Multiresolution Image Segmentation Combining
Region and Boundary Information

A. Bhala Rao, R. Wilson
Department of Computer Science,
University of Warwick,
Coventry CV4 7AL, U.K.

Abstract

A new approach to image segmentation is presented that integrates region and boundary information within a general framework of maximum a posteriori (MAP) estimation and decision theory. The algorithm employs iterative, decision-directed estimation performed on a spatially localised basis but within a multiresolution representation. The use of a multiresolution technique ensures both robustness in noise and efficiency of computation, while the model-based estimation and decision process is both flexible and spatially local, thus avoiding assumptions about global homogeneity or size and number of regions. The method gives accurate segmentations at low signal-to-noise ratios and is shown to be more effective than previous methods in capturing complex region shapes.

I. Introduction

Over recent years there has been a growing interest in the use of multiresolution, or scale space, methods in the segmentation of images. Such methods can broadly be classed as either region-based or edge-based. Among the former, notable examples were the ‘split-and-merge’ algorithm of Chen and Pavlidis [1], and the linked pyramid methods of Burt et al [2]. More recently, Spagnuolo and Wilson described a quad-tree based algorithm employing classification at a low spatial resolution combined with downward directed boundary estimation, for the segmentation of images containing arbitrary numbers of regions of homogeneous grey level or texture [3]. An attempt to place such methods in the more precise statistical framework of Markov Random Fields (MRF) was reported recently by Gidas [4]. Because of the inherent limitations of region-based methods, a number of authors have preferred edge-based techniques, for example the ‘edge-focusing’ technique of Bergholm [5], as well as methods based on spatial frequency analysis, such as that of Calway and Wilson [6].

While either methodology can prove satisfactory under certain conditions, it would clearly be beneficial to combine region and edge information in a single structure to make maximum use of all the available information relevant to a given segmentation. Such approaches have recently been described in [7]. It is the purpose of this paper to describe a new method of integrating region and boundary information within a general framework of maximum a posteriori (MAP) estimation and decision theory. The algorithm employs iterative, decision-directed estimation performed on a spatially localised basis, but within a multiresolution framework. The use of a multiresolution technique ensures both robustness in noise and efficiency of computation, while the model-based estimation and decision process is both flexible and spatially local, thus avoiding assumptions about global homogeneity or size and number of regions which characterise some of the earlier algorithms. A description of the algorithm is followed by some results on the segmentation of objects with complex shapes against a high level of noise (0 dB SNR) and on a natural scene. The results show that the new technique promises to be more effective than previous methods in capturing complex region shapes.

II. The Algorithm

The overall structure of the algorithm is depicted in figure 1. It has four main components: first a lowpass pyramid is constructed which forms the input to the two autonomous region and boundary
processes. The fourth component is an interaction between the two processes which aims to exploit the complementary nature of regions and their boundaries.

The overall computational paradigm is essentially that of producing a ‘least-cost’ or minimum mean square error (MMSE) fit to the data, within the constraint of a model of region and boundary properties. In this respect, it shares some features with the Bayesian methods described in [7]. However, at this stage it has not proved necessary to resort to stochastic relaxation, as deterministic procedures have produced satisfactory results. The signal and noise parameters are estimated using local processing within the pyramid structure for both the region and boundary processes.

A. Low-Pass Pyramid

The first step is a fast smoothing operation performed using a lowpass pyramid [8] which trades off noise reduction against spatial resolution. Given an image \( x_{ij}, 0 \leq i, j < N = 2^M \), and a lowpass kernel \( A_{mn}, -K \leq m, n < K \), the general form of the processing is

\[
x_{ij}(l) = \sum_{m=-K}^{K} \sum_{n=-K}^{K} A_{mn} x_{(2i-m)(2j-n)}(l-1)
\]

where \( 0 \leq l \leq M \) and \( x_{ij}(0) = x_{ij} \). The kernel used in this work was selected by a least squares approach to minimise the effects of errors in orientation estimation, which forms a significant part of the overall algorithm [9].

