

Recursive Algorithms for Image Segmentation Based on a Discriminant Criterion

Bing-Fei Wu, Yen-Lin Chen and Chung-Cheng Chiu

Abstract— In this study, a new criterion for determining the number of classes an image should be segmented is proposed. This criterion is based on discriminant analysis for measuring the separability among the segmented classes of pixels. Based on the new discriminant criterion, two algorithms for recursively segmenting the image into determined number of classes are proposed. The proposed methods can automatically and correctly segment objects with various illuminations into separated images for further processing. Experiments on the extraction of text strings from complex document images demonstrate the effectiveness of the proposed methods.¹

Keywords— image segmentation, multilevel thresholding, clustering, discriminant analysis

I. INTRODUCTION

In image processing, objects must usually be segmented from an image to facilitate further processing. Accordingly, many researchers have developed valuable techniques for segmenting interesting foreground object from an image [1]-[11], to facilitate applications that include image segmentation, pattern recognition and document analysis, among others. In document analysis, the characters or text strings are preferably well-separated from non-text decorated objects or background images to enable further processing.

The rapid development of multimedia technology has led to increasing numbers of real-life documents, including stylistic text strings with various illuminations, decorated objects and colorful backgrounds. Traditional bi-level thresholding methods [2]-[5] cannot effectively segment all important objects into separate regions. Hence a proper segmentation method is required for segmenting an image into multiple object regions with respective homogeneous features. Tsai [6] developed a method using moment-preserving to select a specified number of thresholds to segment an image into a specific number of object regions. This approach is fast and convinced when the number of thresholds is less than or equal to four. However, the numerical methods are needed when the number of thresholds is more than four. Reddi *et al.* [7] presented an efficient method for thresholding an image into a specified number of object regions. However, this method is performed adequate only when the image contains exactly

three or fewer objects. Cheng *et al.* [8] applied the concept of fuzzy set theory into the principle of maximizing the objective entropy function for determining a set of the optimal thresholds to partition the image into corresponding object regions. Tsai [9] presented a method using Gaussian kernel smoothing to repeatedly perform on the histogram of the image until the histogram being divided into desired number of classes. When the desired number of classes is much lower than the number of peaks in the original histogram, the computation time to find the solutions of threshold values is expensive. Cao *et al.* [10] developed a fast entropy-based approach using the local optimization concept to selecting a specified number of thresholds. The main problem associated with the aforementioned methods is that the number of segments into which the gray-level image should be segmented cannot be automatically determined. Hence all the interesting foreground objects such as characters cannot be automatically and adequately separated from other non-text objects to facilitate further analysis.

Discriminant criterion analysis, as stated in Otsu's literature, [2] is attractive for the simplicity in computation with which it measures the separability among segmented images. It can be implemented as a strategy for recursive segmentation of multiple objects in the image, as in Chereit *et al.*'s method [11]. However, Chereit *et al.*'s method is based on the assumption that the text are located in darkest foreground object region with sufficient differences in illumination with other non-text object regions, and hence its strategy cannot yield sufficient results when bright text objects are concerned or characters lay over other objects with little difference in luminance. In this study, our interest is to develop some image segmentation methods to automatically segment an image into several objects according to its statistical contents based on certain clustering criteria. We develop and test two methods for automatic image segmentation, the recursive optimal multilevel thresholding method and the automatic clustering analysis method, and apply a new discriminant-criterion for measuring the separability among classes, to enable objects of interest to be extracted from complex images, and especially document images.

II. ALGORITHMS FOR RECURSIVELY IMAGE SEGMENTATION

In segmenting objects from a given image, different objects with homogeneous properties should be separated into different segmented images. For this purpose, a criterion

¹ Manuscript received June 24, 2004.

The authors are with Department of Electrical and Control Engineering, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 30050, Taiwan, (phone: +886-3-5131538; fax: +886-3-5712385; e-mail: bwu@cc.nctu.edu.tw).

for measuring the separability among the segmented images with different objects is introduced in this section. By evaluating the proposed separability criterion, the number of objects that the image should be segmented can be automatically determined. In this section, two recursive image segmentation methods based on this criterion are also presented as follows.

