IMAGE SEGMENTATION WITH SCALABLE SPATIAL INFORMATION

C.K. LEUNG[†] and F.K. LAM[‡]

[†] Department of Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong.

[‡] Department of Electrical & Electronic Engineering, The University of Hong Kong, Hong Kong.

ABSTRACT

In this paper, a general approach is proposed for the design of image segmentation algorithms utilizing spatial information which is the combined properties of a collection of neighborhood pixels. With different types of properties and different number of neighborhood pixels being utilized, segmentation algorithms with different speed and accuracy performance can be designed. Six algorithms have been implemented with their performance investigated and compared.

1. INTRODUCTION

Image segmentation is the process of classifying image pixels, which is an important step towards higher level image operations such as pattern recognition and object identification [1]. A common high-speed approach to image segmentation is thresholding, where image pixels having a gray-level value larger than a threshold value are classified into one class; otherwise into another [2]. Segmentation performance may be improved if spatial information, such as the combined properties of a collection of some neighborhood pixels, is taken into account. To incorporate spatial information in image segmentation, an image is usually modeled as a Markov random field (MRF) such that the class a pixel belongs to depends on the properties of its neighborhood pixels only [3]. Based on this model, a pixel is segmented into a class if the *a posteriori* probability that it belongs to this class under the condition that the properties of its neighborhood pixels have been observed is maximum [4]. In general, the speed and accuracy of such segmentation algorithms will depend on the number of neighborhood pixels and the type of their properties. In this paper, a general approach to the design of segmentation algorithms utilizing different number of

neighborhood pixels and different types of neighborhood pixels properties is proposed.

2. THEORY

2.1. Notations

A digital image is a collection of spatially ordered picture elements (pixels). To describe the properties of a general pixel X, two parameters are defined. The *scene class* parameter x is defined to denote which scene class pixel X belongs to. The fact that X belongs to scene class j is denoted by "x=j". The *true scene* is the collection of all x parameters for the entire image. The gray-level value g is defined to denote the intensity, or shade of gray, of pixel X assuming that the image is monochrome. The *gray-scale image* is the collection of all g parameters for the entire image, and is an observable property. Without loss of generality, it is assumed that there are J scene classes. Hence for each pixel, its scene class parameter $x \in \{0, 1, \cdots, J-1\}$. We also assume that a digital image is to be processed and the gray-level value g is digitized from 0 to L-1 into L unit steps, i.e. $g \in \{0, 1, \cdots, L-1\}$.

2.2. MAP segmentation with spatial information

Consider the segmentation of a general pixel X. It is assumed that some properties about its neighborhood pixels have been observed, and this is known as the *spatial information* S about X. If the image is modelled as a Markov random field, the *a posteriori* probability that X belongs to scene class j will depend on S, which can be written as P(x=j | S). Applying Bayes' theorem,

$$P(x=j \mid S) = \frac{P(S \mid x=j) \times P(x=j)}{P(S)} \quad j=0,1,\cdots,J-1 \quad (1)$$

Using the maximum *a posteriori* probability (MAP) principle, pixel X is to be segmented into class j* where

P(X=j*| S) is maximum among the J probabilities as described in eq. (1). To determine the maximum value among the J probabilities, both P(x=j) and P(S | x=j) need to be evaluated, but there is no need to evaluate P(S). If the true scene is known, the term P(x=j) can be evaluated as the ratio of the size of scene class j to the image size, and the term P(S | x=j) can also be evaluated as the ratio of number of simultaneous occurrences of S and x=j to the number of pixels belonging to scene class j. In practical situations where the true scene is unknown, a possible way to estimate P(x=j) and P(S | x=j) is to approximate the true scene by a segmented image. With these probabilities evaluated, the MAP process can be performed to obtain a new segmented image, which is taken as the approximate true scene for another run of the MAP process. In this way, an iterative segmentation algorithm is formulated [5]. To initialize the iterative procedure, Otsu's thresholding method [6] is employed. The segmented image obtained in a certain iteration cycle is compared to that obtained in the previous cycle. If they are identical, the iterative procedure is terminated. To prevent endless iteration, the iterative procedure will nevertheless be terminated after 50 iterations. To identify the best segmented image from those produced in the course of iteration, the maximum segmented scene entropy (MSSE) criteria [7] is employed.

