INT. J. REMOTE SENSING, 1998, VOL. 19, NO. 14, 2685–2704

Classification by progressive generalization: a new automated methodology for remote sensing multichannel data

J. CIHLAR†¶, QINGHAN XIAO‡, J. CHEN†, J. BEAUBIEN§, K. FUNG† and R. LATIFOVIC‡

†Canada Centre for Remote Sensing, Ottawa, Ontario, K1A 0Y7, Canada
‡Intermap Technologies, Ottawa, Ontario, K2E 1A2, Canada
§Canadian Forest Service, Sainte-Foy, Quebec, G1V 4C7, Canada

(*Received 21 April 1997; in final form 3 December 1997*)

Abstract. A new procedure for digital image classification is described. The procedure, labelled Classification by Progressive Generalization (CPG), was developed to avoid drawbacks associated with most supervised and unsupervised classifications. Using lessons from visual image interpretation and map making, non-recursive CPG aims to identify all significant spectral clusters within the scene to be classified. The basic principles are: (i) initial data compression using spectral and spatial techniques; (ii) identification of all potentially significant spectral clusters in the scene to be classified; (iii) minimum distance classification; and (iv) the use of spectral, spatial and large-scale pattern information in the progressive merging of the increasingly dissimilar clusters. The procedure was tested with high- (Landsat Thematic Mapper (TM)) and medium- (Advanced Very High Resolution Radiometer (AVHRR) 1 km composites) resolution data. It was found that the CPG yields classification accuracies comparable to, or better than, current unsupervised classification methods, is less sensitive to control parameters than a commonly used unsupervised classifier, and works well with both TM and AVHRR data. The CPG requires only three parameters to be specified at the outset, all specifying sizes of clusters that can be neglected at certain stages in the process. Although the procedure can be run automatically until the desired number of classes is reached, it has been designed to provide information to the analyst at the last stage so that final cluster merging decisions can be made with the analyst's input. It is concluded that the strategy on which the CPG is based provides an effective approach to the classification of remote sensing data. The CPG also appears to have a considerable capacity for data compression.

1. Introduction

Image classification, i.e. categorization of pixels based on their spectral (or other) characteristics, is one of the fundamental analysis techniques for remotely sensed data, with land cover mapping arguably being the most frequent application. Based on the tradition of surveys and mapping by means of aerial photography, numerical image classification methods were developed when data became digital and calibrated and as computer power increased. Currently used methods fall into two basic categories, supervised and unsupervised (Duda and Hart 1973). Even recent classification strategies such as evidential reasoning (Peddle 1993) and neural networks (Benediktsson *et al.* 1990) are forms of supervised classification, while classification trees (Hansen *et al.* 1996) represent an unsupervised classification procedure.

¶E-mail: josef.cihlar@ccrs. nrcan.gc.ca.

Supervised classification assumes that the locations of some land cover classes/surface conditions are accurately known so that they can be used to establish a pattern through which other similar locations can be found. In the case of unsupervised classification, no knowledge of surface conditions is assumed (apart from the statistical characteristics, such as the probability distributions, and not in all cases). Unsupervised classification algorithms attempt to identify the 'natural' groupings in the data, under the constraints of the classification process specified by the analyst.

If enough is known about the surface, supervised classification is likely to be the tool of choice because it permits effective exploitation of the data content. Unfortunately, 'enough' may be too demanding because the surface information must be sufficient to decide *a priori* which parts of the spectral content are significant and which should be ignored. For example, there may be spectral variations within cover type, not all of which may be known in advance. Although for limited areas it is possible to have sufficient knowledge of ground conditions to make such decisions, this becomes increasingly more difficult over larger areas (e.g. $100\,km^2$) because of the cost of field data collection and other reasons. Furthermore, the spectral characteristics of a cover type are likely to change with distance, as many signature extension studies have concluded in the past. Given the current interest in land cover classification over large areas (e.g. Townshend *et al.* 1994) the drawbacks of supervised classification are evident.

Since unsupervised classification aims to identify all important spectral clusters without necessarily knowing at the outset which among them are thematically significant, this classification procedure has an inherent advantage over supervised classification—that is, one need not specify *a priori* the classes and their spectral expression. Another important feature is operator independence. Given that the same result can be obtained for the same data set by various analysts or that consistent results can be obtained by one analyst over different areas, this method has inherent advantage for scientific studies encompassing large areas and/or periods of time. However, unsupervised classification also has disadvantages. Since it strives for an overall optimum spanning of the entire data set it may miss specific detailed, but relevant, information. Such loss would not occur in supervised classification once the analyst knows of the existence of the classes of interest, however small they may be.

Another important disadvantage of unsupervised classification is the dependence of the result on the parameters guiding the classification process. For example, Bryant (1978) reviewed the numerous choices an analyst must make in using unsupervised classification, pointing out the need to optimize the parameters for a given data set through trials. The limitations of clustering methods have been well recognized (Ball and Hall 1965, Duda and Hart 1973, VanderZee and Ehrlich 1995). In fact, by a choice of the clustering parameters it may be possible to produce various numbers of clusters for the same data set, ranging from one to the number of spectral value combinations present in the input data. The spectral clusters should be thematically most meaningful at some intermediate number which is almost always higher than the number of desired thematic classes but is not known *a priori*.

An ideal classification method should be: accurate; reproducible by others given the same input data (failing of supervised classification); robust (not sensitive to small changes in the input data) yet able to exploit fully the information content of the data; applicable uniformly over the whole domain of interest; and objective (not dependent on the analyst's decisions). This paper describes a new classification

procedure that aims to satisfy as many of these criteria as possible. The philosophy behind the approach is outlined in the next section, followed by a description of the steps in the classification process. The performance of the new method is assessed on various data sets and in comparison with a common unsupervised classification algorithm.

2. Principles

In examining the possible approaches to digital classification it is instructive to consider visual image interpretation, the most successful operational classification methodology to date. Starting with an image (often enhanced to make the perception of colour differences easy), an interpreter uses spectral (colour) and spatial (texture, pattern, shape, context and other) clues to identify dominant classes in the image (Rabben 1960). He (or she) aims to detect the largest classes based on the area represented, and the classification legend will thus have no more than several dozen classes. In other words, larger classes are 'more important'. He then identifies other classes to make sure all the relevant conditions are captured by the analysis. Furthermore, the final map will not contain very small individual areas because of the accepted practice, both in field mapping and in cartography, of a minimum mapping unit or equivalently, the largest contiguous area of any class to be ignored as a separate spatial entity in the final product. The strategies of finding the largest clusters, using spectral as well as spatial measures, and merging classes which are spatially highly interspersed can be usefully employed in digital analysis as well.

It should be noted that a human interpreter does not use the spectral information as thoroughly as a computer. While humans perceive colour differences very well they can distinguish only a relatively limited number of radiometric levels (e.g. Cihlar and Protz 1972). The success of the visual image interpretation is primarily due to the importance of spatial clues such as shape, texture, pattern, shadows, height, context, etc. However, these spatial clues have different meanings and importance as the sensor resolution and area coverage change. For example, coarse resolution data imaging large areas have texture and some broad pattern information but beyond that the major clues are spectroradiometric. In such cases, digital classification should outperform visual analysis, especially if texture and pattern can be quantified through computer calculations.

