which are segmented as text. Our objective is to minimize total error, i.e., the sum of false negatives and false positives, subject to the constraint that false negatives are less than false positives. We give greater emphasis to false negatives, because that way we ensure that text is compressed using the correct compression parameters / algorithm. It has been our experience that ensuring good text quality is of key importance in mixed-mode images. For our analysis, the correct text blocks in an image were selected manually. For robustness, our main criterion is the sensitivity of the segmentation algorithm to the threshold used. For DCT based segmentation, choice of which coefficients to choose is also critical.

4. Results

In this section we evaluate the various segmentation schemes by first discussing their characteristics individually and then comparing them. For space reasons, while summary results are presented for all images, detailed results are presented only for image T2.

4.1 Variance Based Approach

Figure 2.a presents the probability distribution of standard deviation for blocks in image T2. As can be clearly seen from the figures, the standard deviation for text blocks is generally higher than that of non text blocks. However, there is some overlap of the two distributions, marking regions of indecision in which the block cannot be conclusively classified as text or non-text.

Figure 2.b plots the number of false-positive blocks and total erroroneously classified blocks as a function of segmentation threshold for image T2. It can be seen that the total error (solid line) is minimized for threshold=10. Figure 7.a shows the results of segmentation with this approach for a threshold of 10. It can be seen that while the number of false-negatives is very small (5), the number of false-positives is enormous (278); the approach unfortunately also classifies the border of the blackboard as text blocks.

Similar experiments as above were also performed to obtain the best threshold for other images.Table 2 shows the results for rest of the images. Overall, we also see that the best threshold varies quite a bit for the different images; the range is between 10 and 25.

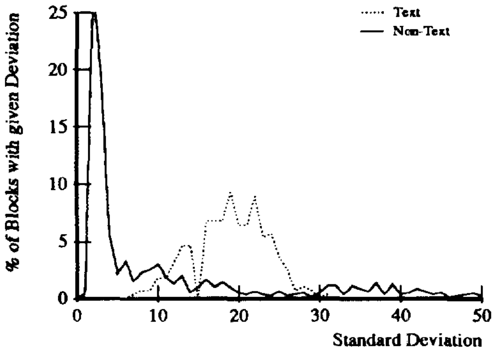


Figure 2.a Pdf of standard deviation of blocks in image T2.

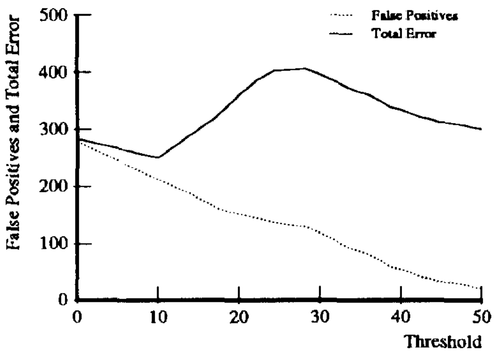


Figure 2.b Number of false-positive and total error blocks in image T2 as a function of threshold.

4.2 Absolute Deviation Based Approach

Figure 3.a shows the probability distribution of absolute deviation for blocks in image T2. As can be clearly seen, the distribution of block absolute deviation for text and non-text blocks has overlap similar to that found for block variance.

Figure 3.b plots the number of false-positive blocks and total erroroneously classified blocks as a function of segmentation threshold for image T2. It can be seen be seen from this figure that the total error is minimized for threshold=9. Figure 7.c shows the results of absolute deviation based segmentation with a threshold of 9 for image T2. It can be seen that while the false-negatives are small (34), the false-positives are very high (246); this approach again captures the edges of the blackboard in the background as text blocks.

Similar experiments were performed to obtain the best thresholds for absolute deviation approach on different images. Table 2 shows the results for the rest of the images. The data show that the results for the absolute deviation based algorithm are highly dependent on the best threshold used for segmentation; we see the best threshold varies between 7 and 19 for the different images.

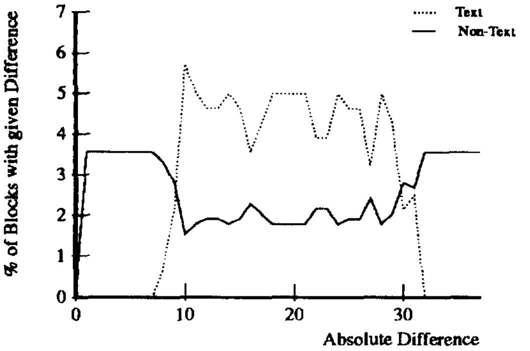


Figure 3.a Pdf of block absolute deviation for image T2.