2.3. Scalable spatial information

We formulate a scalable spatial information scheme by incorporating different number of neighborhood pixels and different types of spatial information. We shall incorporate, respectively, 1, 2, 3, and 4 neighborhood pixels, and three different types of spatial information to design six schemes of spatial information S. These schemes are, namely, S_1=(g, g1), S_2=(g, g1, g2), S_3=(g, g1, g2, g3), S_4=(g, g1, g2, g3, g4), S_a=(g, ga), and S_x=(g, x1, x2, x3, x4). As illustrated in Fig.1, g1, g2, g3, and g4 are the gray-level values of the neighborhood pixels; x1, x2, x3 and x4 are their scene class parameters, ga is the average of g1, g2, g3, g4, i.e. ga=¼(g1+g2+g3+g4), which is also known as the local average [8]. Based on these six spatial information schemes, six iterative MAP segmentation algorithms have been implemented. Since their implementations are quite similar, the algorithms using scheme S_4 and S_x are discussed in following as

the general illustration. With spatial information S_4, P(S | x=j) can be written as P(g, g1, g2, g3, g4 | x=j). This term is hard to estimate confidently due to its huge dimensionality. However, if it is assumed that the gray-level value distribution of a pixel depends on its scene class only [4], it is relatively simple to evaluate P(g, g1, g2, g3, g4 | x=j) as:

$$P(g, g1, g2, g3, g4 \mid x=j) =$$

$$\sum_{a=0}^{J-1} \sum_{b=0}^{J-1} \sum_{c=0}^{J-1} \sum_{d=0}^{J-1} (P(x1=a, x2=b, x3=c, x4=d \mid x=j) \times$$

$$f_j(g) f_{x1}(g1) f_{x2}(g2) f_{x3}(g3) f_{x4}(g4)) \qquad (2)$$

As for the case of S_x, the spatial information is (g, x1, x2, x3, x4), the conditional probability P(g, x1, x2, x3, x4 | x=j) may be evaluated as:

$$P(g, x1, x2, x3, x4 \mid x=j) =$$

$$P(x1, x2, x3, x4 \mid x=j) \times f_j(g) \qquad (3)$$

In eqs. (2) and (3), $f_j(\cdot)$ is the probability density function (pdf) of the gray-level values of those pixels belonging to scene class j. These pdfs are assumed to be Gaussian and their means and variances are estimated by making reference to the gray-scale image and the approximated true scene. With all these probability terms evaluated, the MAP process can be carried out.

3. IMPLEMENTATION

The general procedure of each algorithm can be described formally as follows:

Step 1: Threshold the image by Otsu's method; record the segmented image as `OptImage`

Step 2: Take the segmented image generated in the previous cycle as the true scene; perform MAP segmentation as described in Section 2.3; calculate the SSE of the segmented image, if it is greater than the SSE of the `OptImage`, record this segmented image as `OptImage`;

Step 3: If the present segmented image is identical to the previous one, or the number of iterations exceeds 50, go to step 4; otherwise go to step 2;

Step 4: `OptImage` is the optimum segmented image.

For performance comparison, a segmentation algorithm by the Iterative Conditional Mode [3] has also been implemented. This algorithm is initialized by applying Otsu's thresholding method, and the optimum segmented image is identified by applying the MSSE criteria.