One example of a successful combination of visual and digital analysis has been produced in recent years (Beaubien 1994, Beaubien *et al.* 1997, 1998). The essence of Beaubien's methodology is visual identification of important colours (regarded as 'training areas') in an image which had been computer enhanced in a standardized manner. This identification is made more reproducible and consistent by a customized transformation in which specific colours represent distinct, known cover types (typically needleleaf and broadleaf forest and bare ground). The second mechanism for increasing the consistency of the choice of training areas is the simplification of the spectral space: the original spectral channels are quantized to retain a limited number of possible colour combinations, typically $\leqslant 1024$, without losing significant information (as judged by visual examination of the original and quantized data). Pixels representative of individual colours/training areas are selected in the quantized image based on visual analysis and the analyst's judgement; one or more pixels (adjacent or not) may be selected to represent a colour/create a training area. The quantized image is thus treated as a classification with a very large number of classes and the subsequent steps are designed to identify the key spectral classes which account for

most of the imaged area. This methodology was proven to be significantly more powerful than a standard supervised classification because it allows the selection of all the training areas which are important for the scene of interest and the objectives of the classification (Beaubien *et al.* 1997, 1998). The principal disadvantage of this methodology is operator dependence, albeit more limited that in a supervised classification, in the selection of areas for radiometric normalization and in selecting pixels to represent individual colours/training areas.

The crux of the spectral classification problem is to determine the location and size of the clusters in multi-dimensional spectral space, or for simplicity their means and standard deviations in each of the input data channels. This is the major challenge in using the unsupervised classification method which would otherwise be the preferred method for applications where ground data are unavailable or limited (except for cases where subtle distinctions are important). The method development discussed below thus includes an objective way of identifying meaningful and spectrally representative cluster centroids, in both number and location in the spectral space.

The procedure developed in this study is based on the following assumptions.

1. In classifying an image, one should consider all the spectral combinations in the input data ('clusters') as entities and, using appropriate criteria, combine these to fit the desired classification legend.
2. Larger spectral clusters in the data set are more important. Small clusters do not 'matter', i.e. they will not be retained in the final classification, regardless of their spectral uniqueness.
3. Histogram quantization helps identify significant spectral clusters/regions in the data set without loss of significant information.
4. Adequate seed clusters can be established by combining small, pure, spectrally adjacent clusters (i.e. by spectrally broadening the cluster in the spectral space).
5. Two clusters should be combined if they are close together spatially and/or spectrally; which of these measures are used and how should depend on the number of clusters remaining.
6. By allowing only cluster merging and by limiting the number of clusters through the minimum size of cluster that can be ignored (smallest number of pixels) one avoids important disadvantages of the unsupervised classification, i.e. how many clusters should be retained and when to split or to merge clusters.
7. Since classification is a human construct it is not generally feasible to devise a fully automated computer algorithm and the process must allow for human judgement to influence/confirm the final results.

The procedure developed on the basis of these assumptions has been dubbed Classification by Progressive Generalization, or CPG. Using a combination of image enhancement, unsupervised classification and visual analysis, CPG aims to identify all potentially significant spectral clusters in the data and to group these until a reasonable number of thematically meaningful clusters is obtained for labelling. The new procedure has the following features.

(i) Use of histogram quantization to reduce the maximum possible number of spectral clusters.
(ii) Use of spatial filtering to identify spatially dominant spectral clusters.
(iii) Use of the largest pure clusters (i.e. one level in each of the quantized input channels) as seed clusters.

(iv) Use of progressive decrease of radiometric resolution to merge (thus reducing the number of) initially small clusters to be used as additional seed clusters.
(v) Use of the minimum pixel-to-cluster centroid distance to assign pixels to clusters.
(vi) Use of a combination of spectral and spatial cluster measures to aid cluster merging decisions.
(vii) No interference with the cluster merging process until the large-scale patterns become important (the penultimate step, i.e. the procedure is automatic until that step).
(viii) No need for ground information until the final step (labelling of clusters).

3. CPG procedure

In principle, the CPG procedure is very simple: find means for representative spectral clusters in the data set, assign every pixel to a cluster and combine similar clusters until the remaining clusters can be assigned thematic labels.

The procedure consists of the following steps (figure 1).

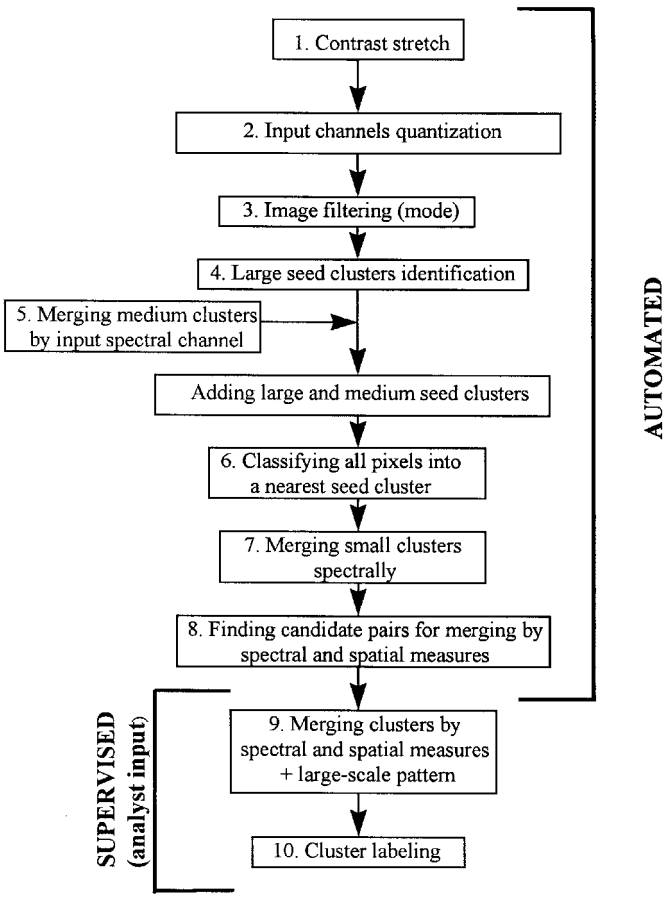


Figure 1. Flowchart of the CPG steps.

Step 1: Contrast stretch. The range of values in each input channel is stretched to encompass the entire range of values permissible (typically 8 bits).

Step 2: Quantization. This step reduces the number of grey levels in each input channel. The number of retained grey levels is as low as possible without losing significant information. In practice, this can be derived by decreasing the number of levels until a visually observable difference appears in a side-by-side comparison with the original, contrast-stretched data. For example, we have found that for boreal environments, ten or eleven levels are sufficient to capture the information contained in Landsat Thematic Mapper (TM) data (Beaubien *et al.* 1997, 1998). As a safety margin, a larger number of levels can be used than is implied by the visual assessment. The number of levels varies with the spectral band and scene type, less with sensor type. After the quantization, each unique combination of the spectral values represents a pure spectral cluster with only one possible digital level in each input channel.

