ELSEVIER

Pattern Recognition Letters 20 (1999) 329–336

Pattern Recognition Letters

Two-dimensional entropic segmentation [1]

E. David Jansing [a,*], Thomas A. Albert [b], Darrel L. Chenoweth [a]

[a] *University of Louisville, Speed Scientific School, Department of Electrical Engineering, Louisville, KY 40292, USA*

[b] *University of Louisville, School of Medicine, Center for Applied Microcirculatory Research, Louisville, KY 40292, USA*

Received 11 February 1998; received in revised form 27 November 1998

Abstract

A novel method of 2-D entropic segmentation using a linear discriminant function is presented. This segmentation method automatically highlights desired objects against background with no user input or parameter specification. Improved class separation, and therefore, reduced classification error can be obtained by adding a second feature to the gray level histogram. Examples of aerial and medical images were segmented using features such as intensity, and fractal error or standard deviation with lower segmentation errors than 1-D and 2-D entropy-based methods. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Entropy; Segmentation; Classification; Remote sensing; Automatic thresholding; Texture

1. Introduction

Segmentation is an important task in many image-processing systems. Automatic target recognition often uses segmentation to separate the desired target from the background. Segmentation is also used to highlight certain parts of the image which may be difficult to see with the human eye, such as man-made objects in synthetic aperture radar (SAR) imagery. The segmentation process is an example of a linear classifier with two classes, i.e., object and background. Thresholding intensity values is the simplest method of segmentation. This type of thresholding implies that the object or objects in question have intensities that are distinctly different from the background or other objects in the scene, e.g., a black part on a white conveyor belt.

Threshold selection methods can be grouped into two categories, local methods and global methods. Segmenting an entire image with a single value using the gray-scale histogram is an example of a global method. Local methods partition an image into a group of sub-images and select a threshold point or set of threshold points for each of the sub-images. Global thresholding techniques are easy to implement, but some methods tend to be inaccurate, especially with complex images. The reader is encouraged to references (Sahoo et al., 1988; Wong and Sahoo, 1989) for excellent surveys of global thresholding methods using information theory.

This paper presents a novel segmentation method after a technique presented by Sahoo et al. (1997). This method automatically finds a thresholding line using the information from a 2-D frequency distribution of the chosen image features.

* Corresponding author. Address: Apt. 8403, 10207 Baltimore Avenue, College Park, MD 20740, USA. Tel.: +301-614-3220; fax: +301-614-3220; e-mail: dr_jansing@mindspring.com

[1] Electronic Annexes available. See http://www.elsevier.nl.locate.patrec

PII: S0167-8655(98)00151-2

Sahoo et al. proposed an entropic global thresholding method that selects a threshold point based on minimizing the difference between entropies of the object and the background distributions of the probability density function. The method presented here uses this thresholding point along with statistical information from the assumed data sets of the object and the background to produce a second point. The line produced from the two thresholding points is then used to separate the object data set from the background data set. To demonstrate the robustness of this method, data were derived from an aerial image in which the object data set is the cultural objects present, and an MRI image of a cross section of the brain where the ventricles represent the objects of interest.

2. General theory

Sahoo et al. (1997) published an excellent review of the theory behind their entropic thresholding method. It will be summarized here.

Thresholding typically uses the gray-level histogram of an image. Assuming a binary image, let $B = b_0, b_1$ be the pair of binary gray levels. B is the partition that separates the object from the background in the image. Mathematically, this partition is defined as

$$j(r,c) = \begin{cases} b_0 & \text{if } i(r,c) \leqslant t, \\ b_1 & \text{if } i(r,c) > t. \end{cases} \tag{1}$$

where $i(r,c)$ is the gray-scale image, t is the gray-scale threshold, and $j(r,c)$ is the resulting binary image. In general, thresholding methods often determine the value of t, using statistical information from the image. The gray-scale histogram can approximate the density function

$$h(x) = \mathrm{Prob}[f(m,n) = x] \tag{2}$$

by using the expression

$$\hat{h}(x) = \frac{g_x}{g_{\mathrm{total}}}, \tag{3}$$

where $\hat{h}(x)$ is the resulting histogram of the image, g_x is the number of pixels with the gray-scale value x, and g_{total} is the total number of pixels in the image. g_{total} is equal to MN if the image is $M \times N$.