B. Region Estimation Process

Inhomogeneous Tessellation

Pyramid nodes are selected based on parent-child variances \( v_{ij}^2(l), l > 1 \)

\[
v_{ij}^2(l) = \sum_{m} \sum_{n} A_{mn} \left[ x_{(i-m)(j-n)}(l) - \bar{x}_{(i-m)(j-n)}(l) \right]^2
\]

\[
\bar{x}_{ij}(l) = \sum_{m} \sum_{n} B_{mn} x_{(i/2-m)(j/2-n)}(l+1)
\]

where \( B_{mn}(l) \) is an appropriate interpolation filter. The sample variances \( v_{ij}^2(l) \) are then compared with the average variance from the level below \( \bar{v}^2(l-1) \). If \( v_{ij}^2(l) > \alpha \bar{v}^2(l-1) \), then the associated node \( (i,j,l) \) is marked. This process is performed bottom-up. When it is complete, the lowest marked nodes are selected to represent a square block \( (2^l \times 2^l) \) of the image (figure 5(b)). The assumption is that \( \bar{v}^2(l-1) \) is a good estimator of the image ‘within region’ variance on level \( l \) of the pyramid so that any variation which is significantly greater than this will be due to the node \( (i,j,l) \) representing more than one region. The factor \( \alpha \) is used to control the extent of the variability allowed.
Region Adjacency Graph

A graph \( \mathcal{R} = \{\xi_0, \xi_1, \ldots, \xi_n\} \) of \( n \) selected nodes \( \xi = (i, j, l)^T \) is created. A first order neighbourhood \( \mathcal{N}_k \) for a node \( \xi_k \in \mathcal{R} \) can be defined as being all nodes that represent image blocks bordering the image block represented by node \( \xi_k \) (figure 2). This structure is a modification of the conventional region adjacency graphs used in segmentation [10] which takes advantage of the scale invariance and computational properties of the quad-tree tessellation.

Iterated Decision-directed Estimation

This part of the processing is based on a normal model for the data, in which regions are assumed to have independently selected means. Each node \( \xi_i \in \mathcal{R} \) is iteratively averaged with its neighbours in the neighbourhood \( \mathcal{N}_i \):

\[
x_i^t = \frac{w_i x_i^{t-1} + \sum_{j \in \mathcal{N}_i} w_j l_{ij} x_j^{t-1}}{w_i + \sum_{j \in \mathcal{N}_i} w_j l_{ij}}
\]

(4)

where \( t \) is the iteration number and \( w_i \) is a weight function equal to the area represented by the node \( w_i = 2^d \), which corresponds to the inverse of the noise variance at that node. The switching function \( l_{ij} \) is determined by the MAP test that a given link is ‘on’, given the data at the two nodes:

\[
l_{ij} = \begin{cases} 1 & \text{if } P_{ij}(1)p(\hat{x}_i - \hat{x}_j \mid 1) > P_{ij}(0)p(\hat{x}_i - \hat{x}_j \mid 0) \\ 0 & \text{otherwise} \end{cases}
\]

(5)

where \( p(\hat{x}_i - \hat{x}_j \mid 1) \) is the conditional probability density of the difference between the grey level at two neighbouring nodes assuming that nodes \( \xi_i \) and \( \xi_j \) are connected and therefore have the same mean. The conditional densities under the two hypotheses are

\[
p(\hat{x}_i - \hat{x}_j \mid 1) \propto \exp \left( \frac{-(x_i^0 - x_j^0)^2}{2(\sigma_i^2 + \sigma_j^2)} \right)
\]

\[
p(\hat{x}_i - \hat{x}_j \mid 0) \propto \exp \left( \frac{-(x_i^0 - x_j^0 - (x_i^t - x_j^t))^2}{2(\sigma_i^2 + \sigma_j^2)} \right)
\]

(6)

where \( x_i^t \) is the estimated grey level at the node \( i \) on iteration \( t \). It is assumed, therefore, that if the nodes \( i \) and \( j \) belong to the same region, then their difference is noise, but if they belong to separate regions these have independently chosen means, whose difference is approximated by \( (x_i - x_j) \). \( P_{ij}(1) = 1 - P_{ij}(0) \) is the probability that the link between the nodes \( \xi_i \) and \( \xi_j \) is on.
TABLE I
3 × 3 EDGE KERNELS.

| \(-0.074\) | \(-0.095\) | \(0.000\) |
| \(-0.095\) | \(0.000\) | \(0.095\) |
| \(0.000\) | \(0.095\) | \(0.074\) |
| \(-0.074\) | \(-0.095\) | \(0.000\) |

These link probabilities are derived from modelling the neighbourhood as an edge. Given an estimate of the orientation \(\theta_i\), position \(r_i\) and probability of the edge \(P_e(\text{edge})\) in the neighbourhood of the link \(l_{ij}\), the probability that a given link is ‘off’ is calculated as a function of these parameters

\[
P_{ij}(0) = F(\theta_i, r_i, P_e(\text{edge}))
\]  

(7)

This introduces a dependence of the region link state on the data in the neighbourhood, through an estimate of the local orientation. The motivation behind the approach is that the parameters required to determine \(P_{ij}(0)\) are available either from some region based estimate of local orientation, or directly from boundary information. In the experiments presented below an estimate of the local edge orientation and position was obtained by least squares fitting to a simple piecewise linear model of an edge in the neighbourhood.