Step 3: Spatial image filtering. The purpose of this step is to identify spatially dominant spectral clusters in the scene, based on the number as well as the spatial arrangement of the pixels from various clusters. In principle, a smaller cluster could be distributed in such a way that it becomes spatially dominant over a larger but more compact one. Mode filtering is used, typically within a 5×5 or 7×7 window. This filter finds the most abundant cluster within the window and places it in the centre pixel.

Step 4: Identification of large seed clusters. Large clusters are considered to be all those containing more than a pre-defined minimum fraction of all pixels (MinLSC = Minimum Large Seed Cluster) to be classified. For example, 0·1% can be used as a conservative value, assuming that no single spectral value could form the basis of a thematic class unless it had at least 0·001 pixels to start with. MinLSC should be smaller than the minimum mapping unit. After Step 3, the size of all clusters is determined. They are then sorted in decreasing size and the large seed clusters are identified as those larger than MinLSC; their number is not otherwise restricted.

Step 5: Merge medium-sized pure clusters. It is possible that the large seed clusters may not contain all the thematically relevant classes (or subclasses). For example, when many different values of pixels (and therefore many clusters of similar size) are present the large clusters might represent only a portion of the spectral space. For this reason, CPG provides for the creation of additional seed clusters, using medium-sized pure clusters as input. Medium-sized clusters are those containing between MaxNC and MinLSC pixels, where MaxNC (= Maximum Neglected Cluster) is the number of pixels in the largest cluster to be neglected in identifying seed clusters (e.g. 0·002% as a conservative value). However, the medium-sized clusters are combined before being considered for candidate seed clusters, i.e. their radiometric resolution is further reduced. The rule is simple: the smallest remaining cluster above MaxNC is added to the closest and largest medium-sized medium cluster which differs by only one quantized level in one spectral dimension. Since this operation is performed iteratively and sequentially for each spectral dimension, the pixels eventually combined in such a seed cluster could differ by as many quantized levels as twice the number of input channels. Note that the order of spectral dimensions in the merging operation matters, so the sequence should start with the channel that, overall, offers the least amount of important information. All pixels merged into one seed cluster are given the same label. Once this process is completed, only combined

clusters which contain more than MinLSC pixels are selected to be seed clusters. Pure clusters smaller than MaxNC are ignored in constructing the seed clusters.

Steps 1–5 yield a set of seed clusters with the quantization and filtering (Steps 2, 3) used as aids in this process. Steps 6–10 are based on the original, full-resolution data.

Step 6: Classification. Using seed clusters from Steps 4 and 5 above, the means are computed for each cluster using the original input channels from Step 1. All pixels in the image are assigned to one of the seed clusters based on the minimum Euclidean distance. This algorithm is used because of its conceptual simplicity and computational speed. Note that the number of resulting clusters could be less than the number of seed clusters, depending on the distribution of the large and medium seed clusters in the spectral space.

Step 7: Merging clusters using spectral similarity. The spectral similarity SS_{ij} is computed between all pairwise combinations of the clusters from Step 6. SS_{ij} is defined as

$$SS_{ij} = \frac{S_{ij} + S_{ji}}{SD_{ij}} \tag{1}$$

$$S_{ij} = \frac{\sum_{k=1}^{n} (\cos_{ijk} S_{ik})}{\sum_{k=1}^{n} \cos_{ijk}} \tag{2}$$

$$\cos_{ijk} = \frac{|M_{ik} - M_{jk}|}{SD_{ij}} \tag{3}$$

$$SD_{ij} = \left(\sum_{k=1}^{n} (M_{ik} - M_{jk})^2 \right)^{1/2} \tag{4}$$

where $i \neq j$ and M = arithmetic cluster mean, S_{ij} = standard deviation of cluster i in the direction of the cluster j centroid, S_{ik} = standard deviation of cluster i in spectral channel k, SD = spectral distance between clusters, cos = cosine of the angle between clusters, i, j = cluster number, k = spectral channel, and n = total number of spectral channels.

SS_{ij} is thus computed as the sum of the standard deviations of the two clusters i, j along the vector (of length SD_{ij}) connecting the centroids of the two clusters (figure 2). It is a measure of the spectral proximity of the two clusters, both in terms of their mean values and at the margins. Equations (1)–(4) are based on the simple concept that clusters which overlap in the spectral space should be combined first, before non-overlapping clusters are considered for merging.

All the clusters from Step 6 are sorted according to decreasing size. Starting with the smallest cluster i, the cluster j which has the lowest SD_{ij} is found. Next, all clusters r with $SD_{ir} \leqslant 1{\cdot}1 SD_{ij}$ are found. Cluster i is then identified to be merged with cluster p provided that $SS_{ip} > SS_{iq}$ for $p, q \in r$. That is, if several clusters have a similar distance in the multispectral space to i, the one spectrally closest overall is merged in preference to those that are more distant. Note that the size of the larger cluster is not changed at this stage. This process continues until the smallest remaining cluster contains MinCM pixels, where MinCM is the Minimum Cluster size for

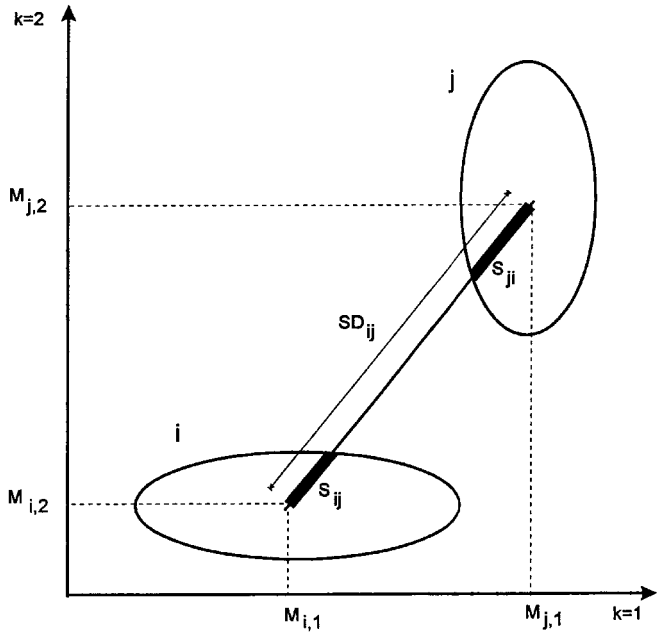


Figure 2. The parameters considered in Step 7 (automatic merging of clusters). The initial choice is based on SD. If several clusters have similar SD values, the standard deviation S of the smaller cluster in the direction of the larger one (and vice versa) are also considered. See text for detailed explanation.

Merging. Then the identified candidates are merged and assigned to the identified clusters p (or j).

Step 8: Identifying candidates for merging using spectral and spatial similarity. The means, standard deviations, SD_{ij} and SS_{ij} are recomputed for clusters resulting from Step 7. In addition, spatial adjacency SA_{ij} is computed as a measure of the intermixing of pixels from clusters i and j. The assumption is made that two clusters that are mutually highly interspersed should be combined in preference to clusters that are spatially distinct or clumped. The spatial adjacency is defined as

$$SA_{ij} = \frac{NA_{ij}}{\mathrm{Min}(NP_i, NP_j)} \tag{5}$$

where NA_{ij} = number of cases where pairs of pixels from clusters i and j are adjacent (including diagonals) in a 3×3 window and NP = number of pixels in the cluster.

