

Supervised Segmentation Using a Multiresolution Data Representation

Isaac Ng*, J. Kittler and J. Illingworth
Department of Electronic and Electrical Engineering
University of Surrey
Guildford GU2 5XH
United Kingdom

Abstract

We present a supervised segmentation scheme in which a Bayesian approach incorporating a pyramid data structure is used. This formulation leads to a significant simplification of the Spann and Wilson quadtree segmentation algorithm [7] under the assumption that image classes are normally distributed. A method for efficiently acquiring the parameters of class distributions at each resolution level has been developed. It involves estimating the class statistics on training sites at full image resolution. The corresponding parameters at lower resolutions are computed by predetermined scaling factors. The segmentation scheme is validated on synthetic data and natural textures obtained from the Brodatz album [1].

1 Introduction

Segmenting an image into spatially disjoint regions of uniform property has been the subject of extensive research over the past decades. The prime aim is to provide a symbolic description of the constituent parts of the image for scene interpretation. The many different approaches to this topic can be classified into distinct categories on the basis of either mathematical framework or image phenomenon used for segmentation. Among them, statistical approaches such as per-pixel classification have been widely applied [5].

However criticism has been raised against the approach of classifying pixels based solely upon the statistical distribution of individual pixel intensity or pixel features, on the ground that the spatial contextual information conveyed in an image is not taken into consideration. Consequently, any spatial coherence and localization of the regions which result from the classification is entirely fortuitous. This may be the reason why these procedures tend to produce an abundance of false regions.

Hence incorporating contextual or spatial information for improving classification reliability has been a major aim of recent research into this problem. A number of methodologies have emerged of which multiresolution processing, in which processing is carried out over a range of spatial scales of the input image, has proved very promising in terms of computational efficiency [6, 7].

*Supported by a scholarship from the Croucher Foundation

We propose a new supervised segmentation scheme in this paper which is closely related to the Spann and Wilson quadtree segmentation algorithm [7] in terms of processing steps involved. For instance, both schemes involve three processing stages such as generating a multiresolution representation of an image, applying statistical classification to coarse resolution images and refining the segmentation result in a 'coarse-to-fine' fashion in order to obtain a full spatial resolution segmentation.

The key difference between the two algorithms lies in the way the segmentation problem is formulated. For instance, a non-supervised clustering technique is used in the Spann and Wilson algorithm for the statistical classification process which makes the algorithm ideal for general segmentation problems. However, in many image analysis applications requiring routine processing, each image region represents one of a finite set of imaged phenomena whose statistical properties are approximately invariant over a large set of images. In such situations the general segmentation scheme of Spann and Wilson is not only computationally unnecessarily complicated but also may, as reported in [4], fail to yield consistent segmentation owing to the unpredictable effect of some of the parameters of their method which are automatically selected in a data dependent manner.

Our objective in the present paper is to demonstrate that a supervised formulation leads to a significant simplification of the Spann and Wilson algorithm under the assumption that image classes are normally distributed and this gives more consistent segmentation results. The paper is organized as follows: we begin by describing the image pyramid formulation in Section 2. In Section 3 we discuss the potential problems in estimating the class statistics for reduced spatial resolution image which are essential for implementing Bayes classifier at each resolution level of the pyramid. Then we present a method for efficiently acquiring the parameters of class distributions at each resolution level. Section 4 details the segmentation scheme. Section 5 shows some results on synthetic data and natural textures obtained from the Brodatz album [1]. Finally we conclude with several remarks about the proposed approach.

2 Image pyramid

An image pyramid [6] is a hierarchical data structure in which successively reduced spatial resolution versions of a given image $I(x, y)$ are stacked to form a pyramid as shown in figure 1. The value held by an entity can be an integer or a vector. The progressively reduced spatial resolution images are generated recursively according to the following equation

$$I_l(x, y) = \sum_{i,j=-1}^2 M(i, j) I_{l-1}(2x + i, 2y + j) \quad (1)$$

where $M(i, j)$ is a 4×4 smoothing kernel or *the generating kernel* according to Burt [2]. $I_l(x, y)$ denotes the image value at coordinates (x, y) and pyramid level l . It is referred to by the term *entity* or *node* in this paper.