Let $p_i = \hat{h}(i) = n_i/MN$ define the estimate of the probability gray-level value, where n_i is the number of elements in the event $[f(m,n) = x]$. The a priori entropy is defined as

$$H_T = -\sum_{i=0}^{255} p_i \ \ln \ p_i \tag{4}$$

for the entire image. Assuming two classes of pixels, where b_0 denotes the class of "black" pixels (gray-scale value of 0) and b_1 denotes the class of "white" pixels (gray-scale value of 255), the a priori entropies for each class of pixels are

$$H_{b_0}(t) = -\sum_{i=0}^{t} \frac{p_i}{p(b_0)} \ln \frac{p_i}{p(b_0)}, \tag{5}$$

$$H_{b_1}(t) = -\sum_{i=t+1}^{255} \frac{p_i}{p(b_1)} \ln \frac{p_i}{p(b_1)}, \tag{6}$$

where

$$p(b_0) = \sum_{i=0}^{t} p_i, \tag{7}$$

$$p(b_1) = \sum_{i=t+1}^{255} p_i. \tag{8}$$

Note that $p(b_0) + p(b_1) = 1$. Kapur et al. (1985) defined the information between the two classes to be

$$\begin{aligned} \Phi(t) &= H_{b_0}(t) + H_{b_1}(t) \\ &= \ln[p(b_0)p(b_1)] + \frac{H(t)}{p(b_0)} + \frac{H_T - H(t)}{p(b_1)}, \end{aligned} \tag{9}$$

where $H(t)$ is defined as

$$H(t) = -\sum_{i=0}^{t} p_i \ \ln \ p_i. \tag{10}$$

Kapur et al. selected the threshold to be the value of t^* at which $\Phi(t)$ is maximum. They noted that since some p_i will be very small, care should be taken in evaluating the term $\ln p_i$.

Sahoo et al. (1997) defined the optimal threshold t^* by minimizing the difference of the entropies $H_{b_0}(t)$ and $H_{b_1}(t)$. The function

$$E(t) = [H_{b_0}(t) - H_{b_1}(t)]^2 \tag{11}$$

measures the difference in the a priori entropies of classes b_0 and b_1. They pointed out that the smaller $E(t)$, the more homogeneous are b_0 and b_1. Thus, the global minima of $E(t)$ can be used to find t^*. This is expressed as

$$t^* = \arg\min E(t). \tag{12}$$

This method proved to be more accurate in some applications than other entropic thresholding methods, such as (Kapur et al., 1985). One drawback to this idea is that there is an implied assumption that the a priori probabilities of the object and background classes are approximately equal. The quality of the resulting segmentation decreases as the difference between these probabilities increases.

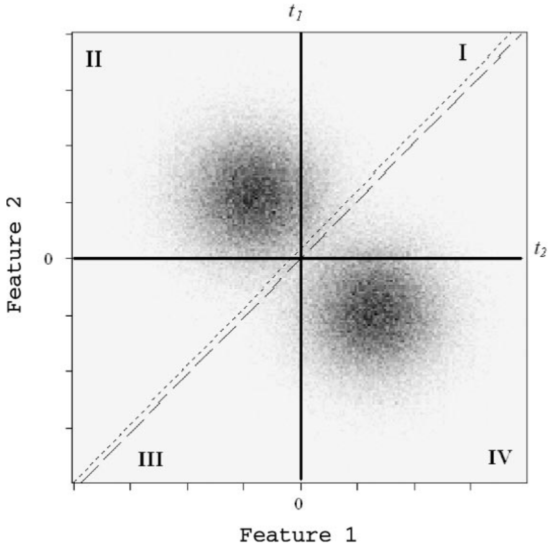

Fig. 1. Joint frequency distribution for two Gaussian data sets. The dashed line represents the Bayesian decision line, and the dotted line the 2-D entropic decision line.