NA_{ij} is computed by counting, for each pixel in cluster i, the number of instances when a pixel from cluster j is adjacent (located within a 3×3 window centred on the pixel i). The denominator represents the smaller cluster because the goal of this step is to facilitate merging smaller clusters with the larger ones.

Step 9: Merge clusters using spectral, spatial and pattern similarity. The values of SD_{ij} and SA_{ij} from Step 8, together with the cluster size, provide the information needed to decide on further cluster merging. Although such merging could also be accomplished by automated decision rules, such an approach might produce undesirable results. This is because the distinctions retained in the final classification do not necessarily correspond to the magnitude of the spectral and spatial differences among

the clusters. Therefore, the computer only suggests the clusters to be merged and the decision is left to the analyst as follows.

For clusters from Step 8 the cluster pair i, j which has the lowest SD_{ij} is found. Next, all clusters r with $SD_{ir} \leqslant 1{\cdot}1 SD_{ij}$ are found. Cluster i is then identified to be merged with cluster p provided that $SA_{ip} \geqslant SA_{iq}$ for $p, q \in r$. That is, if several clusters are spectrally similar to i, the one closest spatially overall is merged in preference to those that are spatially separate. The identities of clusters i and j (or p), their sizes, SD_{ij}, and SA_{ij} are provided as inputs to the analyst. This search for candidate pairs for merging is repeated as long as desired, nominally until the number of remaining clusters (not identified for further merging) equals the number of classes in the classification legend.

The pairs of clusters suggested for merging are then reviewed on the computer screen after displaying them in contrasting colours. Besides SD and SA values, the similarity of patterns of the two clusters over the entire classified area is considered. The proposed merge is accepted only if, in the judgement of the analyst, the two patterns are very similar, in both spatial extent and in local pixel density variations. The rationale is that if the two clusters follow each other that closely, they should be combined, even if they are less spectrally similar than clusters combined earlier (Step 7). No knowledge of ground conditions is required for this decision because the decision is based on a visual assessment of the similarity between two patterns. However, it is possible that the analyst may have other independent information which may lead him to override the proposed merging. The primary reason for using visual assessment is that a comparison of the two patterns over larger distances in the image is not feasible with current computer technology, especially for areas of some size and an appreciable number of clusters.

Step 10: Labelling. Knowledge of surface conditions is required to put thematic class labels on the clusters from Step 9. This is no different from using other unsupervised classification methods. Means and standard deviations computed for the final clusters can effectively aid the labelling process, provided spectral characteristics of various expected classes are known. Another very effective aid consists of displaying the resulting clusters using their mean values in selected spectral channels. The selection is based on the ease of visual interpretation or the familiarity of the analyst with that particular presentation of the data.

4. CPG evaluation

We have evaluated the performance of the CPG methodology in several ways:

1. assessment of the Advanced Very High Resolution Radiometer (AVHRR) classification accuracy, using classification derived from Landsat Thematic Mapper (TM) data as the surface 'truth';
2. assessment of the sensitivity of the classification results to the control parameters, including a comparison with the ISOCLASS method;
3. applicability to data from various sources, by using AVHRR and TM input data.

In all the CPG tests below which are based on AVHRR data we used four channels (AVHRR channels 1 (C1) and 2 (C2), normalized difference vegetation index NDVI (N_m) and area under the NDVI curve (N_a)) representing the growing season means (the first three parameters) or totals (N_a). These measures are similar to those used

by DeFries *et al.* (1995) but differ in the definitions, mainly to avoid the periods with snow on the ground. The data were based on 10-day composites assembled over the 1993 growing season. AVHRR channels 1 and 2 were corrected for atmospheric and bidirectional effects, and the NDVI was corrected for solar zenith angle effects. More detail on these corrections is provided by Cihlar *et al.* (1997 b) and Cihlar *et al.* (1997 c). The length of growing season was established using AVHRR channel 4 data corrected for atmospheric and surface emissivity effects, as the period during which surface temperature was above 10° C. Since a water mask was used to eliminate water bodies from the AVHRR classification process (the original composite images contain many cloud-contaminated pixels over water), only land pixels were used in the tests involving AVHRR data.

4.1. *Classification accuracy*

The assessment of the accuracy of the CPG was performed using the 1993 AVHRR data over Ontario and Quebec (NW corner at −77° 58′ 57·5″, 50° 57′ 31·2″; SE corner at −73° 29′ 24·3″, 45° 12′ 34·3″) presented in the Lambert Conformal Conic projection. The control parameters used were MinLSC = 0·1%, MaxNC = 0·002%, MinCM = 0·5%. This procedure resulted in 81 spectral clusters.

The accuracy assessment was carried out using TM-derived classification over the Matagami, Quebec. This classification was prepared using an image from 20 August 1991 (Royer *et al.* 1994); no significant changes in land cover occurred in the area between 1991 and 1993. Supervised classification was used, together with substantial site information obtained from air photographs and limited field observations (including aircraft overflights). The procedure has been used extensively in similar environments in different parts of Quebec and has been shown to produce accurate results from the thematic classes employed (Beaubien 1994).

The AVHRR and TM data sets were co-registered as follows. Using ground control points, the classified TM image was registered to the same map projection as the AVHRR. Next, the AVHRR classification was re-sampled to the TM pixel spacing using nearest neighbour re-sampling algorithm. The 81 AVHRR clusters were labelled on the basis of the highest correspondence with TM-derived classes. This approach avoids the potential optimistic bias in accuracy assessment (Hammond and Verbyla 1996), although conservative bias underestimating the actual accuracy (e.g. due to image mis-registration and errors in reference data (Verbyla and Hammond 1995)) is difficult to avoid because it is embedded in the TM classification. Cihlar *et al.* (1996) provide more information on the classified TM data set used.

Table 1 shows the confusion matrix for the two data sets for all land pixels. It shows that only the most ubiquitous coniferous class was estimated with good accuracy. The accuracy for the remaining classes was quite poor, and the diagonal accuracy DiAc (see equation (7)) was also low, with only 51% of pixels correctly identified relative to the TM classification.

AVHRR classification accuracy is influenced by the mixed land cover in most AVHRR pixels (Penner 1995, Cihlar *et al.* 1996). The accuracy tests were thus repeated by considering only AVHRR pixels with more than a threshold percentage of TM pixels from one cover type. As the AVHRR pixels become more homogeneous the classification accuracy increases correspondingly (table 2). Cihlar *et al.* (1997 a) obtained similar results with CPG for the southern and northern BOREAS study areas, with diagonal accuracies of 84–89% for pixel purities above 80%.

Tables 1 and 2 show that if the AVHRR pixels are relatively homogeneous the

Table 1. Confusion matrix for AVHRR (rows) and TM (columns) for the Matagami test site*.