A. Discriminant Criterion for Separability Measurement

To segment text, foreground objects, and background components from an image x , the pixels of x must be partitioned into suitable number of classes. For k classes, pixels of x are segmented by applying $k-1$ thresholds $-t_1, \dots, t_n, \dots, t_{k-1}$. These classes are represented by $C_0 = \{0, 1, \dots, t_1\}, \dots, C_n = \{t_n + 1, t_n + 2, \dots, t_{n+1}\}, \dots, C_{k-1} = \{t_{k-1} + 1, t_{k-1} + 2, \dots, L\}$, where L is the upper-bound of the gray values of x .

An efficient criterion for evaluating the results of segmentation is developed to measure the separability among all classes. An automatic multilevel thresholding method can be implemented to segment objects from the image x recursively, regardless of the number of objects and the complexity of the image x . First, based on discriminant-criterion analysis [2], the between-class variance can be derived as,

$$v_{BC} = \sum_{n=0}^{k-1} w_n (\mu_n - \mu_T)^2 \quad (1)$$

where k is the number of classes of pixels and w_n is the cumulative probability mass function of class C_n ; μ_T denotes the overall mean of pixels in x ; μ_n represents the mean of pixels in class C_n ; They are defined as,

$$w_n = \sum_{i=t_n+1}^{t_{n+1}} P(i), \quad (2)$$

$$\mu_n = \sum_{i=t_n+1}^{t_{n+1}} iP(i) / w_n \quad (3)$$

$$P(i) = \frac{m_i}{N}, \quad \mu_T = \sum_{i=0}^L iP(i) \quad (4)$$

where the dummy threshold $t_0 = 0$ is utilized for convenience in simplifying the expression of equation terms; m_i is the number of pixels with gray-level i , and N is the total number of pixels in the x , or equivalently, $N = \sum_{i=0}^L m_i$; and

$P(i)$ is the normalized probability of the occurrence of gray-level i . Equation (1) yields a measure of discriminant of all existing classes decomposed from the x , denoted by the "separability factor" - SF , and defined as,

$$SF = \frac{v_{BC}}{v_T}, \quad v_T = \sum_{i=0}^L (i - \mu_T)^2 P(i) \quad (5)$$

where v_T is the total variance of the gray-level values of image x . The SF value measures the separability among all existing classes, and the SF value lies within the range $0 \leq SF \leq 1$. Maximizing the SF value can be the objective, to

optimize the segmentation result. Intuitively, when SF approximates 1.0, all classes of pixels decomposed from x are ideally and completely separated.

B. Automatic Multilevel Thresholding Method

Hence, a new automatic multilevel thresholding method is developed, based on the analysis of SF . The details of the proposed method are presented below.

Step 1: Initially, only one class C_0 of the image x exists; Let $q = 1$.

Step 2: Currently, q classes exist, having been decomposed from the x . Compute the histogram, the class-mean μ_n , the cumulative probability mass function w_n , and the standard deviation σ_n , of each existing class C_n decomposed from the x .

Step 3: From all classes C_n , determine the class C_p with the maximum standard deviation σ_{\max} , which is to be partitioned in the following step to achieve maximal increment of SF .

Step 4: Partition $C_p : \{t_p + 1, t_p + 2, \dots, t_{p+1}\}$ into two classes C_{p0} and C_{p1} , by applying the optimal partition threshold t_S^* . The C_{p0} and C_{p1} comprise the subsets of gray values derived from C_p and can be represented as: $C_{p0} : \{t_p + 1, t_p + 2, \dots, t_S^*\}$, and $C_{p1} : \{t_S^* + 1, t_S^* + 2, \dots, t_{p+1}\}$. The t_S^* is determined by maximizing the v_{BC} with respect to t_S . Based on the aforementioned definitions, the t_S^* is computed as,

$$t_S^* = \text{Arg Max}_{t_p \leq t_S < t_{p+1}} v_{BC}(t_S) \quad (6)$$

$$v_{BC}(t_S) = w_{p0} (\mu_{p0} - \mu_p)^2 + w_{p1} (\mu_{p1} - \mu_p)^2, \quad (7)$$

$$w_{p0} = \sum_{i=t_p+1}^{t_S} P(i), \quad w_{p1} = \sum_{i=t_S+1}^{t_{p+1}} P(i), \quad (8)$$

$$\mu_{p0} = \sum_{i=t_p+1}^{t_S} iP(i) / w_{p0}, \quad \mu_{p1} = \sum_{i=t_S+1}^{t_{p+1}} iP(i) / w_{p1}, \quad (9)$$

$$w_p = \sum_{i=t_p+1}^{t_{p+1}} P(i), \quad \mu_p = \sum_{i=t_p+1}^{t_{p+1}} iP(i) / w_p \quad (10)$$

where w_p and μ_p are the class-probability and class-mean of C_p , respectively.