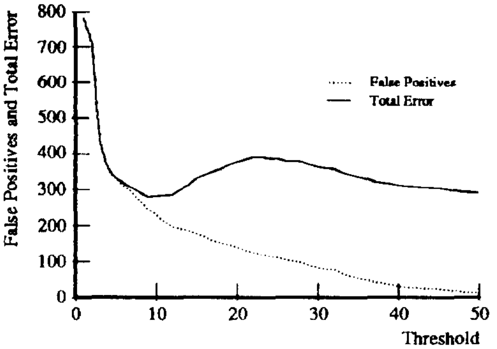


Figure 3.b Number of false-positive and total error blocks in image T2 as a function of threshold.

4.3 Edge Based Segmentation

Figure 4.a shows the results of experimental determination of the two thresholds (*edge-threshold* and *number-threshold*) on Image T2. The edge-threshold is used for deciding if a pixel is an edge-pixel or not. The number-threshold is used for deciding if the block is a text-block or not. It is found that the total error for image T2 is minimized for edge-threshold=17 and number-threshold=3. Figure 4.b shows the two dimensional search for thresholds on all images to obtain an average threshold. The total error across all images in minimized for edge-threshold=15 and number-threshold=9. This thresholds will be used for comparative evaluation of robustness later.

Figure 7.e shows the results of edge based block segmentation with an edge-threshold of 17 and number-threshold of 3 on image T2. It can be seen from Figure 7.e that the false-negatives for the segmentation are small (9) but the false-positives are very high (222), this approach also captures the background as text blocks.

Table 2 shows the results for rest of the images. The first threshold in the table corresponds to the number-threshold while the second-threshold corresponds to the edge-threshold. We also see that the best threshold varies quite a bit for the different images; the range is between 3 and 29 for number-threshold and 7 and 29 for edge-threshold. Overall, we find the three approaches we have considered so far, variance, absolute deviation and edge based, of limited use because of the variation of the best threshold across images. We now look at the DCT based approach.

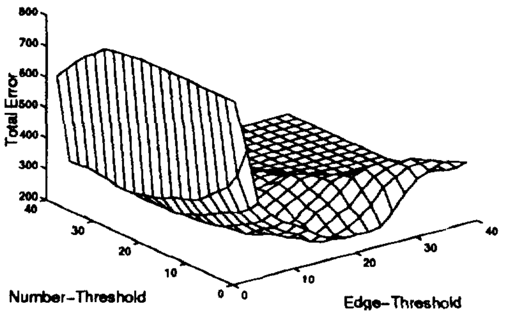


Figure 4.a 2-dimensional search for thresholds in edge based scheme for image T2.

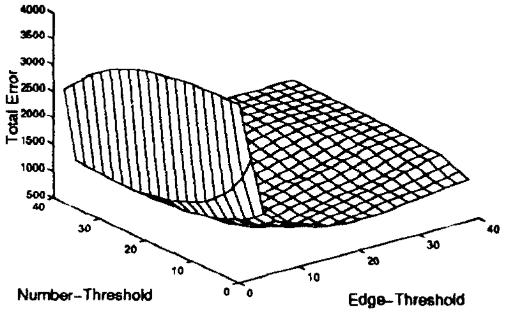


Figure 4.b 2-dimensional search for thresholds across all images.

4.4 DCT Based Approach

In the DCT based approach we are relying on the fact that the energy distribution of text blocks amongst some DCT coefficients is significantly higher compared to the non-text blocks. To show this is indeed the case, Figure 5.a shows the average energy for each DCT coefficient for text and image blocks averaged over Image T2. Figure 5.b shows the absolute sum for each DCT coefficient for text and image blocks averaged over Image T2.

It can be clearly seen that the energies and the absolute sum for some of the textual blocks are significantly higher than the corresponding energies for the image blocks. Thus, a simple and effective method of classifying textual data is to find the sum of the energies or absolute sum of an appropriately selected subset of the DCT coefficients which capture the characteristics of textual data well, and if it lies above a certain threshold, then classify the block as being text.

Determining which coefficients to choose turns out to be an interesting problem. For this purpose we did significant experimentation to determine which set of coefficients could be used as a signature for text blocks. After observing the coefficients at which the textual energy was much higher than the non-textual energy for different images we chose different set of coefficients summarized in Table 4 for an 8x8 block. The coefficients were chosen in two ways. First by doing round robin among the peaks for the energy distribution for text blocks. The second by sorting the maximum difference between the text and non-text DCT coefficients and taking the first few coefficients. The different number of coefficients were chosen to see how many coefficients were required to represent the signature of text blocks accurately. The coefficients are numbered in such a way that the coefficients 0 to 7 represent the first row of coefficients and so on. The set of coefficients which minimized the total error was chosen for DCT-based segmentation. Figure 6 shows results for the different sets of coefficients that we evaluated. It can be seen that the total error is smallest for C1. Thus this set of coefficients is used for our results presented below.