	TM	TM	TM	TM	TM	TM	TM
AVHRR	Conifers 1	Decid. 2	Mixed 3	Disturb. 4	Fen 5	Water 6	Regen. 7
1	**83·6**	50·3	60·7	17·0	47·4	71·6	24·5
2	0	**0**	0	0	0	0	0
3	1·4	3·9	**7·9**	0·5	1·2	0·9	2·1
4	0·1	0·0	0·1	**4·6**	0·2	0·0	0·8
5	6·7	4·0	4·1	35·9	**30·3**	6·6	13·6
6	0·1	0·2	0·1	0·0	0·1	**10·3**	0·0
7	8·2	41·6	27·1	42·0	20·8	10·6	**59·0**

* Note: the classification accuracies were computed using equations (6) and (7). The original 20 classes were grouped into seven: 1 = coniferous (including density classes >60%, 40–60%, 25–40%); 2 = deciduous (density >40%, with or without coniferous understorey); 3 = mixed (>55% coniferous, >55% deciduous); 4 = disturbed (recent cuts, recent cuts more or less covered by vegetation); 5 = fen (wetland with 10–25% conifers, wetland with <10% conifers, grass-dominated wetland); 6 = water (note that pixel assignment to water in the AVHRR classification was based on the U.S. Department of Commerce (1977) database); 7 = regeneration (various stages after disturbances).

Table 2. CPG accuracy at various pixel purity thresholds for the Matagami test site. *DiAc* and *Khat* were computed using equations (6)–(8).

Purity threshold (%)	*DiAc* (%)	*Khat* (unitless)
None	51·1	0·304
50	56·8	0·371
60	63·9	0·456
80	80·1	0·704

CPG accuracy can be high. The values in tables 1 and 2 are also similar to those achieved for several areas in Canada using AVHRR data and the unsupervised classifier ISOCLASS (Cihlar *et al.* 1996).

4.2. *Sensitivity to control parameters*

Ideally, the final results should be completely independent of the parameters guiding the classification process. The sensitivity of the results to the (usually arbitrarily chosen) control parameter is of critical importance. The sensitivity can be low if the classes are spectrally distinct but in practice many classes overlap and the results can thus depend strongly on the controlling parameters. The requirement is then that these parameters be few and their impact on the final classification results be minimal.

To perform an assessment of the CPG sensitivity we used AVHRR data for an area in Ontario and Quebec (NW corner at $-77°\,58'\,57{\cdot}5''$, $50°\,57'\,31{\cdot}2''$ and SE corner at $-73°\,29'\,24{\cdot}3''$, $45°\,12'\,34{\cdot}3''$). The CPG method was applied with various combinations of the control parameters (Steps 4, 5, 7). The accuracy tests used CPG clusters after Step 7; this was done to eliminate the analyst's influence which enters in Step 9. In all the tests, the spectral clusters resulting from each classification were labelled as belonging to a land cover class specified in a reference classification. The

labelling was based on the largest overlap between the CPG cluster and the land cover class in the reference classification.

An independent classification was used to test the sensitivity of the CPG to control parameters and to allow comparisons of CPG (and of ISOCLASS, as discussed below). The reference classification was produced from a similar AVHRR data set (1993 growing season; channels 1, 2, NDVI for Canada). This classification was prepared using an interactive procedure named the Enhancement-Classification Method (ECM, Beaubien 1994, Beaubien *et al.* 1997, 1998) and labelled based on a visual comparison with about 100 Landsat scenes. When tested against Landsat classification the absolute accuracy of the ECM classification was found similar to that of the CPG (Beaubien *et al.* 1997). It should be noted that the absolute accuracy of the reference classification is less important since only relative changes are of interest here. The accuracies were computed as follows:

$$DiAc(i,i) = \frac{100 \sum_{i=j}^{q} P(test)_i}{P(ECM)_i} \tag{6}$$

where $P(test)_i$ is the number of 1 km pixels in the test classification labelled as class i (i.e. column total) in the CPG or ISOCLASS classification, $P(ECM)_i$ is the number of 1 km pixels in the reference classification labelled as class i, and j to q are the spectral clusters in the test classification labelled as class i.

$DiAc$ thus measures the proportion of ground cover that was 'positively identified' as that cover in the test classification, assuming that the reference classification is correct. The overall classification accuracy for a given data set is then computed as

$$DiAc = \frac{100 \sum_{i=j}^{n} P(i,i)}{NP} \tag{7}$$

where $P(i,i)$ is the diagonal entry in the confusion matrix (number of pixels) and NP is the total number of pixels in the matrix. $DiAc$ is thus the most stringent measure, counting only positive matches as correctly classified.

The Khat distance was also computed for each confusion matrix (Congalton 1991), as follows:

$$Khat = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_i \, x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \, x_{+i})} \tag{8}$$

where x_{ii} is the total number of pixels in row i, column i, N is the total number of pixels, r is the number of rows (columns) and x_{i+} and x_{+i} are row and column totals respectively.

As seen from equation (8) $Khat$ measures the dispersion outside of the diagonal in the confusion matrix in relation to the concentration along the diagonal axis. For a perfect classification, $Khat = 1{\cdot}0$.

Table 3 shows the results of the tests for various combinations of control parameters. $CL\,COR$ is defined as the proportion of CPG clusters in which $>67\%$ of the pixels corresponded to only one reference class. $PICOR$ represents the percentage of the pixels in $CL\,COR$ clusters relative to all land pixels in the image. Considering

Table 3. Sensitivity of the CPG to control parameters.

CPG test	Parameters*	*DiAc* (%)	*Khat*	CLCOR (%)	*PICOR* (>67%)
1	0·1/0·1	76·6	0·743	68·9	57·7
2	0·05/0·1	78·9	0·768	75·2	64·2
3	0·05/0·5	76·8	0·746	71·7	58·7
4	0·1/0·5	72·6	0·700	61·7	51·7
5	0·2/0·5	71·7	0·689	60·3	48·7
6	0·4/0·5	65·1	0·613	45·4	29·1

* Parameters: values *x*/*y* refer to threshold values (in per cent of the number of land pixels) of MinLSC used in Steps 4/5 and Step 7 of the CPG. MaxNC was 0·002% in all cases. *DiAc* and *Khat* were computed using equations (6)–(8).

CL COR indicates the proportion of CPG clusters in which >67% of the pixels corresponded to only one reference cover type.

PICOR is the percentage of pixels in *CL COR* clusters relative to all land pixels.

the MinLSC = 0·1% combination as the baseline (test 1), it is evident that changes in control parameters had relatively small effect on the results. Changing the seed cluster threshold (Steps 4, 5) by a factor of four (tests 2, 5, 6) changed the accuracy by <5% absolute and that *Khat* by <0·08, and much smaller changes occurred for a narrower range (tests 1, 2 and 4, 5). Changes in MaxNC had a minimal effect on *DiAc* and *Khat* (compare tests 2 and 3). *CL COR* and *PICOR* generally mirrored *DiAc* and *Khat* but were more pronounced. The proportion of well matched clusters changed by up to 15·2% and *PICOR* by up to 15·5% due to variations in the seed cluster thresholds. Step 7 threshold had somewhat stronger effect here as well, a fivefold change leading to a change in *CL COR* (*PICOR*) of 3·5% (5·5%). Considering the extreme combinations (tests 2 versus 5), the parameters changed by a factor of 40 while the *DiAc* decreased by 17·5%. Overall, it is evident that CPG is able to tolerate an appreciable range of control variables.