4. RESULTS AND DISCUSSIONS

To study and compare the performance of these algorithms in a controlled manner, they are employed to segment different synthetic images. In each synthetic image, there are two scene classes, namely the object and the background. The gray-level values of the object pixels and background pixels are modeled by a Gaussian process, each with a certain mean and standard deviation. As an example, consider the segmentation results for eight synthetic images. The true scene of these images is shown in Fig. 2. The relative size of the object is 0.2. The standard deviation of the gray-level values of both the object pixels and the background pixels are 18. The mean gray-level value of the background pixels is 180. The mean gray-level value of the object pixels ranges from 100 to 170 in steps of 10. In this way eight images have been synthesized. Each synthetic image is segmented by each of the seven algorithms and the segmentation error percentage recorded, which is plotted against the object gray-level value for each algorithm in Fig. 3. From these results, it is seen that the more neighborhood pixels to be considered, the better is the segmentation result. But it is also noted that the segmentation process will be slower if more neighborhood pixels are considered since the amounts of computation become greater. Based on these results, the ICM algorithm seems to have the best compromise between speed and performance. The second best is the algorithm employing S_x as the spatial information. The runners-up are the algorithms employing S_a and S_4 as the spatial information. These two algorithms have demonstrated comparable performance. But since the S_a algorithm is faster than the S_4 algorithm, it seems that S_a is a better compromise in speed and performance than S_4. The remaining three algorithms employing S_3, S_2, and S_1 have achieved progressively worse performance, but with progressively shorter time of processing, as is to be expected.

5. CONCLUSIONS

A general approach to the design of iterative image segmentation algorithms utilizing scalable spatial information has been proposed. It is shown that by varying the number of neighborhood pixels and the type of neighborhood information, segmentation algorithms with different speed and accuracy performance can be designed.

REFERENCES

[1] A. Rosenfeld and A. Kak, Digital Picture Processing, 2nd ed., (Academic Press, NY, 1982)

[2] P.K. Sahoo, S. Soltani, A.K. C. Wong and Y.C. Chen, "A survey of thresholding techniques," Comp. Vis. Graph. Image Proce., 1988, Vol. 41, pp. 233-260.

[3] J. BESAG, "On the statistical analysis of dirty pictures", J. R. Stat. Soc., series B, 1986, Vol. 48, pp.259-302.

[4] C.K. Leung and F.K. Lam, "Maximum a posteriori spatial probability segmentation," Proc. IEE Vis., Image and Sig. Proc., 1997, Vol. 144, pp. 161-167.

[5] LEUNG, C.K., and LAM, F.K.: 'An iterative image segmentation algorithm utilizing spatial information', Proceedings of IEEE TENCON 1996, Perth, 1996, vol. 1, pp.141-146.

[6] OTSU, N.:'A threshold selection method from grey-level histograms', IEEE Trans., 1979, Vol. SMC-9, (1), pp. 62-66

[7] LEUNG, C.K., and LAM, F.K.: 'Maximum segmented-scene spatial entropy thresholding', Proceedings of IEEE international conference on Image processing, Switzerland, 1996, pp. 963-966.