3. 2-D extension of the entropy crossover method

This section will outline the 2-D extension of the entropy crossover method described in (Sahoo et al., 1997) as well as the novel technique for the proposed extension. The information in the gray-level frequency distribution (histogram) is often not enough to accurately segment a given image. In these cases, it is not uncommon to incorporate additional information in the histogram. The additional information results in a 2-D histogram which is the frequency distribution of the second feature combined with the gray-level frequency distribution. This 2-D histogram is an estimate of the joint probability density function of the two features. The co-occurrence matrix is a commonly used example of this type of 2-D histogram. Obviously, the choice of features is extremely important in determining the separability of the object and background classes in the joint probability density function. By choosing image features which are linearly uncorrelated, maximum separation of classes may be obtained.

Abutaleb (1989) proposed that the best threshold point for a 2-D histogram can be found by locating an optimal threshold for each feature in the histogram, which results in two separate thresholds, $t1^*$ and $t2^*$. When the threshold for each feature has been found, the orthogonal lines segmenting each feature divide the matrix into four quadrants. The intersection of the orthogonal lines produces the overall thresholding point $(t1^*, t2^*)$, as shown in Fig. 1.

Brink (1992) used two of the quadrants for segmentation. He segmented quadrant II pixels as background and quadrant IV pixels as foreground (see Fig. 1). However, he discarded the pixels located in quadrants I and III, which may ignore

Table 1
Automatic 2-D entropic segmentation algorithm

1.	Calculate the optimal thresholding point $(t1^*, t2^*)$ using the entropy crossover method as described in (Sahoo et al., 1997).
2.	Divide the 2-D histogram into quadrants around $(t1^*, t2^*)$.
3.	Determine the quadrant, j, with the largest number of points and calculate its central moment, $(\bar{x}, \bar{y})_j$.
4.	Compute the line, $l(x, y)$, that intersects $(t1^*, t2^*)$ and $(\bar{x}, \bar{y})_j$.
5.	Define $L(x, y)$ to be the line perpendicular to $l(x, y)$, passing through $(t1^*, t2^*)$.
6.	Segment the image using $L(x, y)$ as the optimal thresholding line.