The sensitivity of CPG was also assessed in comparison to that of ISOCLASS (Tou and Gonzales 1974), a commonly used unsupervised classification algorithm. The same Ontario–Quebec data as for the above CPG tests (table 1) were used. ISOCLASS requires five parameters to be specified (Loveland *et al.* 1991), and only some combinations were tested here (table 4). Tests 1–4 are variations on the combination employed by Cihlar *et al.* (1996), while test 5 is the combination used by

Table 4. Sensitivity of ISOCLASS to control parameters.

ISOCLASS test	Parameters*	*DiAc* (%)	*Khat*	Larger *CL COR* (%)	All *CL COR* (%)	*PICOR* (>67%) (%)
1	81/12/3·2/4·5/250	72·2	0·695	56·3	58·8	48·9
2	81/12/1·6/4·5/250	72·2	0·695	54·3	59·3	48·9
3	81/12/3·2/2·3/250	69·2	0·661	45	46·3	40·3
4	40/12/3·2/4·5/250	60·9	0·653	40	45	32·5
5	70/15/2·6/3·5/30	67·2	0·640	44·3	52·9	40·8

* Values *a*/*b*/*c*/*d*/*e* represent: *a* = maximum number of clusters; *b* = maximum number of iterations; *c* = cluster-combining distance; *d* = cluster splitting distance; *e* = minimum number of pixels in a cluster.

Other measurements are as described in table 3.

Loveland *et al.* (1991) for classifying NDVI monthly composites over North America. An additional accuracy measure has been introduced to reflect the fact that some final clusters had very few pixels and were therefore clearly noise etc.; thus the distinction between 'larger *CL COR*' and 'all *CL COR*' clusters. In all cases, the results have been compared with the reference classification, using the same approach as for the CPG.

Results of the comparisons show that in most cases, the performance of ISOCLASS was significantly more sensitive to the degree of change of the control parameters than the CPG. The results were not sensitive to changes in the ISOCLASS cluster combining distance but responded substantially to reduced cluster splitting distance (tests 1 versus 3). They were highly sensitive to the maximum number of clusters (tests 1 versus 4) when all the accuracy measures decreased. Comparing the various tests to test 1, it is evident that change in the control parameters by a factor of two resulted in reducing the ISOCLASS accuracy by 0–15·7%, depending on the parameter involved.

4.3. *Robustness*

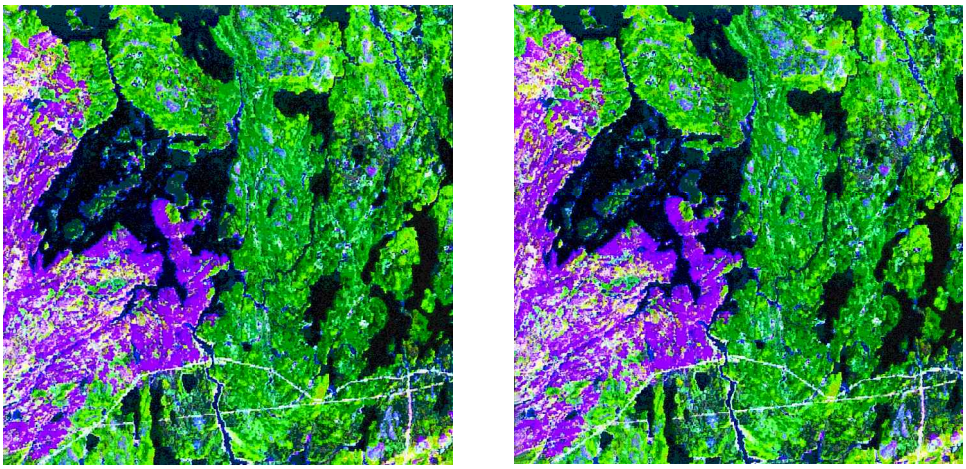
The CPG was initially developed for large-area applications, with AVHRR data as the primary data source. However, the principles should apply to other data as well. A detailed assessment of the CPG performance should be carried out using the reference data set with 100% accuracy which does not suffer from the mixed-pixels problem in the way AVHRR data do. High-resolution satellite image data would thus be appropriate for this purpose. Unfortunately most classifications of such data are erroneous to various degrees, and cannot thus be used as a rigorous standard. Consequently, we are left with an imperfect but plausible test—does the classified image retain most, or all, of the information visible in the input data?

A trial classification was therefore carried out for TM data of part of the BOREAS Northern Study Area (scene centre at $-99°\,34'\,8{\cdot}8''$, $55°\,53'\,35{\cdot}5''$ with 58·5 km in EW and 49·5 km in SN directions). TM channels 3, 4 and 5 were employed, and no radiometric corrections were made to the data. As before, the procedure was carried out automatically, with the following control variables: MinLSC = 0·3%, MaxNC = 0·002%. This reduced the 2679 clusters (after Step 2) to 71. The TM image was re-created from the CPG classification by replacing each of the 71 clusters with the cluster mean in each spectral channel. The re-created three-channel image was then visually compared with the original TM image (prior to quantization). Only one substantial difference could be found between the two images, namely bright bare rock and gravel road (covered with locally dug material) were combined into one cluster. Figure 3 shows a part of the two images in which a noticeable difference occurred along the road. The small difference between the original and the classified TM data accompanied by a reduction of the number of clusters from 2679 to 71 shows that the CPG is an effective classification procedure.

5. Discussion

5.1. *Comparison with supervised and unsupervised classifications*

Compared to supervised classification, the CPG has the advantages of an unsupervised classification, i.e. it does not require knowledge of surface conditions and thematic classes involved until the labelling step (Step 10). This is important for classifications over areas greater than the analyst's personal knowledge encompasses. Because of the potentially changing signatures with distance, the analyst would need



(*a*) (*b*)

Figure 3. Comparison of the original and classified TM data for part of the BOREAS Northern Study Area west of Thompson, Manitoba, Canada. (*a*) Original TM image (band 5 = red, 4 = green, 3 = blue). (*b*) Classified image into 71 clusters, each cluster displayed using mean values derived from the original image.

to know the distribution of all classes across most of the area of interest, at least as samples for each class. When the area is sufficiently small, this knowledge can be obtained with confidence but the task becomes difficult for larger (e.g. >100 km) or unfamiliar areas, and impossible at the continental scale. Importantly, with a fairly low sensitivity to only three control parameters, CPG is much less subject to the analyst's influence than supervised classification.

In comparison to the various unsupervised classification algorithms, CPG does not require prior knowledge of the number of distinct spectral clusters (which is typically higher than the number of desired thematic classes). This is a fundamental advantage because the number of distinct clusters is very difficult to know for a specific data set and because classification results are sensitive to it (table 4). When clouds are present, some of the distinct clusters are mixtures of surface and cloud signatures and the number of clusters may thus be highly variable, even for the same area. The choice of the number of clusters for a particular unsupervised classification task thus becomes arbitrary, although statistically optimal numbers can be defined through trial runs (e.g. Masselli *et al.* 1996). A high number of clusters is not necessarily advantageous (Loveland 1996, personal communication) because the smaller clusters may be thematically meaningless. CPG avoids these problems by finding the maximum number of natural clusters in the initial data set.

