SPIE Vol. 2315, Image and Signal Processing for Remote Sensing, pp. 214-224, 1994

Refining region estimates for post-processing image classification

Paul L. Rosin

Institute for Remote Sensing Applications
Joint Research Centre, I-21020 Ispra (VA), Italy
email: paul.rosin@jrc.it

ABSTRACT

This paper describes a method for post-processing classified images to enable generalisation to be performed whilst maintaining or improving the accuracy of region boundaries. This is achieved by performing region growing, and incorporates both spatial context and spectral information. In contrast, few classifiers use any spatial context, and many post-processing techniques, such as iterative majority filtering, discard all spectral information. If class models are available these can also be included in the region growing process, otherwise, the algorithm operates in a data-driven mode, and locally estimates models for each region.

1 INTRODUCTION

The most common classification methods applied to remotely sensed images operate on a per pixel basis. Since no spatial information is incorporated these approaches often produce many noisy, isolated, small regions due to local variations in the measured spectral values. For a range of applications this is undesirable for the end-users of these thematic maps. Generally it is preferable to "generalise" the map so that the overall information is retained but unnecessary or spurious detail is removed. Not only is this detail visually unpleasing, but it is unsuitable for input to Geographic Information Systems. Poor segmentation may also prove problematic for further processing of the map (e.g. automatic detection of roads, airports, etc.) since many regions in the image will not correspond to significant objects in the scene. These limitations have been tackled in several ways:

- Pre-processing of the image prior to per-pixel classification. This usually takes the form of smoothing.[1]
- Post-processing of the classification map. There are a variety of methods to remove small regions. These include iterative majority filtering based methods[1,17,20] erosion followed by dilation,[4] region merging,[3,9] and mixed class texture labelling.[3] Alternatively, rather than consider only the class labels, relaxation techniques can be applied to update class probabilities.[1,18] A disadvantage of such techniques is that in addition to small regions they also tend to remove more significant long, thin regions. Furthermore, while incorporating spatial information, most of these methods ignore spectral information, and consequently the region boundaries can be substantially distorted from the true image boundaries.
- Segmentation of the image followed by region based classification.[7] A problem with such techniques is the validity of the segmentation. Since segmentation is a difficult task the algorithms tend to be very dependent on user supplied parameters, and are prone to errors.

- Incorporating spatial context within the classification process. A limitation of such methods is that the spatial context is normally restricted to a small fixed size window about each pixel,[2,10,14,15,19] although this can be alleviated by applying the classification process recursively.[6]

To overcome the problems of the above techniques we describe an algorithm that starts with an initial segmentation of the image which is iteratively modified to produce a better partitioning of the image. In comparison to the some of the above approaches it operates on regions which have a closer relationship with the objects in the image rather than arbitrarily sized and shaped windows. In addition, it uses both the class labels and the image intensities, and therefore more of the available information is used in the relabelling processes than those methods than discount the image intensities and use only the labels.

2 REGION GROWING

The algorithm starts with an initial partitioning of the image which could have been produced by any segmentation method. For instance, the image could be segmented into homogeneous regions using region growing or split-and-merge type algorithms. Examples of our method applied to this type of input are given elsewhere.[12] Alternatively, the image can be classified (using a per pixel classifier) and maximally connected regions with identical classification extracted by running a connected component labelling algorithm. This is the approach described in this paper.

Given the partitioning of the image into roughly homogeneous regions the algorithm operates by modifying the pre-existing region boundaries. But, in contrast to majority filtering type approaches, class relabelling at a pixel is performed with reference to the adjacent image regions rather than a fixed size square window centred on the pixel.

Each point along the boundary of each region can be thought of as exerting an outwards force. In other words, every region is trying to grow outwards. Apart from points on the image border, which are naturally unable to extend beyond the image, every region boundary pixel is adjacent to another region boundary pixel, both of which are trying to grow into each other. The one exerting the greatest force wins, and that pixel is removed from the weaker region and assigned to the stronger region. This process occurs in parallel at every pixel in the image that is on the boundary of a region (excepting the image borders), and is iterated until the regions' forces reach an equilibrium and no more changes (re-assignments) occur.