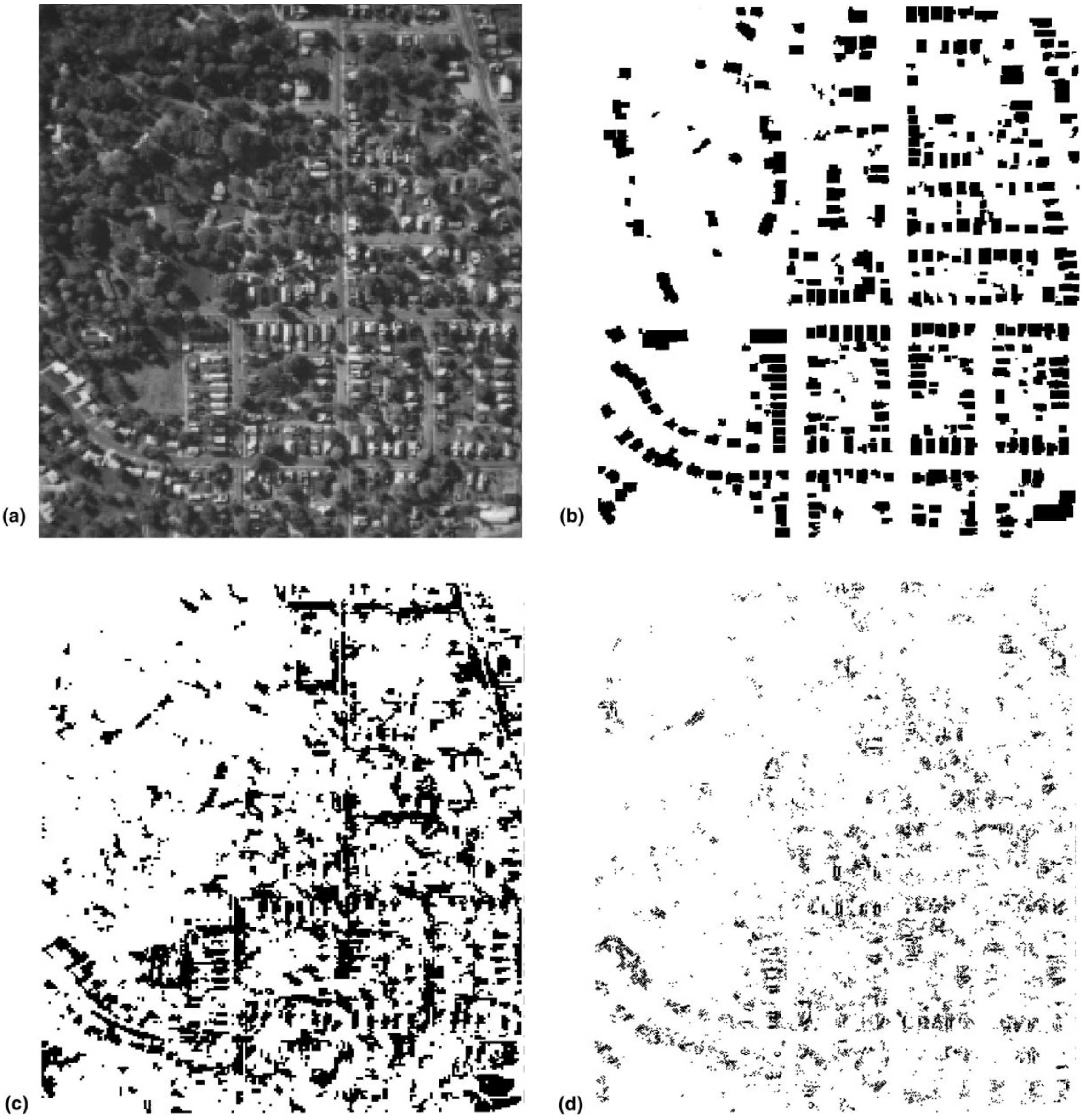

Fig. 2. Segmentation of suburban aerial image. (a) Original image. (b) Ground truth. (c) 1-D entropic segmentation using intensity. (d) 1-D entropic segmentation using fractal error.

important information concerning the objects to be segmented. Sahoo et al. recognized that the selection of a single point to classify the 2-D pdf may completely disregard some sections of the desired distribution that reside near, but not in, the foreground or background quadrants. Instead of thresholding in quadrants, they proposed a thresholding line in the 2-D histogram plane to provide better segmentation. Such a line is shown in Fig. 1. They suggested finding the optimal