After each iteration, any block with a high certainty of an edge, \(P(\text{edge}) > 0.5\), becomes a candidate for splitting. The quadtree structure is maintained and each such node is split into 4 with the new nodes being linked into the existing region adjacency graph. The test for splitting compares the prospective children with the parent and data gathered from the current neighbourhood. If a block is split, the data for the children are taken from the parent neighbourhood, thus ‘focusing’ the region information around the edge.

C. Boundary Estimation Process

An orientation estimation procedure is performed on the lowpass pyramid of the image. There are two aspects to the process: the initial orientation estimation on each level of the pyramid, followed by a top-down propagation of estimates from higher levels to lower levels, to reduce estimation errors. The procedure is a variation of the MMSE estimator developed by Clippingdale [11].

The initial orientation estimate is obtained by the use of a pair of optimised (3 × 3) spatial kernels shown in table I [8]. The outputs from the spatial convolution, \(g_k(l), 0 \leq k \leq 1\), are combined to yield an estimate of the local orientation \(\hat{\theta}_{ij}(l)\), which can be expressed as the vector:

\[
\hat{\theta}_{ij}(l) = \begin{pmatrix}
2g_{0,ij}(l)g_{1,ij}(l)
g_{0,ij}^2(l) - g_{1,ij}^2(l)
\end{pmatrix}
\]  

(8)

The magnitude of this vector is a measure of the local edge energy and the argument twice the local orientation. The doubling of the the angle, originally proposed by Knutsson [12], is essential to make operations such as averaging of the vectors meaningful in terms of orientation.

Estimation of the orientation vector of a node on child level \(l\), based on a linear pyramid model of the data, is defined recursively as a linear combination of the data \(\hat{\theta}_{ij}(l)\) and the estimate of the parent \(\hat{\theta}_{ij}(l + 1)\) [11]:

\[
\hat{\theta}_{ij}(l) = \beta(l)\hat{\theta}_{ij}(l) + [1 - \beta(l)]\sum_m\sum_n B_{mn}\hat{\theta}_{ij/2,m}(l + 1)
\]  

(9)

\[
\hat{\theta}_{ij}(l) = \hat{\theta}_{ij}(l) + \eta_{ij}(l)
\]  

(10)
where the data $\hat{Q}_{ij}(l)$ come from orientation estimate on the pyramid of the noisy image, and $Q_{ij}(l)$ is the ‘signal vector’, $\phi_{ij}(l)$ the vector representing the errors in the initial orientation estimate and the coefficients $\beta(l)$ are a function of the signal and noise variances [11].

The propagated orientation estimate $\hat{\theta}_{ij}(l)$ is sequentially filtered anisotropically by a 2-d Gaussian kernel $h_{ij}(\theta)$ in a manner used by a number of authors (e.g., [13]). The iterated filtering is of the form

$$\hat{\theta}_{ij}^k(l) = \sum_m \sum_n h_{nm}(\hat{\theta}_{ij}^{k-1}(l-1)\hat{Q}_{(i-m)(j-n)}^k(l)$$ \hspace{1cm} (11)

where $h_{mn}(\theta)$ is an oriented smoothing kernel depending on the last estimated orientation

$$\hat{\theta}_{ij}^0(l) = 0.5\text{arg}(\hat{Q}_{ij}^0(l))$$ \hspace{1cm} (12)

and $\hat{Q}_{ij}^0(l)$ is given by equation (10).

The parameters that determine the filter shape are derived by posing the problem as a constrained linear MMSE estimation. The correlation function of the orientation data is assumed to be separable in the orientation of the feature, and like the filter to be used, Gaussian in shape because it is primarily the result of the smoothing operation of the pyramid. By least squares fitting, it is possible to estimate its shape and hence determine the parameters of the required MMSE filter $h_{mn}(\theta)$. Table II shows results of the enhancement process on the orientation estimate of the ‘shapes’ image (figure 5(a)). The vector estimation errors are given as a SNR, as a function of level, from the starting level (4) to full spatial resolution. To give an idea of scale, a 1 deg. error in angle corresponds to a SNR of 32dB. It can be seen that there is a significant improvement in SNR as a result of the processing.