Since the CPG needs fewer control parameters than a typical unsupervised classifier, it introduces less analyst bias. On the other hand, it allows the analyst to influence the classification process in the last stage when small spectral or spatial differences may be thematically significant. A comparison with ISOCLASS (tables 3 and 4) shows the CPG accuracies to be higher and overall less sensitive to variations in the control parameters, consistently with the above expectations. The various tests, albeit limited, thus indicate the overall accuracies achieved are comparable or superior to other present unsupervised classification methods.

While unsupervised classifiers typically employ only spectral information, CPG

uses a combination of spectral and spatial measures. This is preferable because not all the relevant information is in the spectral domain. It is also consistent with the rich history of photointerpretation where spatial clues are very important (Rabben 1960). The same logic applies to involving the analyst in the last cluster merging phase (Step 9). Strictly speaking, the results of Step 8 (suggested merging of clusters) could be applied automatically. However, the present measures of spatial adjacency are inadequate to capture the similarity of patterns at various spatial scales. The spatial adjacency (equation (5)) quantifies the relation only for the immediate neighbourhood of the cluster. Although a larger window could be used, it would remain a fixed-size one while the differences in spatial patterns can occur at different spatial scales for various clusters. At present, the human eye–brain combination is most efficient at evaluating these differences. The subjectivity of the analyst's influence is minimized in CPG by accepting or rejecting choices for pairwise grouping proposed by the algorithm. The recommendation of the algorithm can be overridden by the analyst but presumably only in cases where it improves the final result.

Unlike some unsupervised classifiers, CPG is not recursive. This is an advantage because at every step, a qualitatively different process takes place and its results can be analysed and assessed. Furthermore, different characteristics of the data space are exploited so that the maximum possible information is extracted before the subjective element enters the process. While a 'safe' number of iterations in a recursive classification can be readily defined as 'many', it is more difficult to specify a minimum required number of iterations; usually, preliminary data runs are needed to find it for a particular data set. CPG avoids the need to pre-determine the number of iterations. At the same time, it can be implemented as a continuous process in which the three parameters are specified at the outset and the program halts at the end of Step 8.

5.2. *CPG performance*

Comparisons of CPG results for AVHRR and TM data sets described above indicate that the method effectively reduces the dimensionality of the original data, yet retains the significant spectral information. The tests indicate that the CPG procedure meets many of the requirements for an ideal classification algorithm stated at the outset, provided that the desired classes can be identified primarily using radiometric information. The CPG result resembles the original image, especially when the classes are colour coded appropriately (e.g. figure 3 (*b*)). They do not have a salt-and-pepper appearance, indicating noise in the classification, as results of unsupervised classifications tend to have. This is because, by design, the classes are based on generalizations of the digital levels in the original images to retain most of the original pixel information.

Ideally, a classification scheme would be completely independent of the analyst's influence and would yield results with 100% accuracy. These are fairly contradictory requirements and thus to achieve high accuracy, some guidance to the classifier is clearly necessary. CPG requires only one type of information at the various stages preceding the labelling step, namely the smallest spectral cluster that should be ignored. This is probably the minimum that must be specified in a digital classification process, otherwise the classification is not likely to produce practical results. It is also consistent with the minimum mapping unit concept.

Since CPG is designed to detect the larger spectral clusters it may suppress small but significant ones. For example, urban areas (AVHRR) or roads (TM) may disappear in the clustering process. Similarly, classes of interest that are spectrally similar

to others may be grouped. If any of these are important, they can be identified through post-classification steps, e.g. by splitting clusters or using supervised classifier (applied to the original data) for these classes.

The algorithm used in the quantization process (Step 2) influences the results of the clustering, especially for the large seed clusters. Various options exist, e.g. quantization of the original image, quantization of an equalized image, etc. For example, equalized quantization provides higher resolution in spectrally similar areas, while 'direct' quantization is preferable when small clusters (in the wings of the histograms) are likely to be important. The decision regarding the type of quantization procedure can be made from these considerations, from prior experience, or based on the absence of a significant difference between the original and the quantized image. The fact that the original multispectral image can be reduced to relatively few digital levels without visible loss of land cover type information suggests that the more detailed data represent differences within land cover types, not among them. This and contextual information appear to allow interpreters to classify successfully land cover from hard copy images (e.g. de Boissezon *et al.* 1993), even though the radiometric content is reduced compared to the digital data.

The merging of medium sized clusters (Step 5) is intended to ensure that smaller but potentially significant clusters are not omitted from the classification; the large clusters will be selected in any case. Since it affects only the selection of the smallest seed clusters there is a relationship between MinLSC threshold and these two steps. In particular, if MinLSC is sufficiently low Step 5 may not be required. The need for it can be readily ascertained by comparing the means of large and medium seed clusters. For example, we found for a 4-channel AVHRR data set of all Canada that the large seed clusters with MinLSC = 0·1% spanned the entire spectral space. It should be noted that spatial image filtering (Step 3) also affects only the small seed clusters but its effectiveness is probably less predictable as it will vary with the heterogeneity of the scene being classified.

The computer implementation of CPG consists of various image processing and set ordering/searching operations. When the number of initial clusters is high, the searches can be time consuming. The number of initial clusters depends on the number of input channels and the number of digital levels in each channel after quantization (Step 2). The overall process thus is more computer-demanding than for other algorithms such as supervised classification. This could be reduced by optimizing the software used, and it should become of less concern as computing speeds increase. For the same reason CPG would not be a practical classification approach for imaging spectrometer data, unless a dimensionality reduction scheme (such as principal component analysis) is applied first. The computational demand may not be a big disadvantage since the number of independent channels in remote sensing data of natural scenes tends to be limited, even for hyperspectral data (Price 1994). Conversely, CPG is less sensitive to the spatial resolution/area size of the input data set. This is because most of the operations handle clusters, and the pixel classification algorithm itself (Step 6) is fast.

Since CPG assigns each pixel to one cover class it is a 'hard' classifier. Various operations are possible that improve the final result of an unsupervised classification, e.g. an adjustment of the map legend in relation to the identified spectral classes (Lark 1995), fuzzy techniques (Foody 1996), contextual correction (Groom *et al.* 1996), additional cluster splitting based on ancillary information (Loveland *et al.* 1991), etc. These could be applied to the CPG results as well, provided that the requisite ground data are available.

CPG appears to have considerable capacity for data compression. For example, in the TM classification for NSA the number of spectral clusters was reduced by 97·5% (from ≫2679 to 71), yet the vast majority of the visible spectral information was retained (figure 3).

6. Summary and conclusions

Land cover type classification is arguably the most widespread application of remote sensing data for terrestrial environments. Although many different classification approaches have been developed to date, they have substantial disadvantages in practical use. Specific drawbacks vary with the methodology but may include one or more of the following: prior knowledge of the number of distinct spectral clusters; prior knowledge of the approximate spatial distribution and the spectral variability of the land cover classes of interest; knowledge of the statistical properties of spectral clusters and their distribution in the multi-dimensional spectral space; prior specification of parameters controlling the classification process; and others. Most of these are not known exactly, and some are unknowable in practice. Thus, in many cases where existing supervised or unsupervised classifiers are used, digital land cover classification becomes a trial-and-error procedure with several runs and modifications of the intermediate results.