The force F_A exerted at the pixel P by region A is inversely proportional to its deviation from the expected value, and is given by:

$$F_A = \frac{1}{|E_A(P) - I(P)|}$$

where $E_A(P)$ is the expected value at the pixel P given the region model A, and $I(P)$ is the image value at P. The force F_B exerted at the pixel P by region B is calculated likewise. The deviation between the measured and expected value is calculated as the Euclidean distance between the two values, and so can be applied to any dimension of data (e.g. multi-spectral images).

Since a binary decision for the labelling is made, only the relative strengths of the forces need to be compared and their absolute values are not considered. This qualitative approach eliminates the need for error tolerances (on homogeneity for instance) that most other region growing techniques use. Additionally it makes the decision rule invariant under uniform scalings and offsets of the image values. In fact, we do not need to explicitly calculate the force, but use the deviations instead. Thus, the decision rule to label pixel P currently in region A and adjacent to region B becomes:

$$L_{i+1}\left(P \mid L_i(P) = A\right) = \begin{cases} B & |E_B(P) - I(P)| < |E_A(P) - I(P)| \\ A & \text{otherwise} \end{cases}$$

where $L_i(P)$ denotes the label of pixel P at iteration i. In other words, if the deviation at pixel P from the expected value given region B's model is less than the deviation from the expected value given region A's model then the pixel is relabelled as part of region B, otherwise it remains part of region A. Thus, if the two forces are equal no action is taken.

2.1 Topology and Region Deletion

The algorithm primarily moves region borders and does not explicitly consider region topologies. However, because pixel relabelling is performed locally this can have the effect of producing topological changes such as the splitting, merging, and annihilation of regions. The ability to alter the topology of the initial estimate depending on the image values can be extremely beneficial if the topologies of the objects in the image are not known in advance.[8,16] For more details concerning the topological behaviour and convergence properties of the algorithm see Rosin.[12]

On the other hand, in some situations topological changes are undesirable. In such cases the original region topology is preserved by a two pass process. First relabelling is performed as before, allowing arbitrary changes in topology. Then the second pass "repairs" the image topology. Each region has a unique label determined by the initial connected component analysis. At every iteration of the region growing process another connected component analysis is performed. A mapping is made between the labels of each region before and after relabelling. This allows split regions to be easily identified by checking for a pre-growing label that maps onto more than one post-growing label. The image topology is restored by reducing each set of subparts produced from one originating region to a single region. This is done by keeping the largest subpart and deleting all other subparts. Here we assume that the largest subpart is likely to be the most significant. Other possible criteria for deletion could involve the homogeneity or shape of the region. Deleting a region is implemented by relabelling its pixels with a null region label. During further iterations of region growing such deleted regions are assigned zero outwards forces, and are therefore always relabelled by adjacent regions. The number of iterations required to fill the deleted region will depend on its width.

Regions can also be eliminated for reasons other than preserving topology. For instance, to reduce the effects of the characteristic speckle produced by feature space clustering methods small regions can be deleted. This is related to the region merging method[3,9] except that a deleted region is not necessarily uniformly relabelled. Different parts may be relabelled by the various labels of the adjoining regions.

2.2 Region Models

The expected value at a pixel depends on the region model being used. There are two main approaches to estimating the parameters of the region models: data-driven or model-driven. In the former case no prior information is provided and the parameters of each region model are estimated from the intensity values in the initial region before the boundary modification stage. Alternatively, class models can be derived from the training data. Then each region will be assigned the model for the corresponding class that it (exclusively) contained before the boundary modification stage. In both cases we assume a constant intensity per band within each region/class (although more complex models such as the inclined plane can be incorporated in the algorithm[12]).

The two approaches have their advantages and disadvantages. The data-driven approach is more adaptive to the local image intensities. Thus, if the statistics of the spectral values of a landcover class vary within the image (e.g. due to variations in the underlying ground slope, elevation, orientation, etc) then each region can estimate the local statistics more precisely than a global class statistic. On the other hand, working in a purely data-driven mode means that some of the valuable information available in the training data is lost. For instance, the training data will consist of well chosen examples of each class which may provide a better estimate of the

class statistics. In particular, the data-driven estimates calculated for small regions will tend to be unreliable.

2.2.1 Data-Driven Model Estimation

The model intensity for each region is estimated for each band by the median of the intensity values within the region. The median is a robust estimator which is necessary since the initial segmentation may be inaccurate. If a simple average was used instead then the estimated parameters would be biased by the intensities of neighbouring objects in the image that the region incorrectly overlapped.