### Boundary Graph

A set of candidate boundary nodes $B = \{\chi_0, \chi_1, \ldots, \chi_n\}$ are selected, where each $\chi_k$ denotes a scale-space position vector $(i, j, l)^T$, using a multiresolution peak detection procedure on the enhanced orientation pyramid. This is a top-down, recursive method where each boundary node, selected at a given level, is moved to the ‘peak’ amongst its children. Linking together these boundary nodes creates an approximate ‘dual’ of the region adjacency graph (figure 5(c)).

The enhanced orientation estimate is then used to control the relaxation of the boundary graph $B$. The method used has similarities to the boundary refinement algorithms used in [1] [14] and is related to the contour modification process outlined by Pavlidis and Liow [15]. Each boundary node $\chi_i$ is permitted to move from $\chi_i$ to a position that maximises the cost function:

$$C(i) = \frac{1}{\text{card}\{N_i\}} \sum_{j \in N_i} E(\hat{Q}(\lambda_{ij}))$$ \hspace{1cm} (13)

where $N_i$ is the neighbourhood of the node $\chi_i$, $\text{card}\{N_i\}$ the number of neighbours, and $\lambda_{ij}$ is the link between $\chi_i$ and its neighbour $\chi_j$. The energy function $E(\hat{Q}(\lambda_{ij}))$ is the average projection of the

<table>
<thead>
<tr>
<th>level</th>
<th>Input SNR (dB)</th>
<th>Output SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-4.4</td>
<td>21.9</td>
</tr>
<tr>
<td>1</td>
<td>13.1</td>
<td>28.0</td>
</tr>
<tr>
<td>2</td>
<td>27.5</td>
<td>32.6</td>
</tr>
<tr>
<td>3</td>
<td>38.4</td>
<td>39.8</td>
</tr>
<tr>
<td>4</td>
<td>50.8</td>
<td>50.8</td>
</tr>
</tbody>
</table>

**Table II**

Orientation enhancement results on ‘shapes’ 0dB image.
magnitudes of the orientation vectors ‘under’ the link $\lambda_{ij}$ onto the direction of the link. This function is a maximum when the link $\lambda_{ij}$ lies along the underlying orientation vectors.

The movement of each node is constrained by the local orientation. Nodes may move either perpendicular or parallel to this orientation. After each iteration there is a merging step to merge any nodes that are within some minimum distance set to be the size of the smallest block of the region tessellation. Nodes which lie on ‘straight-line’ segments are also eliminated if their removal does not significantly alter the local cost.

D. Process Interaction

There are two data paths to the interaction process (figure 1): a flow of information from the boundary to the region process and the reverse flow, from the region to the boundary process. The common denominator between the two processes is the estimate of the orientation and position, so this seems like the logical vehicle for the interaction.

The process interaction takes place in a neighbourhood which is defined in terms of the interaction of boundary links and the region blocks. Figure 3 illustrates a typical interaction neighbourhood. In this example boundary link $\lambda_{01}$ is defined to interact with all the blocks that it passes through, namely $\{\xi_0, \xi_1, \xi_2\}$. Similarly, the region node $\xi_2$ is defined to interact with all the boundary links that pass through it, namely $\{\lambda_{01}, \lambda_{12}\}$ and hence the boundary nodes $\{\xi_0, \xi_1, \xi_2\}$.

In the boundary-to-region interaction, the region edge estimate is combined with the local boundary orientation and position. The best linear estimate is used, i.e.

\[
\begin{align*}
\hat{\theta}^i_k &= \beta_r \theta_k + \beta_b \theta_{\xi_i} \\
\hat{\xi}_k &= \beta_r \xi_k + \beta_b \xi_{\xi_i}
\end{align*}
\]

(14)

where $\xi_{\xi_i}$ denotes the interaction neighbourhood of node $\xi_i$. The best combination coefficients $\beta_r$ and $\beta_b$ are derived from the estimates of the variances of the respective region and boundary orientation estimates [16]. Estimates are combined after the region based edge estimation. The boundary information thus influences the calculation of the prior region link probabilities and therefore the link decision process of equation (5).