Step 5: Step 4 yields $q+1$ classes, C_0, C_1, \dots, C_q . Then, SF of all classes is computed using Eq. (5). If $SF < TH_{SF}$, then let $q = q+1$ and go back to step 2; otherwise, go to step 6.

Step 6: Terminate the thresholding procedure and then classify the pixels into separate classes, according to the resultant threshold values.

This study employs $TH_{SF} = 0.92$, determined from the training using numerous images, such that all existing classes are almost completely separated. Consequently, all objects are recursively segmented into individual object images.

C. Automatic Clustering Analysis

In this section, we would like to introduce another choice of automatic image segmentation method, the automatic clustering analysis using the aforementioned discriminant criterion defined in Eq. (5). This clustering algorithm is utilized to cluster the grayscale image into several objects as that of the multilevel thresholding approach described in the previous section. This clustering analysis is an unsupervised method for automatically separating image objects. The gray scale pixels of the original image will be classified into different clusters. Each cluster contains one object with similar gray values. The clustering algorithm is shown below.

Step 1: Calculate the mean value, m , and the standard deviation, σ , of the image x , where $x(i, j)$ denote the pixel located at (i, j) coordinate.

Step 2: Then $x(i, j)$ is split according to mean and standard deviation of the processed sub-block image. Define two centers, C'_1 and C'_2 , by,

$$C'_1 = m + 0.5 \times \sigma \text{ and } C'_2 = m - 0.5 \times \sigma \quad (11)$$

Step 3: Calculate the Euclidean distance from each pixel $x(i, j)$ to C'_1 and C'_2 , using the following equalities, respectively.

$$D'_{ij,1} = |x(i, j) - C'_1| \text{ and } D'_{ij,2} = |x(i, j) - C'_2| \quad (12)$$

Then, $x(i, j)$ is partitioned into two clusters ψ_k ($k=1, 2$) as,

$$\psi_1 : \{x(i, j) | D'_{ij,1} \leq D'_{ij,2}\}, \text{ and } \psi_2 : \{x(i, j) | D'_{ij,1} > D'_{ij,2}\} \quad (13)$$

Step 4: As stated in Eq.(5) of previous section, to compute the separability factor among k existing clusters, the computation of cluster probabilities, w_n (corresponding to Eq.(2)), cluster means, μ_n (corresponding to Eq. (3)) of each cluster ψ_n can be computes as,

$$w_n = \sum_{r \in \psi_n} P(r), \quad \mu_n = \sum_{r \in \psi_n} P(r) / w_n \quad (14)$$

where m_r is the number of pixels with gray-level r and N is the total number of pixels in the cluster ψ_n ; and

$P(r) = \frac{m_r}{N}$ is the normalized probability of the occurrence of gray-level r . Then the separability factor SF among all existing clusters can be determined using Eq. (5).

Step 5: Calculate the standard deviation σ_n of each existing clusters ψ_n ($n=0, 1, \dots, k-1$). If $SF < TH_{SF}$ and $\sigma_{n_{\max}} > TH_{\sigma}$, then determine the maximum standard deviation $\sigma_{n_{\max}}$ among σ_n of all existing clusters, then split the cluster of $\sigma_{n_{\max}}$. Else, terminate the clustering process.

Step 6: Step 5 yields three clustering centers C_k ($k=0, 1, 2$). $x(i, j)$ is partitioned into three clusters ψ_k ($k=0, 1, 2$).

Step 7: Calculate the cluster probabilities, w_n and cluster means, μ_n of each cluster and the separability factor SF , among the clusters ($n=0, 1, \dots, k-1$ and $k > 2$). If $SF < TH_{SF}$ and $\sigma_{n_{\max}} > TH_{\sigma}$, then split the cluster with maximum standard deviation $\sigma_{n_{\max}}$ into two clusters. Next, repeat Step 7. Else, terminate the clustering process.