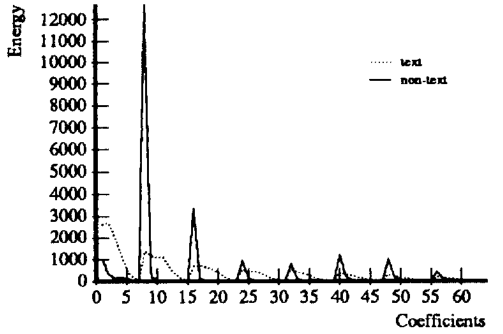


Figure 5.a Average DCT coefficient energy for image T2.

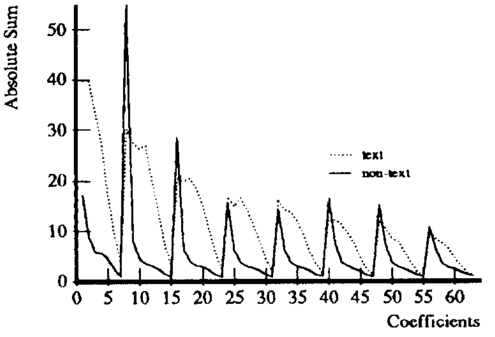


Figure 5.b Average DCT coefficient absolute sum for image T2.

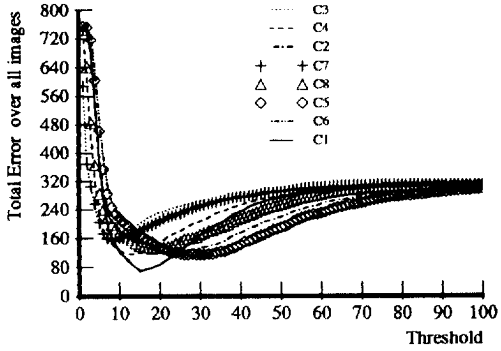


Figure 6. Best DCT coefficients search over all images.

The threshold value of 15 was empirically determined to be the best for segmenting image T2. Figure 7.g shows the results. In contrast to previous approaches, we find that not only are the false-negatives small (6), the false positives are also small (30). Table 2 shows the results for rest of the images. It is seen from this Table that DCT based segmentation is accurate and robust across different images. In fact, as it turns out, the best threshold is 15 for all of the 5 images that we consider.

4.5 Comparative Evaluation of Robustness

In order to compare the robustness of the different segmentation algorithms we choose a constant threshold for each algorithm and then compare results. For our study we choose the threshold to be the average of the threshold for the five images. The resulting thresholds and segmentation statistics are presented in Table 3. In Figures 7.b, 7.d, 7.f, and 7.g we also show the results pictorially for image T2.

It can be seen that for the variance based approach the false-negatives are very high (161) and the false-positives are also very high (161). The absolute deviation approach also gives high (88) false-negatives and very high (197) false-positives. The edge based approach gives high (52) false-negatives and very high (184) false-positives. The DCT based approach gives very small (6) false-negatives and small (30) false-positives. Thus it can be seen that the DCT based segmentation is the least sensitive to the threshold for segmentation.

5. Conclusions

In this paper we have examined four different methods for segmenting text blocks in mixed-mode images. In particular, we have compared our DCT based segmentation approach against more commonly used approaches that rely on block variance, block absolute deviation, and edge detection. For the images evaluated in this paper, DCT based segmentation is shown to have greater success in correctly classifying text blocks, and much higher robustness compared to the other approaches. Although DCT is more expensive than some of the other approaches, it is the most popular transform for block based compression algorithms and has been adopted as the algorithm of choice in JPEG, MPEG and H.261 standards. When used in systems supporting these standards, the incremental computation of summing some of the DCT coefficients will be negligible, thus making DCT-based text segmentation very attractive.

Acknowledgments

Navin Chaddha was supported by an IBM Graduate Fellowship.

References

1. C.T. Chen, "Transform coding of digital images using variable block size DCT with adaptive thresholding and quantization", SPIE vol. 1349, 1990, pp 43-54.

2. P.J. Bones et al., "Segmentation of document images", SPIE vol. 1258, 1990, pp 66-78.

3. S.L. Wood, "Image Processing for segmentation and classification of facsimile images", 24th Asilomar Conference on Signals, Systems and Computers, 1990, pp 887-891.