2.2.2 Model-Driven Model Estimation

Since the model-driven approach estimates the class model parameters from the training data it should not have to cope with the potentially large amounts of incorrect class labelling presented to the data-driven approach by a poor initial segmentation. Therefore, we assume that the covariances between each of the bands can be determined reliably, and these are used to augment the class models. However, even if the training data is not too noisy, because we use a simple class model it is important to robustly estimate its parameters since the training data might not follow a nice uni-modal Gaussian distribution. This is particularly important in our example which uses a neural net for classification since no *a priori* assumptions were made about the class distributions when selecting the training data.

We have experimented with various methods for determining typical band intensities and band covariances for classes. The first is the usual mean and covariance. The covariance of two sets of variables $X = \{x_1, ..., x_N\}$ and $Y = \{y_1, ..., y_N\}$ is given by

$$C_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x}) (y_i - \bar{y})$$

where the mean is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The second method replaces the mean values by their medians

$$\tilde{x} = \underset{i=1}{\overset{N}{\text{med}}} x_i$$

and the covariance is calculated using these median values

$$C'_{XY} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \tilde{x}) (y_i - \tilde{y})$$

Finally, the third method uses the median rather than the average of the product of the variances

$$C''_{XY} = \underset{i=1}{\overset{N}{\text{med}}} [(x_i - \tilde{x}) (y_i - \tilde{y})]$$

The use of medians instead of means should make the estimators more robust. However, they can still be calculated efficiently and easily, in comparison to more complex techniques from robust statistics such as the minimum volume ellipsoid or the minimum covariance determinant estimators.[13]

When performing region growing the estimated class covariances can be incorporated into the calculation of the region forces. Rather than using the Euclidean distance between the pixel under consideration and the region

model as before, the squared Mahalanobis distance is used instead

$$MD^2(\mathbf{x}, c) = (\mathbf{x} - \mu_c)^T \, \mathbf{\Sigma}_c^{-1} \, (\mathbf{x} - \mu_c)$$

where $\mathbf{x}$ is the vector containing the multi-channel spectral values of the pixel, c is the class of the region, and μ_c and $\mathbf{\Sigma}_c$ are the average and covariance of the class calculated as described in section 2.2.2.

3 THE ALGORITHM

The algorithm can now be summarised as follows:

```
perform connected component analysis to identify initial regions
if in data-driven mode then
        estimate parameters of regions from input image
else if in model-driven mode mode then
        estimate parameters of classes from training data
repeat
        at every region boundary pixel do
                calculate opposing forces from current and adjoining regions
                (re)assign pixel with label of strongest region
        if topology is to be preserved then
                perform connected component analysis
                delete multiple regions
until no changes occur
```

4 APPLICATION OF THE REGION GROWING ALGORITHM

Post-processing the classified image to remove small noisy regions can now be approached in either of two ways:

- Perform iterative majority filtering (for example) to remove noise, and then apply the region modification method to correct the region distortion caused by the filtering.
- Delete small regions (i.e. label as unclassified) and then apply the region modification. A deleted region exerts no force and so the gaps will filled by the most suitable neighbouring regions.

Experiments show that both approaches produce similar results. It is therefore preferable to perform IMF first since this requires less computational resources. Since the region growing method must represent each region explicitly (in contrast to IMF) many small regions leads to large memory requirements. If only large regions are to be kept then both approaches can be combined: first applying IMF to remove the fine detail, and then further deleting regions below a size threshold prior to region growing.

There are several possible parameters to the algorithm which can be set to suit the application. For instance, rather than performing the region growing until convergence it is possible to stop at any time. Setting the maximum number of iterations to be performed enables an upper limit on the movement of region boundaries to be enforced. Other parameters are the minimum region size, whether to preserve the initial topology or not, and the data-driven versus model-driven approach.