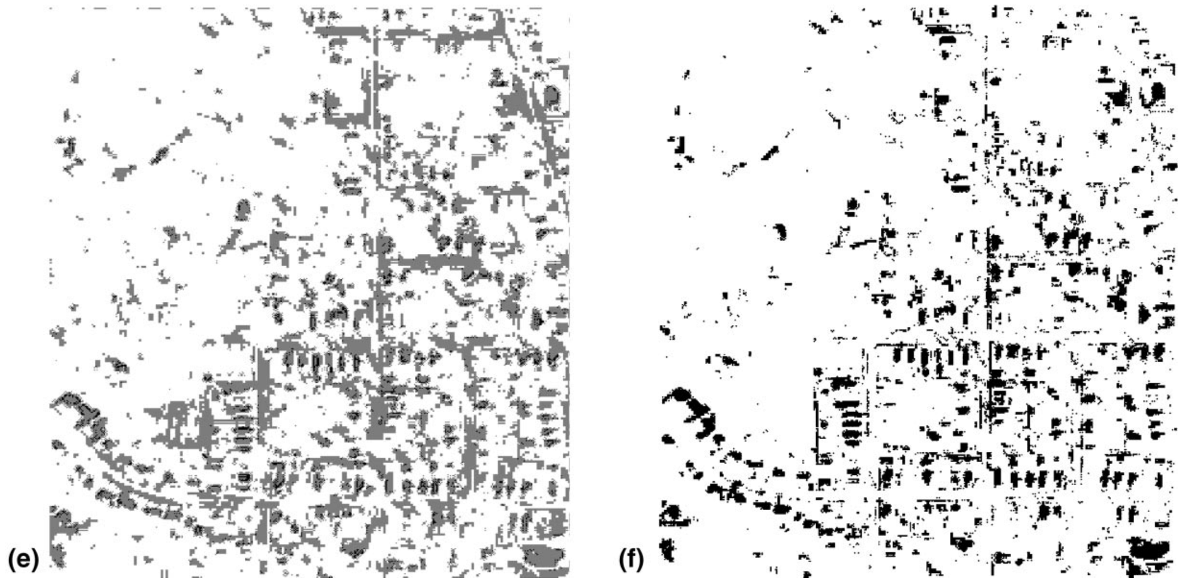

Fig. 2. (e) Segmentation using Brink's Method with intensity and fractal error as features. (f) Segmentation using 2-D entropic segmentation with intensity and fractal error as features.

thresholding line by minimizing the difference in the entropy values on either side of a proposed line in the 2-D histogram plane, not unlike the method used in finding the optimal point in the 1-D histogram. This line represents the least computationally intensive (linear) classifier for a 2-D space. However, this is still a costly method, as there are approximately 4×10^5 lines that bisect a 256×256 histogram.

Instead of an exhaustive search of the potential thresholding lines, it is possible to calculate a thresholding line using the statistics of each quadrant. Let $(\bar{x}, \bar{y})_j$ be the central moment of quadrant $j = \mathrm{I}, \ldots, \mathrm{IV}$. Thus,

$$\bar{x} = \frac{\sum x_i}{Q_j}, \tag{13}$$

$$\bar{y} = \frac{\sum y_i}{Q_j}, \tag{14}$$

where Q_j is equal to the total number of points in quadrant j. The quadrant with the most points is determined, and its central moment is used with the optimal threshold point to form a line. The resulting line is perpendicular to the desired thresholding line. The algorithm is summarized in Table 1.

4. Direct texture measurement

One very useful feature for segmentation is image texture. While there is no formal definition of texture, there are several means by which to measure different properties of texture, such as smoothness, coarseness, and regularity (Gonzalez and Woods, 1992). These properties may be quantified using statistical, structural, spectral, or fractal methods. An example of a statistical texture measure is standard deviation, which quanitifies dissimilarity of pixels within a localized neighborhood (Albert et al., 1991). Cooper et al. (1994) described a fractal analysis approach to texture called "fractal error". This metric is based on the observation that natural features in high-resolution gray scale aerial images fit the fractional Brownian motion (fBm) mathematical model, as described by Mandelbrot (1977).

Cooper (1994) used gray-scale values and fractal error as features in a 2-D statistical classifier to successfully detect objects in aerial images. Jansing et al. (1997) demonstrated that fractal error was useful in identification of cultural objects in SAR imagery. Because of their proven ability to differentiate between textures, fractal error or local

Table 2
Segmentation results for various data sets

Data	Method	Type I error (%)	Type II error (%)
Gaussian data	1-D entropic with feature 1	6.8	6.2
	1-D entropic with feature 2	7.6	4.9
	Bayesian classifier	1.9	1.3
	Brink's method with feature 1 and feature 2 (24.1% unclassified)	5.0	2.0
	2-D entropic with feature 1 and feature 2	2.4	1.0
Aerial image	1-D entropic with intensity	10.5	13.9
	1-D entropic with fractal error	15.4	1.7
	Brink's method with intensity and fractal error (21.5% unclassified)	9.5	0.4
	2-D entropic with intensity and fractal error	4.0	12.0
MRI image	1-D entropic with intensity	1.3	6.3
	1-D entropic with standard deviation	12.4	2.9
	Brink's method with intensity and standard deviation (18.5% unclassified)	1.3	0.8
	2-D entropic with intensity and standard deviation	4.4	2.6

standard deviation, along with the pixel gray level value were used as features in the segmentation algorithm.