The reverse data flow, from the region process to the boundary process, is done through the addition of an extra term to the boundary cost function of equation (13) which uses the current orientation estimate from the region information

\[
C'(i) = 1 / \text{card}\{N_i\} \sum_{j \in N_i} [\beta_b E(\theta_\lambda (\lambda_{ij})) + \beta_r E(\theta_\lambda (\lambda_{\xi_i}))]
\]

(15)
where the subscripts $b$ and $r$ denote the source of the orientation information, and $\mathcal{I}_{ij}$ is the interaction neighbourhood of link $\lambda_{ij}$. This modified cost is thus a linear combination of the two orientation energies. Once more a pair of combination coefficients, also derived from estimates of signal and noise variance of the respective orientation, are used to control the extent of the influence that either process has at a given iteration.

The convergence criterion for the iterative processing is when there is no significant change in the grey level estimate between iterations and there is little average movement of the boundary nodes. This is typically achieved in 10-20 iterations.

III. EXPERIMENTAL RESULTS

Figure 5(d) shows the segmentation result on the ‘shapes’ synthetic image of size $256 \times 256$, (8-bit pixel), to which Gaussian noise has been added with a standard deviation $\sigma = 20$, (Fig 5(a)). The luminance difference between the object and the background is 20 giving an inter-region signal to noise ratio $20/\sigma = 1.0$. The segmentation of the circle compares well with the results obtained by Spann and Wilson [14](Fig 6(a)). Its performance on the objects that have sharp corner features, however, is superior to those reported elsewhere. Figure 5(e) show an almost perfect segmentation result on a 3dB ($\sigma = 14$) version of the ‘shapes’.

The performance of the method over a variety of input SNRs is shown in figure 4. The graphs show root mean square (RMS) boundary error, which is obtained by averaging the squared boundary error, i.e. the square of the distance (in pixels) between the estimated boundary and the true boundary, and taking the square root of this value. An orientation estimation from the noise free ‘shapes’ test image was used to guide the measure. The dashed line shows the comparative results for an implementation of the Spann and Wilson method [14], and confirms the qualitative findings that the method is more accurate at locating object boundaries.

The results applied to two natural images are shown in figure 6 and figure 7. The ‘table’ image was chosen because it has many regions which are essentially ‘flat’, and also boundary shapes which are polygonal, which conforms better to the boundary model. The region result shown in figure 6(e) is not
as easy to interpret as that for the synthetic image. Although the major region have been identified, the local averaging has flattened out these regions. All the major curves have been detected in the boundary result (figure 6(d)).

The ‘Lena’ image (figure 7(a)) is considerably more complex than the ‘table’ image. The boundary result is shown in figure 7(b). There are clearly problems in the textured regions, such as the feathers that hang off the hat and the hair. Nevertheless, the curves of the mirror, hat and the vertical straight edges in the background have all been accurately found. However, less success has been achieved with the facial features, partly because of the high curvature around the eyes and the shallowness of the edge across the nose.

IV. CONCLUSIONS

A new algorithm for image segmentation, having a number of interesting features, has been described and been shown to be effective in locating regions and their boundaries with high accuracy from noisy data. In applications where region shape is important, it has a number of advantages over previously reported methods. It is computationally fast, flexible with respect to region and boundary models and uses only local processing.

There are several areas, however, in which consolidation and extension are required. These relate to the algorithm, model and the underlying feature description. The orientation estimation and subsequent boundary segmentation are currently two processes. There is a need to unify these two stages, which both involve vertical and lateral processing, to parallel the region processing. The results show that the segmentation is limited by the simplicity of the underlying model. The constant region model fails to deal adequately with many natural images because of variation of grey level across a neighbourhood. The extension from grey level to other region properties is a simple one (e.g. [3]) as is that from 2-D to problems in higher dimensions. The model for boundaries could also be extended to allow segments between nodes to be curved, i.e. quadratic arc segments or splines, which would result in a better fit to image curves. The use of a more generalised multiresolution feature description, such as the Multiresolution Fourier Transform (MFT) [6] [17], could provide a rich image description from which grey level, texture, edge, line and branch feature information could be extracted. Such developments are currently under way.

ACKNOWLEDGEMENT

This work is supported by the UK SERC and by Shell Research Ltd.

REFERENCES

Fig. 5. Results on ‘shapes’ image 256 × 256

Fig. 6. Spann-Wilson result on 'shapes', and 'table' results

Fig. 7. Results on 'Lena' image