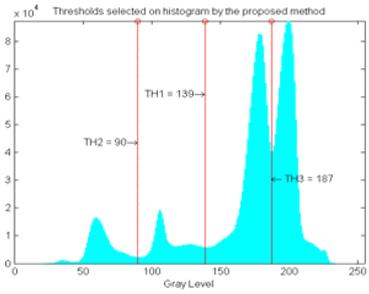
Repeat Steps 2~7 until the image $x(i, j)$ have been segmented into appropriate number of objects. This clustering analysis algorithm in this study utilizes $TH_{SF} = 0.92$ as applied in aforementioned multilevel thresholding approach, and $TH_{\sigma} = 14$.

In the multilevel thresholding and clustering analysis approaches, the parameters TH_{SF} is the threshold values to decide which cluster is convergence when the conditions, $SF > TH_{SF}$ and $\sigma < TH_{\sigma}$, are satisfied. The SF value measures the separability among all existing clusters in clustering analysis as utilized in the multilevel thresholding process. The SF value may lies within the range $0 \leq SF \leq 1$, as described in section 2.1. Hence, when the SF approximates 1.0, all the existing clusters are ideally and completely separated. When the number of the clusters is more than two, the SF value is used to measure the separability among the clusters. In this study, we utilize $TH_{SF} = 0.92$ in both multilevel thresholding and clustering analysis approaches.

In the clustering analysis approach, the standard deviation, σ , measures the compactness of the pixel values of each clusters. Ideally, the σ approximates zero for a monochromatic object. In a pilot experiment, we analyze the widespread distributions, caused by the scanner or the original document, of the pixel values of monochromatic texts in different document images. The average variation of the monochromatic texts with different size or style is around 0~50. In general, the $TH_{\sigma} = 25$ can obtain good preservation of the texts, but it is insufficient for our needs. Therefore, this study employs $TH_{\sigma} = 14$ to obtain better outcome, when the texts overlap a background with rapidly varying texture and similar grayscale.



(a). Original image.



(b). Thresholds selected by multilevel thresholding method



(c). The dark foreground object region obtained by Otsu's method



(d). The dark foreground region obtained by Chereit *et al's* method



(e). The foreground object region 1 (dark characters) by automatic multilevel thresholding method



(f). The foreground region 2 (light characters) obtained by automatic multilevel thresholding method



(g). Segmented region 3 obtained by automatic multilevel



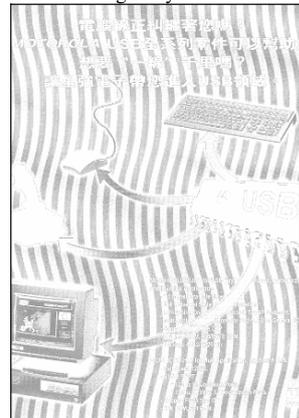
(h). Segmented region 4 obtained by automatic multilevel



(i). The foreground object region 1 (dark characters) obtained by clustering analysis method



(j). The foreground object region 2 (light characters) obtained by clustering analysis method



(k). Segmented region 3 obtained by clustering analysis method



(l). Segmented region 4 obtained by clustering analysis method

Figure 1. Result of segmenting the test image, "USB" (scanned by 300dpi, image size=1600×2304)

III. EXPERIMENTAL RESULTS

In this section, we utilize some real-life, true color, full-page, complex document images to perform the performance evaluation. They are transformed into gray-scale images with 256 gray-levels and then processed by the Otsu's method [2], the Chereit *et al's* method [11], the proposed multilevel optimal thresholding method and clustering analysis method, to analyze and evaluate their performance. Figures 1(a) and 2(a) show two of our test images – "USB" and "Chung-Hsin". They have several text strings with various styles, some decorated objects that overlay the text strings, and a complex and varying background. In this study, we adopt the "uniformity measure" from Levein and Nazif's literature [12] to evaluate the segmentation performance of the methods. The uniformity measure is defined as,

$$U = 1 - \frac{v_{WC}}{v_T}, \quad 0 \leq U \leq 1 \quad (15)$$

where v_{WC} is the within-class variance for all segmented classes of pixels and can be computed as,

$$v_{WC} = w_0\sigma_0^2 + \dots + w_n\sigma_n^2 + \dots + w_{k-1}\sigma_{k-1}^2 \quad (16)$$

where σ_n^2 is the class variance of each class of pixels. If the image is well-segmented, the score of uniformity measure should close to 1.