5. Results

The 2-D entropic segmentation was tested with two classes of Gaussian data, shown in Fig. 1. Class 1 had a distribution of $\mathrm{N}([1.5\ 1.5]^{\mathrm{T}}, I)$ and class 2 a distribution of $\mathrm{N}([-1.5 - 1.5]^{\mathrm{T}}, I)$, where I is the identity matrix. The decision line found by the 2-D entropic method was $f(x) = -0.96x + -0.092$, which is very close to the optimum Bayesian decision boundary of $g(x) = -x$. Table 2 shows the results of segmentation using the 2-D entropic segmentation, as well as the 1-D entropic segmentation, and Brink's method. Note that only in Brink's method is there the possiblity for pixels that may be "unclassified". For the Gaussian data sets, the 2-D entropic segmentation comparable to the Bayesian classifier, which relies on a priori information of the classes to calculate the decision boundary.

Table 2 also shows the results of applying each of these methods to an aerial image and an MRI image. In each case, the 2-D entropic segmentation is compared with 1-D entropic segmentation using each of the features and Brink's method using both features. In each case, the 2-D entropic segmentation produced a lower overall error than the other segmentation methods. Figs. 2 and 3 show the original images, the correctly segmented images, and the results of each segmentation method.

The Type II errors for our method using the suburban aerial image may be due, in part, to the occlusion of houses and other cultural objects by the trees in the image. The ground truth did not include the streets visible in the image, however, these cultural objects were also correctly segmented using the method. Therefore, the results of classification of man-made objects was actually better than 16% for the 2-D entropic segmentation.

6. Conclusions

A novel method of automatic 2-D entropic segmentation was presented. It was shown that this method is comparable to a Bayesian classifier for two-class segmentation, while no a priori information about the desired objects was needed. While the results for the test images are fairly accurate, it is clear that success with real-world images is dependent on an intelligent choice of features. Adding a second feature to the histogram improves the ability to segment objects, but the features must be linearly uncorrelated to provide the best separability of