5 EXAMPLES

The following examples are demonstrated on portions from a Landsat-TM image taken near Lisbon in Portugal. The first example demonstrates how the the classified image can be cleaned up without distorting the region boundaries. Figure 1a shows band 5 of the image with the classification (based on six bands) into sixteen classes produced by a multi-layer perceptron neural network.[5] The noisy appearance of the classification is evident. Performing one iteration of majority filtering (using a 3×3 window) removes many of the smallest regions (figure 1c). However, many noisy patches still remain. Most of these can be removed by further filtering, but this is at the cost of substantially distorting the remaining regions. Figure 1d shows the results of performing IMF until no further changes occur (12 iterations). The results of applying the data-driven region growing using the IMF filtered output as the initial segmentation of the image is shown in figure 1e. After 33 iterations the algorithm converges, and no further changes are made. The distorted regions have been grown back to match the underlying intensities of the image. This can be seen more clearly by overlaying the classification boundaries for each of the three filtered results on top of the intensity image. Even after a single iteration of majority filtering (figure 1f) many boundaries are slightly misplaced (e.g. the dark fields in the top left hand corner). This is further aggravated by the full IMF in which regions are often shrunken and rounded, displacing boundaries. The region growing boundaries shown in figure 1h can be seen to closely track the boundaries of objects in the scene.

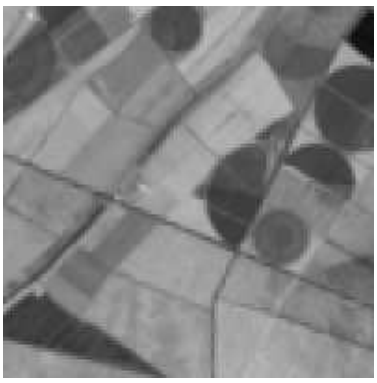
Although the IMF produces a relatively small number of regions as input to the region growing, these regions may then split during the growing process. Therefore the final number of regions is not constrained. Thus, in figure 1e there are various small regions that have "broken-off" from larger ones. These can be eliminated in a variety of ways. For instance, one iteration of majority filtering might be sufficient. Alternatively, figure 1i shows the results of region growing with the topology preserving option. This has prevented small spurious regions being formed, and the final number of regions is the same as its input (the IMF filtered result).

As an alternative to initially applying IMF, figure 1j shows the results of region growing directly from the raw classification in figure 1b and initially deleting all regions less than 22 pixels in size. This size threshold was chosen experimentally to produce the same number of final regions as the previous region growing example in figure 1e.

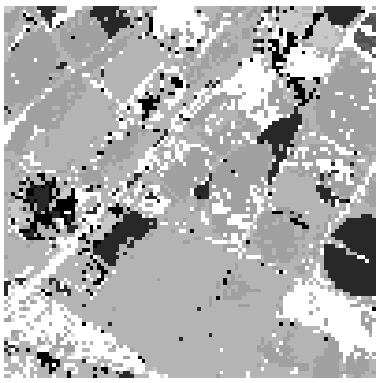
The results of these three different options (figures 1e,i, and j) show similar boundaries, although due to the different region deleting processes there are differences in the final set of regions. For instance, the grey region inside the white triangular region in the bottom left hand corner in figure 1e remains in figure 1i, but has been deleted in figure 1j.

Although the spurious small regions split off in the previous example may be undesirable, in other cases this process may be useful. This is demonstrated in another example of data-driven region growing shown in figure 2. The first principle component of the image is shown in figure 2a, and its classification in figure 2b. It can be seen that the separating patch between the two light fields in the top of the image is fragmented, causing the fields to be incorrectly joined to form a single region. When IMF is applied the join is only strengthed since the fields are so close. (figure 2c). When region growing is applied the fields are better delineated, and are also separated into individual regions (figure 2d). Unfortunately, in the process substantial portions of the image have changed classification. This generally arises because a small region fits the intensity data of most of an adjoining region better than that region itself. This can be seen most clearly by the large black central region (that also extends to both sides of the image). After region growing the smaller light gray coloured regions have "overgrown" the central part of the black region which is forced out into the thin strips corresponding to the borders between the fields. The black region fits the data poorly since the IMF processing has connected various disparate regions to form one large region with substantial variation in intensity. However, this misclassification can be remedied by a further post-processing step in which each region is reclassified according to its median intensity (for each band). As shown in figure 2e this results in the correct classes while maintaining accurate region boundaries.