In this study, we have attempted to develop a methodology that would represent the spectral information content of the data accurately, would facilitate obtaining an accurate classification, would not require knowledge of the data distribution in the spectral space, and could run automatically with minimum input from the analyst. The resulting procedure, Classification by Progressive Generalization (CPG), meets these criteria. CPG is based on the identification of all 'significant' clusters in the data, followed by classification of all pixels and a guided merging of the resulting clusters. It can be run automatically until the desired number of classes is reached. However, it gives an analyst the option of optimizing the results by introducing independent information in the last stages when the merging of large clusters (based on spectral similarity) might be undesirable. It does not require knowledge of spectral space such as the number and proximity of clusters. It needs only three parameters to be specified *a priori*, all of which refer to the size of clusters that can be ignored at a given stage of the analysis and are therefore relatively easy to estimate by a user of the classification product.

Several tests have been carried out to evaluate the CPG accuracy (absolute and in comparison with ISOCLASS, a typical unsupervised classifier), sensitivity to control parameters, and usefulness with various input data. Results of these tests can be summarized as follows.

- The accuracy achievable by the CPG is comparable or superior to other existing unsupervised classification methods (represented by ISOCLASS).
- The CPG is less sensitive to control parameters that is ISOCLASS.
- CPG performs well with AVHRR and TM data.

It is concluded that the classification strategy embodied in CPG represents a viable and effective approach to the digital classification of remote sensing data.

Acknowledgments

We wish to acknowledge the contribution of Ms Fengting Huang to the processing of AVHRR data used in this paper. Drs Zhanqing Li and Frank Ahern from the

Canada Centre for Remote Sensing provided helpful comments on a draft of this paper.

References

Ball, G. H., and Hall, D. J., 1965. ISODATA, a novel method of data analysis and patter classification. Technical Report, Stanford Research Institute, Menlo Park, California, USA.

Beaubien, J., 1994, Landsat TM satellite images of forests: from enhancement to classification. *Canadian Journal of Remote Sensing*, **20,** 17–26.

Beaubien, J., Cihlar, J., Simard, G., and Xiao, Q., 1998, Land cover from thematic mapper data using new enhancement—classification methodology. In preparation.

Beaubien, J., Cihlar, J., Xiao, Q., Chen, J., Fung, K., and Hurlburt, P., 1997, A new, nationally consistent, satellite-derived land cover of Canada: a comparison of two methodologies. *Proceedings of the International Symposium Geomatics in the Era of Radarsat, Ottawa, Ontario, May 25–30* (published on CD-ROM).

Bryant, J., 1978, Applications of clustering in multi-image data analysis. NASA Report no. 18, Contract NAS-9-14689-8S, NASA Johnson Space Center, Houston, Texas, USA.

Cihlar, J., Beaubien, J., Xiao, Q., Chen, J., and Li, Z., 1997 a, Land cover of the BOREAS region from AVHRR and Landsat data. *Canadian Journal of Remote Sensing*, **23,** 163–175.

Cihlar, J., Chen, J., and Li, Z., 1997 b, Seasonal AVHRR multichannel data sets and products for scaling up biospheric processes. *Journal of Geophysical Research—Atmospheres*, **102,** 29625–29640.

Cihlar, J., Ly, H., Li, Z., Chen, J., Pokrant, H., and Huang, F., 1997 c, Multitemporal, multichannel data sets for land biosphere studies—artifacts and corrections. *Remote Sensing of Environment*, **60,** 35–57.

Cihlar, J., Ly, H., and Xiao, Q., 1996, Land cover classification with AVHRR multichannel composites in northern environments. *Remote Sensing of Environment*, **58,** 36–51.

Cihlar, J., and Protz, R., 1972, Perception of tome differences from film transparencies. *Photogrammetria*, **8,** 131–140.

Congalton, R. G., 1991, A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37,** 35–41.

De Boissezon, H., Gonzales, G., Pous, B., and Sharman, M., 1993, Rapid estimates of crop acreage and production as a European scale using high resolution imagery—operational review. *Proceedings of the International Symposium of Operationalization of Remote Sensing, Entschede, The Netherlands, April 19–23, 1993*, pp. 94–105.

DeFries, R., Hansen, M., and Townshend, J., 1995, Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. *Remote Sensing of Environment*, **54,** 209–222.

Duda, R. O., and Hart, P. E., 1973, *Pattern Classification and Scene Analysis* (New York: Wiley).

Foody, G. M., 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data. *International Journal of Remote Sensing*, **17,** 1317–1340.

Groom, G. B., Fuller, R. M., and Jones, A. R., 1996, Contextual correction: techniques for improving land cover mapping from remotely sensed images. *International Journal of Remote Sensing*, **17,** 69–89.

Hammond, T. O., and Verbyla, D. L., 1996, Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, **17,** 1261–1266.

Hansen, M., Dubayah, R., and DeFries, R., 1996, Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, **17,** 1075–1081.

Loveland, T. R., Merchant, J. W., Ohlen, D. O., and Brown, J. F., 1991, Development of a land-cover characteristics database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, **57,** 1453–1463.

Masselli, F., Petkov, L., Maracchi, G., and Conese, C., 1996, Eco-climatic classification of Tuscany through NOAA-AVHRR data. *International Journal of Remote Sensing*, **17,** 2369–2384.

Price, J. C., 1994, How unique are spectral signatures? *Remote Sensing of Environment*, **49,** 181–186.

Rabben, E. L., 1960, Fundamentals of photo interpretation. In *Manual of Photographic Interpretation*, edited by R. N. Colwell (Washington, DC: The American Society of Photogrammetry), pp. 99–168.

Royer, A., Anseau, C., Viau, A., Theriault, A., Bonn, F., O'Neill, N., Granberg, H. B., Singh, B., Schuepp, P., Beaubien, J., and Cihlar, J., 1994, Impact des changements de l'environnement global sur la foret boreale du Quebec. In: *Teledetection de l'Environnement dans l'Espace Francophone* (Sainte-Foy, Quebec: Presses de l'universite du Quebec) pp. 305–330.

Tou, J. T., and Gonzales, R. C., 1974, *Pattern Recognition Principles* (Reading, MA: Addison-Wesley).

Townshend, J. R. G., Justice, C. O., Skole, D., Malingreau, J.-P., Cihlar, J., Teillet, P., Sadowski, F., and Rutenberg, S., 1994, The 1 km resolution global data set: needs of the International Geosphere Biosphere Programme. *International Journal of Remote Sensing*, **15,** 3417–3441.

U.S. Department of Commerce, 1977, *World Data Bank II, Volume 1—North America and General User's Guide* (Springfield, VA: National Technical Information Service).

VanderZee, D., and Ehrlich, D., 1995, Sensitivity of ISODATA to changes in sampling procedures and processing parameters when applied to AVHRR time-series NDVI data. *International Journal of Remote Sensing*, **16,** 673–686.

Verbyla, D. L., and Hammond, T. O., 1995, Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing*, **16,** 581–587.