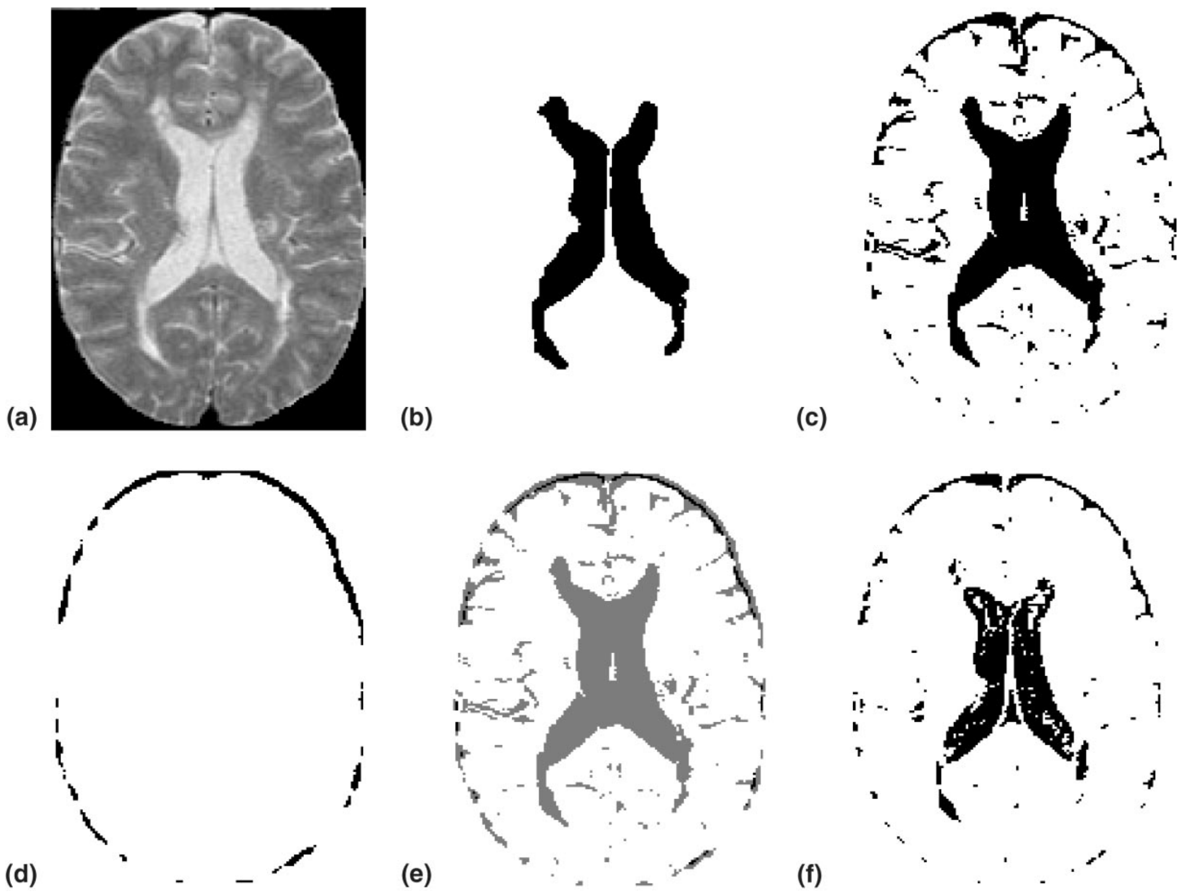

Fig. 3. Segmentation of MRI image. (a) Original image. (b) Ground truth. (c) 1-D entropic segmentation using intensity. (d) 1-D entropic segmentation using 5×5 standard deviation. (e) Segmentation using Brink's Method with intensity and 5×5 standard deviation as features. (f) Segmentation using 2-D entropic segmentation with intensity and 5×5 standard deviation as features.

classes in the joint frequency distribution. A linear classifier can then be rapidly computed which provides a decision surface between the two classes with a minimal amount of error.

References

Abutaleb, A.S., 1989. Automatic thresholding of gray-level pictures using two-dimensional entropy. Computer Vision, Graphics and Image Processing 47, 22–32.

Albert, T.A., O'Conner, C.A., Harris, P.D., 1991. Automatic segmentation of microvessels using textural analysis. Proceedings of the SPIE, 1450, 84–89.

Brink, A.D., 1992. Thresholding of digital images using two-dimensional entropies. Pattern Recognition 25 (8), 803–808.

Cooper, B.E., 1994. Fractional Brownian motion for representation of natural image texture. Ph.D. Thesis, University of Louisville.

Cooper, B.E., Chenoweth, D.L., Selvage, J.E., 1994. Fractal error for detecting man-made features in aerial images. Electronics Letters 30 (7), 554–555.

Gonzalez, R.C., Woods, R.E., 1992. Digital Image Processing. Addison-Wesley, Reading, MA.

Jansing, E.D., Chenoweth, D.L., Knecht, J., 1997. Feature detection in synthetic aperture radar images using fractal error. In: Proceedings of the IEEE Aerospace Conference, Vol. 1, pp. 187–195.

Kapur, J.N., Sahoo, P.K., Wong, A.K.C. 1985. A new method for gray level picture thresholding using the entropy of the histogram. Computer Vision, Graphics and Image Processing 29, 273–285.

Mandelbrot, B.B., 1977. Fractals: Form, Chance and Dimension. Freeman, San Francisco.

Sahoo, P.K., Slaaf, D.W., Albert, T.A., 1997. Threshold selection using a minimal histogram entropy difference. Optical Engineering 36 (7), 1976–1981.

Sahoo, P.K., Soltani, S., Wong, A.K.C., Chen, Y.C., 1998. A survey of thresholding techniques. Computer Vision, Graphics and Image Processing 41, pp. 233–260.

Wong, A.K.C., Sahoo, P.K., 1989. A gray-level threshold selection method based on maximum entropy principle. IEEE Transactions on Systems, Man and Cybernetics SMC-19, 866–871.