The 2-D kernel, $M(i, j)$, is a product of a 1-D generating function $M(\cdot)$ and its transpose, so that $M(i, j) = M(i) \cdot M(j)$ where $M(\cdot)$ is defined subject to the following constraints: *Normalization*: $\sum_{i=-1}^2 M(i) = 1$, *Symmetry*: $M(-1) = M(2) = b$ and $M(0) = M(1) = a$, and *Unimodality*: $a \geq b \geq 0$.

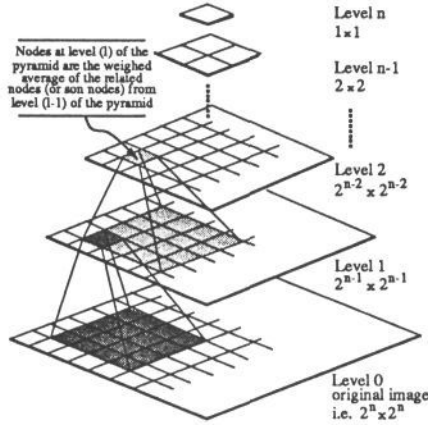


Figure 1: Pyramidal image structure

The corresponding constraints extend to the 2-D kernel, $M(i, j)$. Equation (1) can also be written as:

$$I_l(x, y) = \sum_{i, j=r_l^-}^{r_l^+} E_l(i, j) I_0(2^l x + i, 2^l y + j) \quad (2)$$

where $E_l(i, j)$ is a kernel which will be referred to as *equivalent smoothing kernels* hereafter and where r_l^- and r_l^+ denote the spatial extent of the equivalent kernel. It means that the entity value, i.e. $I_l(x, y)$, can be determined directly from the original resolution image $I_0(\cdot, \cdot)$ by convolving with an appropriate kernel $E_l(i, j)$, instead of generating all the successive levels.

$E_l(i, j)$, like $M(\cdot, \cdot)$, is Cartesian separable, so that $E_l(i, j) = E_l(i) \cdot E_l(j)$ where $E_l(\cdot)$ is related to $M(\cdot)$ by a recursive expression, i.e.

$$E_l(x) = \sum_{i=-1}^2 M(i) E_{l-1}(x - 2^{l-1}i), \quad r_l^- \leq x \leq r_l^+; \quad (3)$$

with the initial conditions stated as $E_0(x) = \begin{cases} 1; & x = 0 \\ 0; & \text{otherwise} \end{cases}$. Note that $E_l(x) = 0$ for x outside the kernel domain $[r_l^+, r_l^-]$. The indices, i and j , lie between r_l^+ and r_l^- , where $r_l^+ = r_{l-1}^+ + 2^l$, $r_l^- = r_{l-1}^- - 2^{l-1}$ with $r_1^+ = 2$ and $r_1^- = -1$. It should be noted that the equivalent smoothing kernel at level 1 is equal to the generating kernel, i.e. $E_1(\cdot) = M(\cdot)$. These one dimensional and two dimensional equivalent kernels inherit all the constraints applied to the original generating kernel.

3 Estimation of the class statistics

To implement the Bayes decision rule with a pyramidal data structure, one must know the class statistics at each level of the pyramid. However, direct

estimate from the data at low resolution level of a pyramid is likely to give biased estimates as we explain later in this section. Instead we introduce here a method which can estimate the class statistics at each reduced resolution level of the pyramid without recourse to the data at that level.

Let us assume that the entities of the full resolution image function $I(x, y)$, or $I_0(x, y)$, constitute a set of random vectors $\{\mathbf{X}\}$ which are statistically independent and identically distributed (iid) according to $N(\bar{\mu}, \Sigma)$ or $N({}_0\bar{\mu}, {}_0\Sigma)$. This implies that, for the moment, we consider a homogeneous image. Since $I_l(x, y)$ is obtained by a linear combination of normally distributed random vectors, $I_0(\cdot)$, as expressed in equation 2, then $I_l(x, y)$ itself is normally distributed as $N({}_l\bar{\mu}, {}_l\Sigma)$ with mean vector ${}_l\bar{\mu}$ and covariance matrix ${}_l\Sigma$. Now the problem is how to obtain ${}_l\bar{\mu}$ and ${}_l\Sigma$

3.1 Potential Problems

In order to obtain an unbiased estimate of a class statistics, we require that the data be statistically independent and identically distributed. However as a result of the pyramid construction process the nodes at higher levels of the pyramid are no longer independent. In figure 2, each node at higher level of the pyramid shares half the population of its son nodes one level below with its neighbouring nodes. Hence the statistical independence assumption holds only when the nodes are separated by a certain distance. The distance required monotonically increases as we move further up the pyramid. As a result, we cannot determine the class statistics from the data at each level since the estimates would be highly biased, but also we cannot estimate them from the subsampled independent data because the sample size is not large enough to provide sufficient statistical confidence for the results.