The first test image, "USB", a real-life, full-page, complex document image, is shown in Fig. 1(a). Figure 1 shows the segmentation results of the first test image processed by the four methods. By evaluating the processing results of this complex document image, the thresholding performance of the proposed can be distinctly demonstrated. As shown in Figs. 1(c) and 1(d), we can see that the dark foreground object regions obtained by Otsu's and Chereit *et al*'s methods are alike. It is because that the segmentation strategy of Chereit *et al*'s method cannot appropriately determine whether the segmented foreground image should undergo further segmentation to completely separate all interesting image objects such as character patterns from the background components. It is obvious that both methods cannot properly segment all interesting foreground objects into separate regions. Ideally, dark black characters and light characters should be segmented into two thresholded images, and the non-text objects overlaying them should be extracted into other segmented regions. For example, the arrow, crossed and superimposed on the black characters in the bottom-right corner of the figure should be separated out. After the proposed optimal multilevel thresholding method has been performed, three suitable thresholds on the histogram are determined and presented in Fig. 1(b). Figures 1(e)-(h) indicate that the thresholds are properly selected by the proposed multilevel thresholding method to segment clearly different objects and background components into several independent segmented images. Figures 1(i)-(l) are the segmentation results of the proposed clustering analysis method. To observe the results of the two approaches, we can find that they both extract the text strings with different illuminations are very well. As shown in Figs. 1(e)-(f) and 1(i)-(j), the black and light characters are successively and clearly separated from the background components and extracted into separate regions. As for the segmented background image of both methods, as shown Fig. 1(g)-(h) and 1(k)-(l), we can see that the background textures segmented by the automatic multilevel thresholding method have somewhat better effects than those of clustering analysis method.

Figure 2 shows another experimental result of the test image 2 -- "Chung-Hsin", which comprises several black characters overlapped with a highly complex background with several objects and textures. Since the extraction of characters is the main focus, we concentrate on the dark foreground regions derived by the four methods in this experiment. They are expected to comprise clear dark characters. Figure 2(b) displays the segmented dark foreground region obtained by performing Otsu's method. This shows that the characters cannot be extracted from complex background objects by only one single threshold.

Figure 2(c) is the segmented dark foreground region derived by Chereit *et al*'s method. Although some bright objects are further segmented, the characters are still not sufficiently separated from other background components. In contrast, as shown in Fig. 2(d), we can find the dark characters are successfully and clearly segmented from the background objects using the proposed automatic multilevel thresholding method, while other decorated non-text objects with different illuminations are separated into the other four segmented regions. The proposed clustering method also segment Fig. 2(a) into five segmented regions. As shown in Fig. 2(e), this segmented region contains characters extracted from Fig. 2(a), although coexisted with slight non-text components, the characters are clearly separated from complex background objects.

TABLE I. Experimental data of Fig. 1(a) performed by Otsu's method, proposed Automatic Multilevel thresholding method(AMT) and proposed Automatic Clustering Analysis Method(ACA)

Method	segmented region no.	Uniformity measure
Otsu's	2	0.784
Cheriet's	2	0.784
AMT	4	0.94
ACA	4	0.761

TABLE II. Experimental data of Fig. 2(a) performed by Otsu's method, Automatic Multilevel thresholding method and Automatic Clustering Analysis Method

Method	segmented region no.	Uniformity measure
Otsu's	2	0.702
Cheriet's	3	0.817
AMT	5	0.926
ACA	5	0.716

Table 1 and 2 show the experimental data of Fig. 1(a) and Fig. 2(a) processed by the Otsu's method, the proposed automatic multilevel thresholding method and automatic clustering analysis method. As seen in Table 1 and 2, we can find that the proposed automatic multilevel thresholding achieves highest scores of uniformity measures, and the objects are well-segmented from original images. Although the uniformity measure scores of the automatic clustering analysis method are similar to those of Otsu's bi-level thresholding method and not outperform Chereit *et al*'s method, the segmented objects are clear and compact, and the visual effect of the segmented objects are as better as those obtained by automatic multilevel thresholding. Both the proposed multilevel thresholding method and clustering analysis method provide fine visual effect of segmentation results and the segmented objects are sufficiently distinct for further processing, such as text extraction.