4. S. Ohuchi et al., "A segmentation method for composite text/graphics documents", Systems and Computers in Japan, Vol. 24, No. 2, 1993.

5. W.K. Pratt, Digital Image Processing, John. Wiley & Sons, New York, 1991.

Table 1. Description of Mixed-mode Images

Image Name	Description
T1: PadH	Pad with hand-written text and hands of a person
T2: ChalkB	Chalk on green board
T3: TextP	Type-written text and photograph of a person
T4: OvP	Text displayed using an overhead projector with hands of a person
T5: VarFont1	Scanned Magazine: text in varying font size and picture of a person

Table 2. Segmentation results based on the best manual adjustment of the threshold.

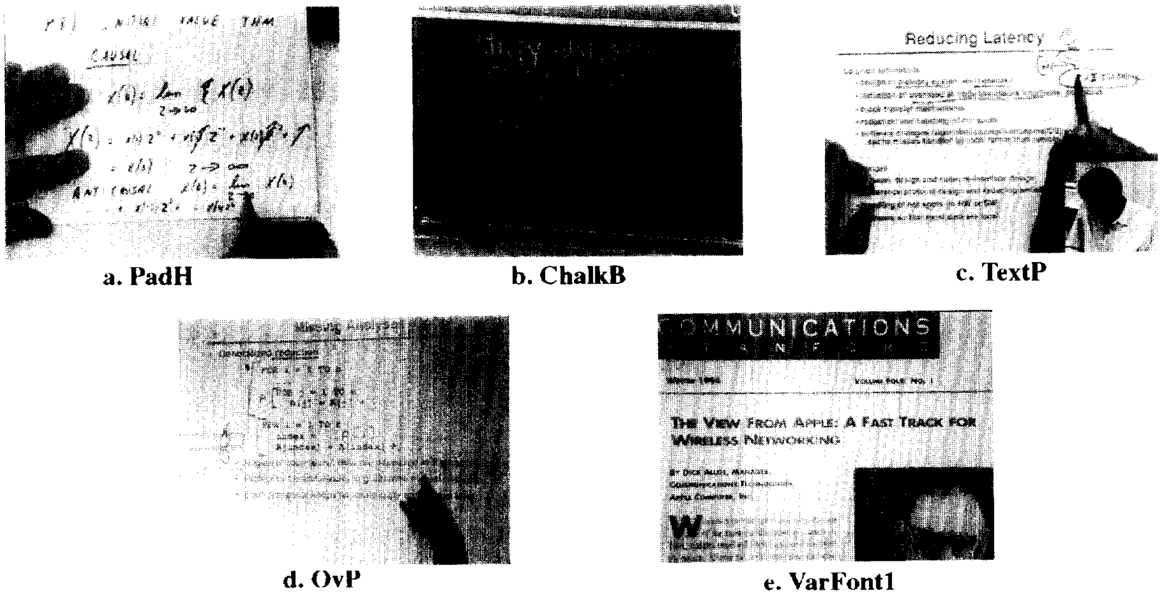
Image Name	Algorithm	Best Thres-hold	No. of misclassified Text Blocks	No. of misclassified Non-Text Blocks
PadH	Var	22	41	42
	AD	16	50	55
	Edge	3, 29	25	34
	DCT	15	7	56
ChalkB	Var	10	5	278
	AD	9	34	246
	Edge	3, 17	9	222
	DCT	15	6	30
TextP	Var	25	19	146
	AD	19	29	140
	Edge	17, 15	35	116
	DCT	15	29	111
VarFont1	Var	14	10	149
	AD	11	20	148
	Edge	25, 7	74	78
	DCT	15	11	31
OvP	Var	10	27	74
	AD	7	25	71
	Edge	13, 7	22	69
	DCT	15	15	48

Table 3. Segmentation results using the average threshold of the 5 images tested.

Image Name	Algorithm	Average Threshold	No. of misclassified Text Blocks	No. of misclassified Non-Text Blocks
PadH	Var	20	34	55
	AD	12	34	75
	Edge	9, 15	23	77
	DCT	15	7	56
ChalkB	Var	20	161	161
	AD	12	88	197
	Edge	9, 15	52	184
	DCT	15	6	30
TextP	Var	20	7	166
	AD	12	10	186
	Edge	9, 15	8	179
	DCT	15	6	30
VarFont1	Var	20	66	126
	AD	12	24	147
	Edge	9, 15	47	129
	DCT	15	11	31
OvP	Var	20	165	21
	AD	12	42	79
	Edge	9, 15	95	40
	DCT	15	15	48

Table 4. Lists of DCT coefficients used for best coefficients search.