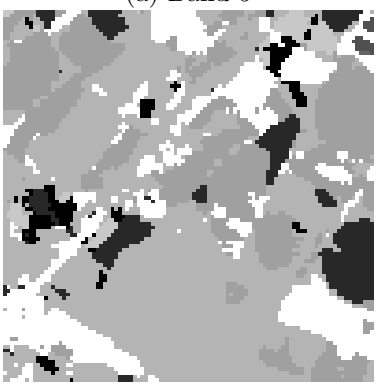
Figure 2f shows the results of applying model-driven region growing rather than the data-driven approach



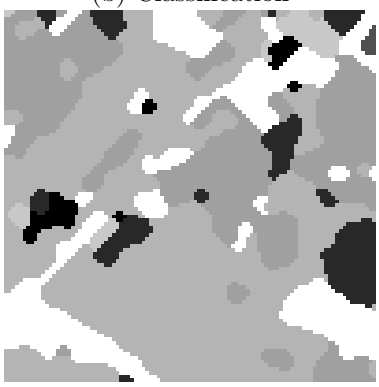
(a) Band 5



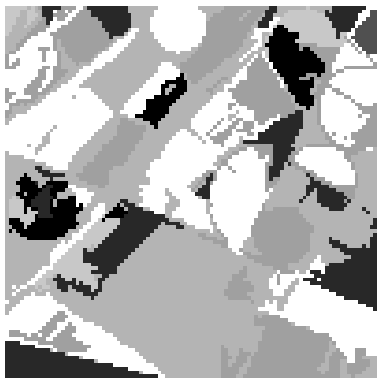
(b) Classification



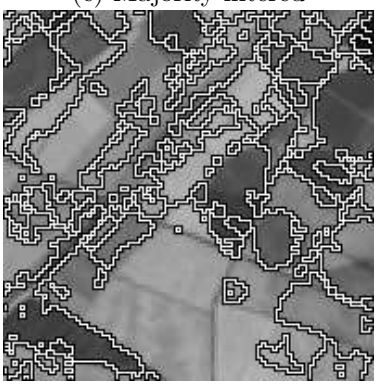
(c) Majority filtered



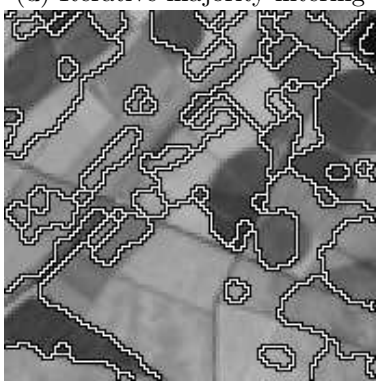
(d) Iterative majority filtering



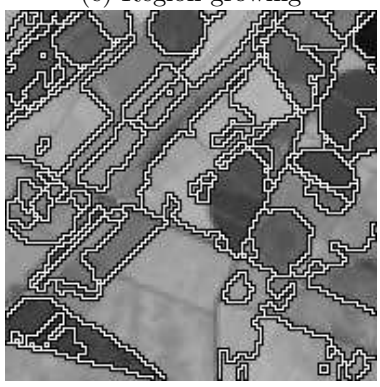
(e) Region growing



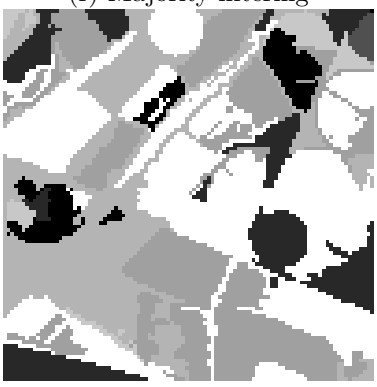
(f) Majority filtering



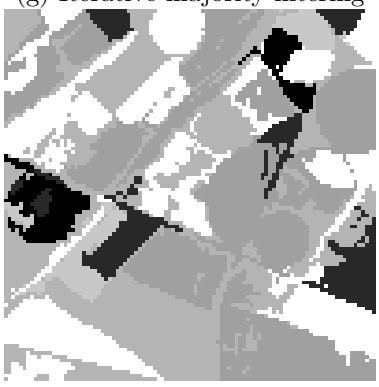
(g) Iterative majority filtering



(h) Region growing



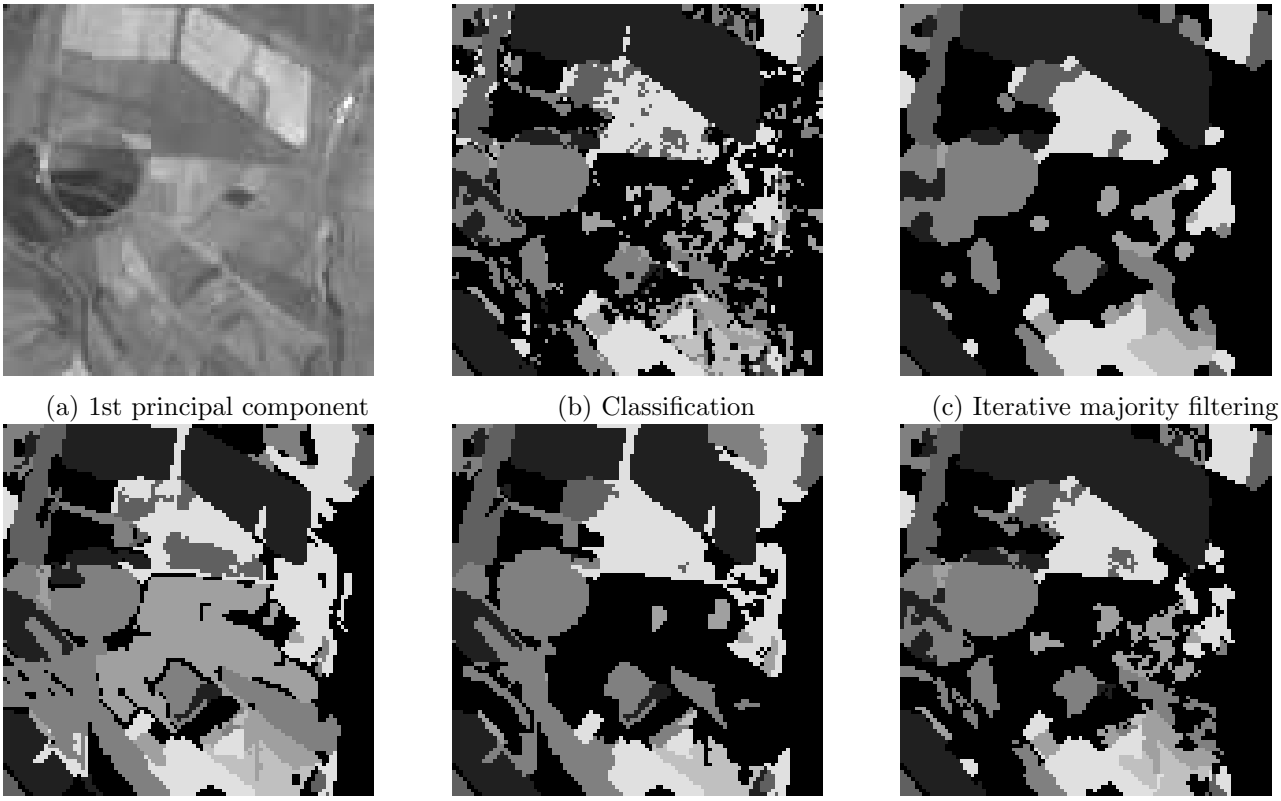
(i) Constraining topology



(j) Deleting small regions

Figure 1: Improving delineation of objects by region growing

using the second method for estimating class averages and covariances. The first method performed similarly, but the third method gave worse results. It can be seen that this time the fields have not been split. The reason is because the light gray coloured region just below the fields has not grown into the field region since its associated class has a higher uncertainty (as estimated by the covariance matrix) than the field region's class. This is known because the fields are still split if the Euclidean distance (from the class averages) is used instead of the Mahalanobis distance.



(a) 1st principal component (b) Classification (c) Iterative majority filtering

(d) Region growing (e) Reclassified (f) Model-based growing

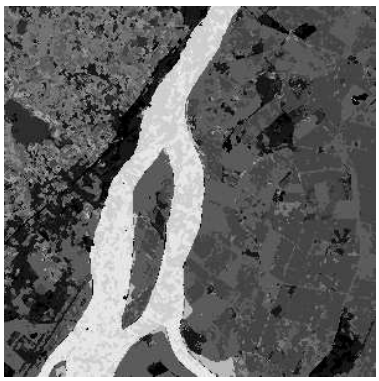
Figure 2: Improving segmentation of objects by region splitting

The final example, taken from the same Landsat image, uses the CORINE landcover map as ground truth. In the area we considered 11 different landcover classes were present (according to the CORINE map). This landcover map covers a large contiguous area, and is useful for assessing the quality of spatial filtering. However, since very broad classes were designed (which are suitable for human photointerpreters) there is much spectral confusion between the classes which causes problems for automatic classifiers.