Summarily, text strings with various illuminations and styles, decorated objects and different backgrounds are effectively separated into different, segmented images by performing the proposed segmentation methods. For further

document analysis, text strings can be easily extracted from the segmented images using a general text extraction method.

Figure 2. Result of segmenting the test image 2, “Chung-Hsin”, (scanned by 300dpi, image size=1842×3310)

IV. CONCLUSION

Two efficient recursive image segmentation methods were developed in this study. An effective discriminant-criterion for determining separability among classes can simplify the problem of determining the number of object images that should be segmented. The original image can be recursively segmented into more detailed object images until the discriminant criterion is met. Accordingly, all the objects of interest can be partitioned into separate segmented images. When applied to real-life, complex document images, the proposed method can successfully extract text strings with various illuminations from overlaying non-text objects or complex backgrounds, as determined experimentally. Furthermore, the proposed method can be applied not only to complex document images, but also to objects with various illuminations in various types of images.

REFERENCES

- [1] Sahoo P.K., Soltani S., Wong A.K.C., Chen Y.C., “A survey of thresholding techniques”, *Computer Vision, Graphics and Image Processing*, vol. 41, pp. 233-260, 1988.
- [2] Otsu, N., “A threshold selection method from gray-level histograms”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-9, pp. 62-66, 1979.
- [3] Kapur, J., Sahoo P.K., Wong A.K.C., “A new method for gray-level picture thresholding using the entropy of the histogram”, *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 273-285, 1985.
- [4] Kittler, J. and Illingworth, J., “Minimum Error Thresholding”, *Pattern Recognition*, vol. 19, pp. 41-47, 1986.
- [5] Li, C.H. and Lee, C.K., “Minimum cross entropy thresholding”, *Pattern Recognition*, vol. 26, pp. 617-625, 1993.
- [6] Tsai W.H., “Moment-preserving thresholding: A new approach”, *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 277-393, 1985.
- [7] Reddi, S.S., Rudin, S. F., and Keshavan, H.R., “An optimal multiple threshold scheme for image segmentation”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-14, pp. 661-665, 1984.
- [8] Cheng, H.D., Chen, Y.H., and Sun, Y., “A Novel Fuzzy Entropy Approach to image enhancement and thresholding”, *Signal Processing*, vol. 75, pp. 277-301, 1999.
- [9] Tsai, D.M., “A fast thresholding selection procedure for multimodal and unimodal histograms”, *Pattern Recognition Letters*, vol. 16(6), pp. 653-666, 1995.
- [10] Cao L., Shi Z.K. and Cheng E.K.W., “Fast automatic multilevel thresholding method”, *Electronics Letters*, vol. 38(16), pp. 868-870, 2002.
- [11] Cheriet M., Said J.N., and Suen C.Y., “A recursive thresholding technique for image segmentation”, *IEEE Transactions on Image Processing*, vol. 7(6), pp. 918-921, 1998.
- [12] Levine, M.D. and Nazif, A.M., “Dynamic measurement of computer generated image segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 155-164, 1985.



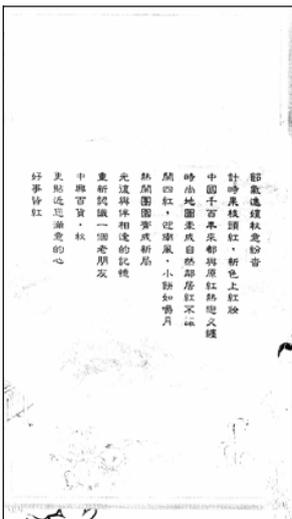
(a) Original image



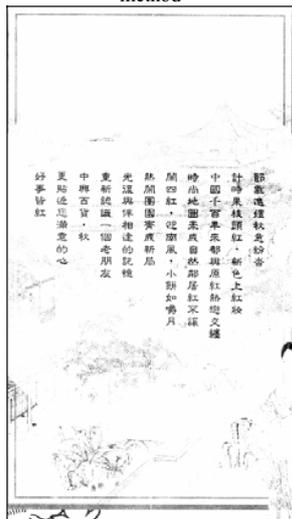
(b). The darkest foreground object region obtained by Otsu's method



(c).The darkest foreground object region obtained by Cheriet et al's method



(d). The darkest foreground object region obtained by automatic multilevel thresholding method



(e). The darkest foreground object region obtained by clustering analysis method