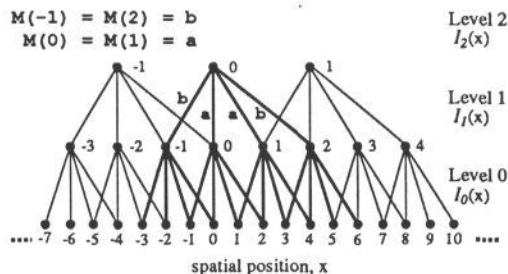


Figure 2: Graphical illustration of the procedure to determine the node values at various levels of the pyramid.

So far we assumed that the entities of the image function $I(x, y)$ are drawn from a single normal distribution. However the image function is usually composed of several segments each filled with samples drawn from a distinct distribution. This gives rise to yet another problem when the image pyramid is constructed, namely the segments at higher levels of the pyramid will contain mixed population pixels. These originate from the kernel smoothing across the boundary of two adjacent segments. The relative proportion of these mixed pixels rapidly becomes significant as one moves up the pyramid and can seriously affect (bias) the estimated class statistics.

3.2 Parameter inference

The idea of predicting class statistics for reduced resolution image from training data at full image resolution originates from the equivalence relation as expressed in equation 2. Here a direct relation between the entity value at any level to the original resolution image has been established. The consequence of this relation between coarse and high resolution image entities is that the class statistics at each resolution level can be expressed in terms of the class statistics estimated from the full resolution training sites, i.e. $N(0\vec{\mu}, 0\Sigma)$ by the expression:

$${}_l\vec{\mu} = {}_0\vec{\mu}; {}_l\Sigma = f_l {}_0\Sigma \quad \text{where } f_l = \begin{cases} \sum_{i,j=r_l^-}^{r_l^+} E_l(i,j)^2 & l \neq 0 \\ 1 & l = 0. \end{cases} \quad (4)$$

Since we can obtain unbiased estimate of ${}_0\vec{\mu}$ and ${}_0\Sigma$ with reasonable confidence, so can we compute the class statistics at each resolution level by applying equation 4.

4 Segmentation Scheme

The proposed supervised segmentation scheme involves three distinct stages of processing as shown in figure 3. The first stage called **image pyramid construction** has been described in Section 2. In the following, the second and third stage of the segmentation scheme are briefly described.

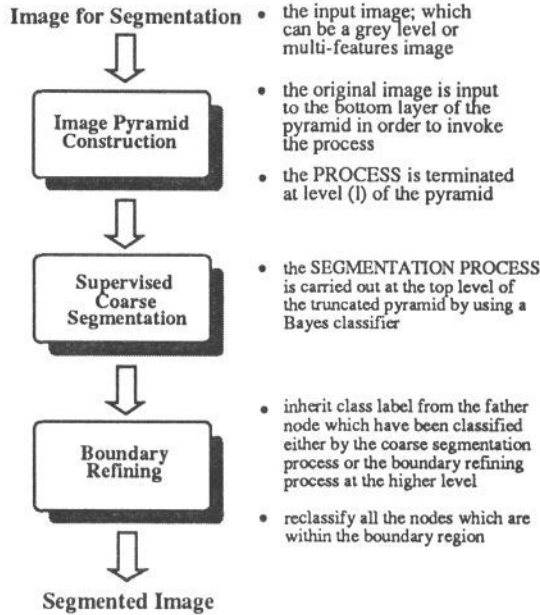


Figure 3: Schematic diagram illustrating the proposed Supervised Segmentation Scheme.