Figure 3a shows band 5 of the area of interest. The associated ground truth is shown in figure 3b. Due to the minimum parcel size of 25 Ha used in the CORINE project the regions have been considerably generalised. In contrast, the per pixel classification is much noisier, and there is substantial class confusion (figure 3c). IMF reduces the fine noise, but there are still many regions below the 25 Ha limit. Applying the data-driven region growing with a minimum region size of 277 pixels (corresponding to the 25 Ha limit) results in the more generalised image in figure 3d. If IMF with a large enough mask size had been used instead to eliminate all the small and medium sized regions the remaining regions would have been extremely distorted. The regions in figure 3d still track the image features very accurately. It was experimentally determined that the optimal region size threshold for this image was slightly larger than 277 (i.e. 400), and these results are shown in figure 3e. The model-driven approach produced similar results (figure 3f). It can be seen from table 1 that the region growing process improved the accuracy relative to the CORINE ground truth over the initial per pixel classification. However, it should

(a) Band 5

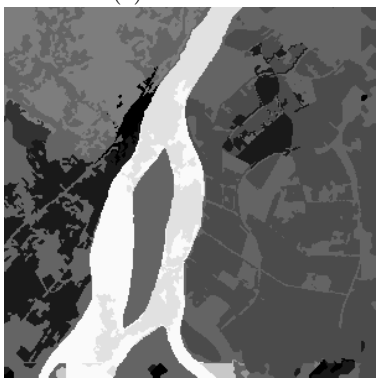
(b) CORINE ground truth



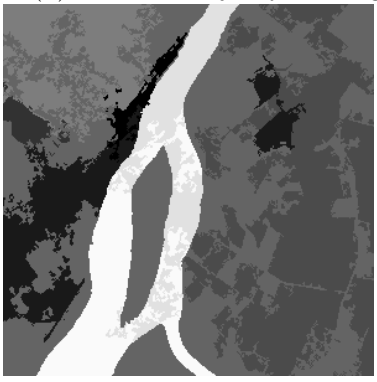
(c) Classification

(d) Iterative majority filtering

(e) Region growing (T = 277)



(f) Region growing (T = 400)



(g) Model based growing

Figure 3: Improving accuracy by deleting regions

be stressed that the aim of the region growing is primarily to perform generalisation and improve boundary localisation rather than improve overall classification accuracy.

Surprisingly, the data-driven approach gave better accuracies than the model-driven approach. Moreover, it can be seen that the standard estimators for class averages and variances fared better than the median based ones. Possibly this arises because primarily unimodal class models were assumed, whereas the CORINE classes do not form well defined unimodal (Gaussian) distributions.

PROCESSING METHOD	**ACCURACY (%)**
raw classification	49.1
IMF	55.0
data-driven region growing ($T = 277$)	59.5
data-driven region growing ($T = 400$)	61.4
model-driven region growing ($T = 400$; using C)	59.3
model-driven region growing ($T = 400$; using C')	56.5
model-driven region growing ($T = 400$; using C'')	44.1

Table 1: Results for classification filtering relative to CORINE map

6 DISCUSSION

We have described a model for region growing based on the similarity between the expected values within the region and the pixel values. In data-driven mode this is a Euclidean distance, while the model-driven mode incorporates covariances between bands, and uses the Mahalanobis distance. It is also possible to add various other terms into the force exerted by each region. Examples include the expected size and shape of a region, and the amount of movement expected/allowed from its original estimated location. Currently each pixel is examined individually during the growing process. To incorporate shape information a larger window would have to be examined to be able to calculate a shape measure (e.g. curvature). The above examples all determine a region's forces independently from the other regions in the image. Alternatively, global statistics could be incorporated. For instance, following Wilkinson,[20] growing could be performed with the aim of removing small regions while maintaining the class statistics (i.e. the proportions of classes in the raw classified image). This would be done in model-driven mode by modifying the force of each class at every iteration to the advantage of the under-represented classes (which have been reduced on previous iterations) to encourage them to expand. There is a similarity between our region growing approach and (closed) active contours[11] since both perform boundary modification driven by various forces. The main difference is that the forces of the active contours are usually restricted to the image gradient at the contour and the shape of the contour without considering the intensity values within the contour.