[8] R.L. Kirby and A. Rossenfeld, "A note on the use of (gray level, local average gray level) space as an aid in thresholding selection", IEEE Trans. Syst. Man Cybern. 1979, Vol. SMC-9, pp. 860-864.
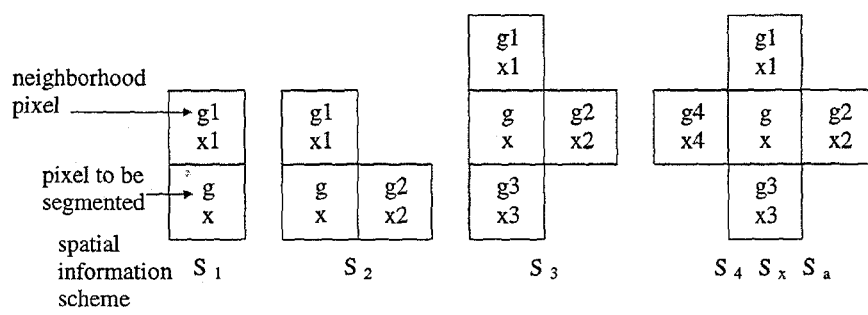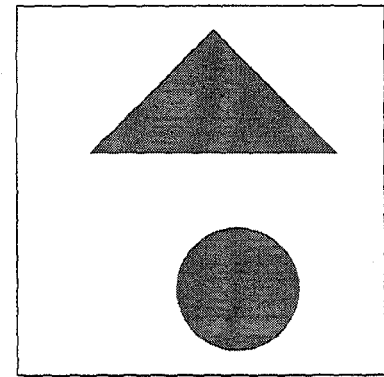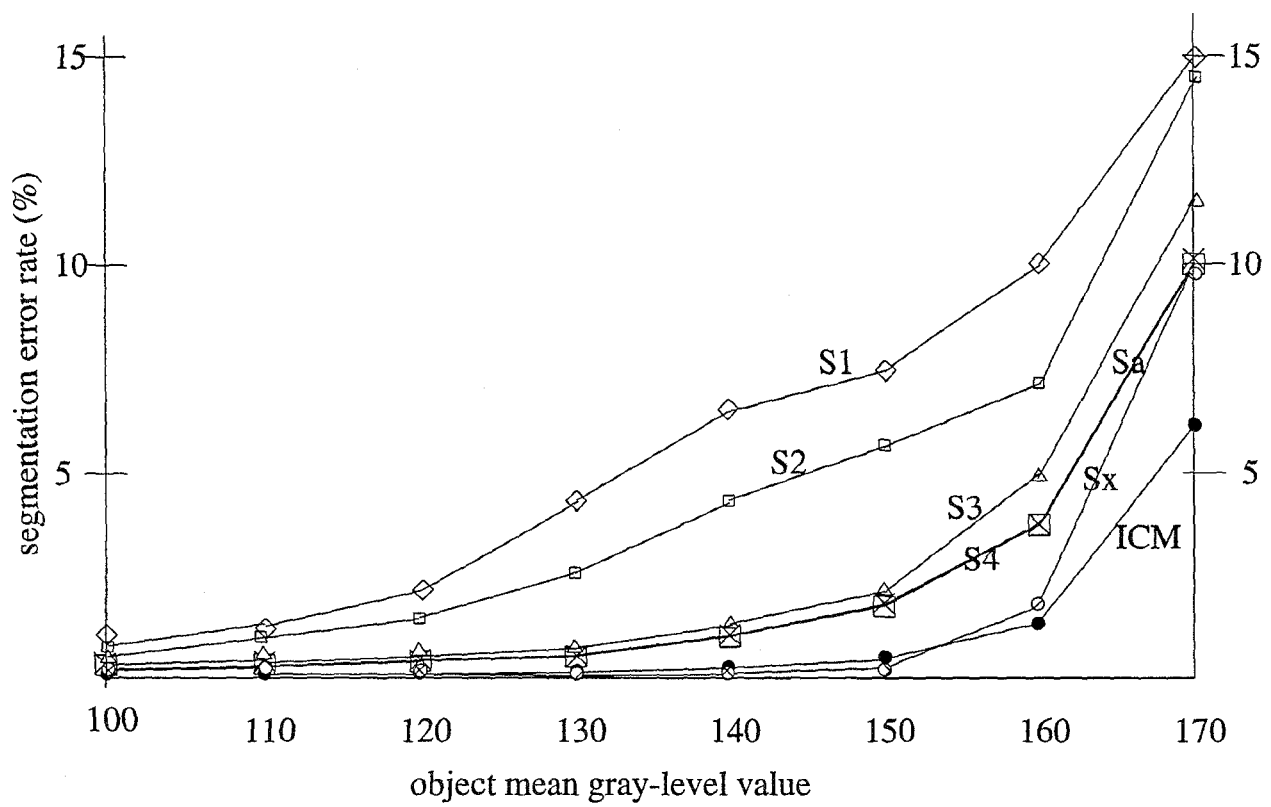
Fig.1 Spatial information defintion

Fig.2 True scene

Fig.3 Segmentation results comparison