Stage 2: Supervised Coarse Segmentation

The supervised coarse segmentation procedure involves the following steps:

- Step 1. Estimate the class statistics ${}_0\bar{\mu}_i, {}_0\Sigma_i$ for each class from the full spatial resolution training data where 0 and i stand for the pyramid level and class identity respectively
- Step 2. Build the pyramid for image $I(x, y)$ to be classified
- Step 3. Choose the level at which the classification is to be carried out, i.e. at level l
- Step 4. Generate the equivalent convolution mask $E_l(i, j)$ by specifying parameter a and then evaluate the corresponding scaling factor f_l
- Step 5. Classify the pixels of the reduced resolution image $I_l(x, y)$ by evaluating the discriminant functions $g_i^l(\cdot)$ given as

$$g_i^l(I_l(x, y)) = \log_e P(\omega_i) - \frac{1}{2} \log_e [(2\pi)^D |{}_0\Sigma_i|] - \frac{D}{2} \log_e [f_l] - \frac{1}{2f_l} [I_l(x, y) - {}_0\bar{\mu}_i]^T \Sigma_i^{-1} [I_l(x, y) - {}_0\bar{\mu}_i]. \quad (5)$$

where D is the dimension of the feature space and $P(\omega_i)$ is the *a priori* probability of class ω_i .

Stage 3: Boundary refining

The boundary refining process is designed to restore the actual spatial location of the region borders and can be outlined in the following general form:

- Step 1. At k th level of the pyramid:
 - if nodes do not lie at the region boundaries
 - then give them the same class label as their fathers
 - else label as boundary nodes
- Step 2. Nodes in the boundary region are classified by the Bayes classifier with the corresponding class statistics at level k
- Step 3. Terminate the process if full resolution has been restored; otherwise proceed downward to $(k - 1)$ th level and go to Step 1

5 Experimental Results

Two experiments are presented to demonstrate the proposed segmentation scheme. The first uses synthetic data but the second shows segmentation results using textured images composed of Brodatz's texture regions.

Experiment 1

Consider figure 4. This 128×128 image is composed of three squares of different size. The entities of these regions are defined by bivariate data vectors which are drawn independently from three normal populations specified in the figure. Figure 5 shows the scatter plot of the pixel vectors in the 2-D feature space. It can be seen from the figure that the class density functions are overlapped quite severely. Figure 6a shows the result of segmenting the bifeature image at full resolution directly using conventional pixel-based Bayes classifier[5]. As expected, the segmented regions are extremely 'spotty'.

In comparison, we show in figure 6b the result of segmenting the above bifeature image using the proposed algorithm. The pyramid representation of

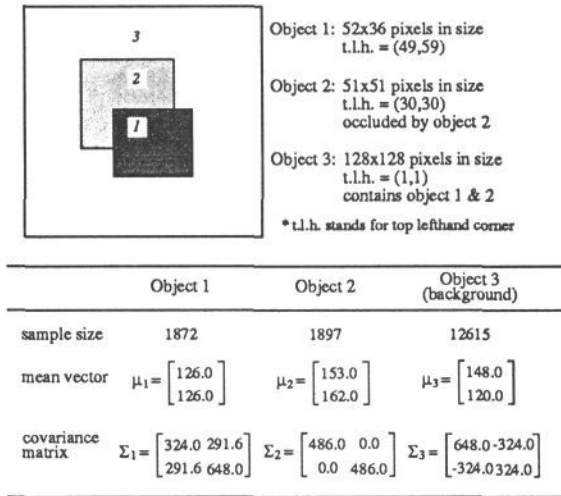


Figure 4: The ground truth of the multiple classes bivariate feature image and the statistical parameters.

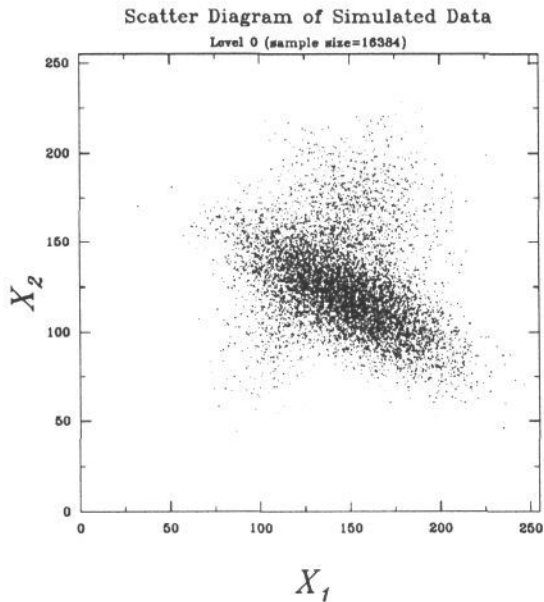


Figure 5: Scatter plot of the data in a two dimensional feature space

this bifeature image is constructed using the convolution mask $M(i, j)$ with $a = 0.37$. The coarse segmentation is carried out at level 3 of the pyramid where the spatial resolution is equal to 16×16 pixels in size. As can be seen, the algorithm has produced a segmentation in good agreement with the given ground truth.