List	Coefficients used
C1	3 4 5 11 12 13 19 20 21 43 44 45 51 52 53 59 60 61
C2	2 3 9 10 11 12 16 17 18 19 26 27 28
C3	2 11 18 26
C4	2 11 18 26 35 43 51 58
C5	2 3 4 11 10 9 12 16 19 18 17 20 5 27 26 25
C6	2 3 4 11 10 9 12 16 19 18 17 20 5 27 26 25 13 28
C7	2 3 10 11
C8	2 3 4 11 10 9 12 16



a. PadH

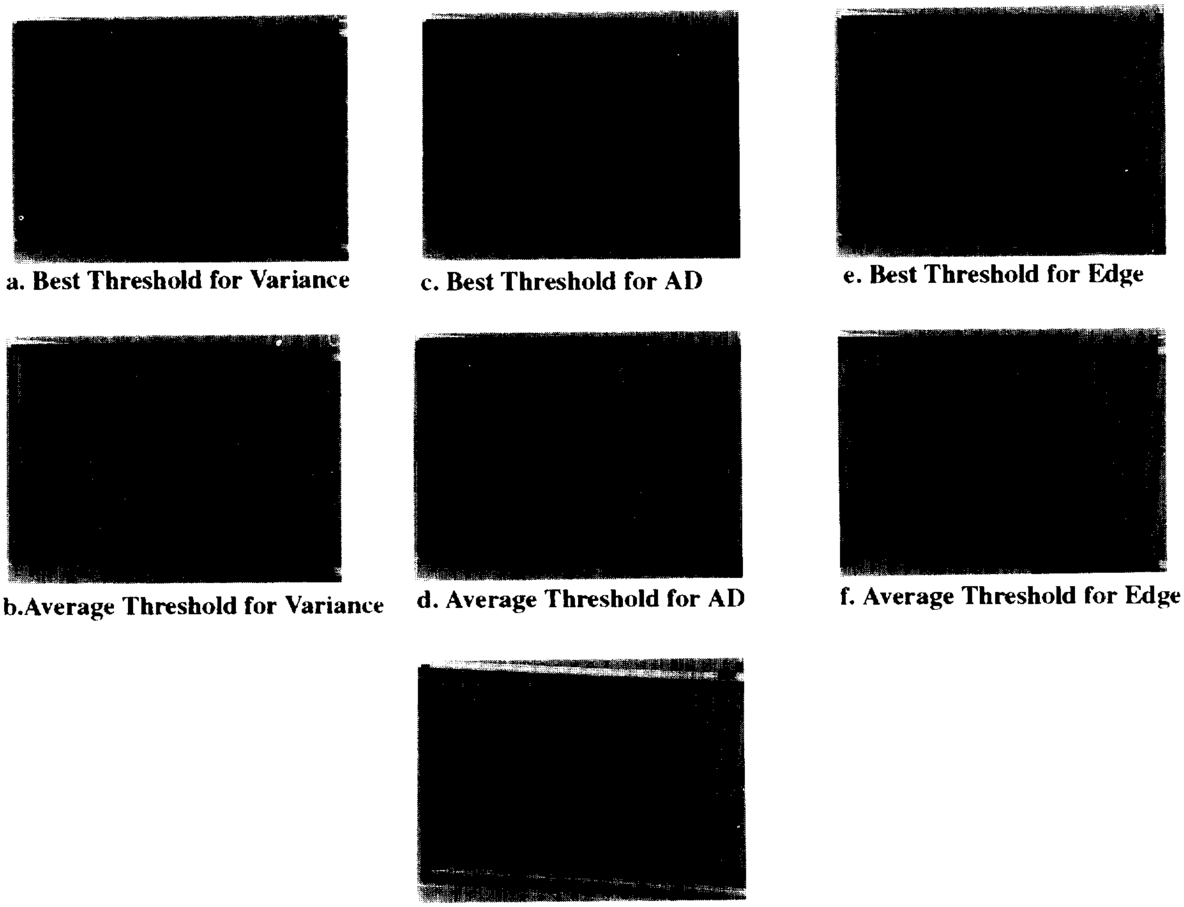
b. ChalkB

c. TextP

d. OvP

e. VarFont1

Figure 1. Test Images



a. Best Threshold for Variance

c. Best Threshold for AD

e. Best Threshold for Edge

b.Average Threshold for Variance

d. Average Threshold for AD

f. Average Threshold for Edge

g. Best and Average Threshold for DCT

Figure 7. Results for Best and Average Threshold on ChalkB