The method described in this paper is constrained by the initial segmentation of the image. If this segmentation is reasonably good then the region growing can be applied without requiring any parameters. However, in the case of poor initialisations (e.g. a noisy classification without any post-processing) the segmentation needs to be improved by eliminating some of the spurious regions. This is done by deleting small regions, and requires a user-set parameter. If additional terms were included in the growing forces as described in the previous paragraph then these also have to be integrated, requiring weights to specify the importance of each in a similar manner to the energy term used by snakes.

Future work will attempt to overcome the restriction of assuming roughly unimodal models. This can be done by performing cluster analysis on the training data and assigning a subclass model to each cluster. The force exerted by a region at each pixel will then be calculated as the maximum force exerted by all of its subclass models.

7 REFERENCES

[1] J.O. Eklundh, H. Yamamoto, and A. Rosenfeld. A relaxation method for multispectral pixel classification. *IEEE Trans. PAMI*, 2(1):72–75, 1980.

[2] M. Goldberg and D.G. Goodenough. Analysis of a spatial filter for landsat imagery. *J. of Appl. Photo. Eng.*, 4:25–27, 1978.

[3] L.J. Guo and J.M. Moore. Post classification processing for thematic mapping based on remotely sensed image data. In *Proc. IGARS, Finland*, pages 2203–2206, 1991.

[4] R. Haralick and L.G. Shapiro. *Computer and Robot Vision.* Addison-Wesley, 1992.

[5] I. Kanellopolous, A. Varfis, G.G. Wilkinson, and J. Megier. Land cover discrimination in SPOT HRV imagery using an artificial neural network. A 20 class experiment. *Int. J. Remote Sensing*, 13:917–924, 1992.

[6] J. Kittler and J. Foglein. Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2:13–29, 1984.

[7] D.A. Landgrebe. The development of a spectral-spatial classifier for earth observational data. *Pattern Recognition*, 12:165–175, 1980.

[8] R. Malladi, J.A. Sethian, and B.C. Vemuri. A fast level set based algorithm for topology-independent shape modeling. *J. Mathematical Imaging and Vision*, 6(2-3):269–289, June 1996.

[9] D.G. Nichol. Region adjacency analysis of remotely-sensed imagery. *Int. J. Remote Sensing*, 11:1089–1010, 1990.

[10] A. Owen. A neighbourhood-based classifier for LANDSAT data. *The Canadian Journal of Statistics*, 12:191–200, 1984.

[11] T. Pavlidis. A critical survey of image analysis methods. In *Proc. 8th Int. Conf. on Pattern Recognition*, pages 502–511, 1986.

[12] P.L. Rosin. Refining region estimates. Technical Report I.94.05, Institute for Remote Sensing, Joint Research Centre, Ispra, Italy, January 1994.

[13] P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection.* Wiley, 1987.

[14] P.H. Swain, H.J. Siegel, and B.W. Smith. Contextual classification of multispectral remote sensing data using a multiprocessor system. *IEEE Trans. Geoscience and Remote Sensing*, 18:197–203, 1980.

[15] P. Switzer. Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Mathematical Geology*, 12:367–376, 1980.

[16] R. Szeliski, D. Tonnesen, and D. Terzopoulos. Curvature and continuity control in particle-based surface models. In *Proc. SPIE 2031: Geometric Methods in Computer Vision II*, pages 172–181, 1993.

[17] I.L. Thomas. Spatial post processing of spectrally classified Landsat data. *Photogrammetric Engineering and Remote Sensing*, 49:1201–1206, 1980.

[18] T. Watanabe and H. Suzuki. Experimental evaluation of classifiers using spatial context for multispectal data. *Systems and Computers in Japan*, 19:33–47, 1988.

[19] S.W. Wharton. A contextual classification method. *Pattern Recognition*, 15:317–324, 1982.

[20] G.G. Wilkinson. The generalisation of satellite-derived thematic raster maps for GIS input. *Geo-Informations-Systeme*, 6:24–29, 1992.