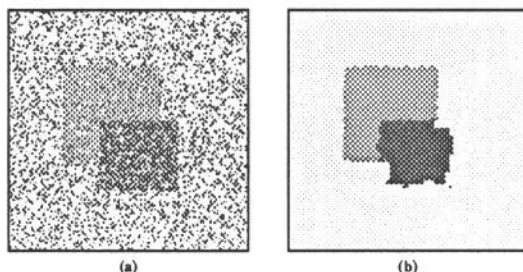


Figure 6: Segmentation results. (a) Segmentation result obtained by using pixel based classification method applied to the full resolution image. (b) Segmentation result obtained using the proposed algorithm.

Experiment 2

Figure 7a–b show a test image which is 128×128 pixels in size and is composed of two Brodatz's textures. All textures are independently histogram equalized into 256 grey levels. Texture features are then extracted by a set of Gabor filters as described in [3]. The segmentation, for instance, is operated in a multidimensional feature space rather than the original image domain. Figures 7c–d show the segmentation results. It can be seen that the algorithm has produced a segmentation in good agreement with the perceived boundaries.

6 Conclusion

A supervised segmentation scheme has been described in which a Bayesian approach incorporating a pyramid data structure is used. The favourable features of this approach are that the supervised Bayes classifier is known to be a very effective tool for classification problems and that the pyramidal data representation can relieve the computational burden involved in the classification process. Furthermore, spatial information can be incorporated into the classification process to overcome the weaknesses of purely statistical methods.

The key question in incorporating a Bayes decision rule with pyramidal data structure is that the class statistics at each level of the pyramid must be available. However, we have pointed out that unbiased estimates cannot be achieved by direct calculation. Instead, a method for efficiently acquiring the parameters of class distributions at each resolution level has been proposed. Experimental results have been presented which demonstrate the power of the method. However, this performance can only be expected when class statistics are normally distributed and image segments are of commensurate size to the level of noise corrupting the image.

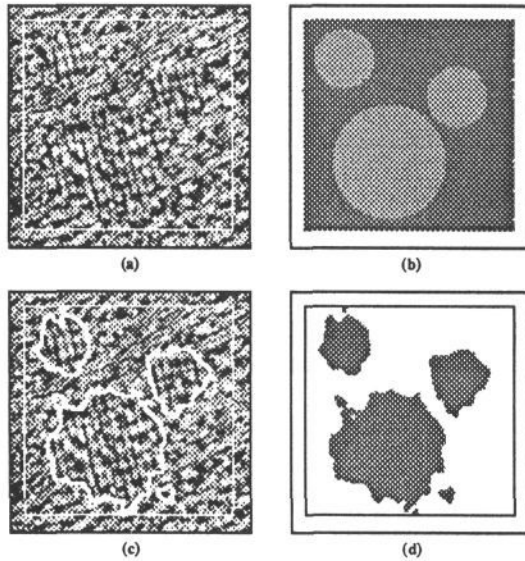


Figure 7: Texture mosaic. (a) and (b) show a texture mosaic and its grey level coded ground truth respectively. (c) Segmentation result. (d) Original image superimposed with the estimated region boundary

References

- [1] Brodatz, P., "*Textures - A Photographic Album for Artists and Designers*," New York: Dover, 1966.
- [2] Burt, P. J., "Fast filter transforms for image processing," *Computer Graphics and Image Process.*, 1981, 16, pp. 20-51.
- [3] Clark, M. R., Bovik, A. C. and Geisler, W. S., "Texture segmentation using a class of narrowband filters," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, 1987, pp. 571-574.
- [4] Daida, J., Samadani, R. and Vesecky, J. F., "Object-Oriented Feature-Tracking Algorithms For SAR Images of the Marginal Ice Zone," *IEEE Trans. Geosci. and Remote Sensing*, 1990, 28, pp. 573-589.
- [5] Rosenfeld, A. and Kak, A. C., "Digital Picture Processing," New York, Academic Press, 1982.
- [6] Rosenfeld, A (Ed.), *Multiresolution image processing and analysis*, Springer-Verlag, Berlin, FRG, 1984.
- [7] Wilson, R., and Spann, M., *Image Segmentation and Uncertainty*, Research Studies Press, 1988.