COMBINING DIVERSE SPECTRAL, SPATIAL AND CONTEXTUAL ATTRIBUTES IN SEGMENT-BASED IMAGE CLASSIFICATION

B. Guindon
Canada Centre for Remote Sensing
588 Booth Street
Ottawa, CANADA, K1A 0Y7
Email: bert.guindon@ccrs.nrcan.gc.ca

ABSTRACT

Conventional per-pixel classifiers will have limited effectiveness in the interpretation of very high resolution satellite imagery. More flexible approaches will be needed which rely on a broader set of scene attributes, especially spatial and contextual ones. This raises issues related to methods for combining such diverse attributes in a consistent decision-making framework. Evidential reasoning provides an alternative approach. A methodology, involving segment-based labelling is described. The implications for attribute selection, integration into complex recognition modules and the derivation of evidential weights from training are discussed.

BACKGROUND

The availability of high resolution (1-3 m) commercial satellite imagery presents the civilian remote sensing community with new challenges in the area of information extraction. First, the highest spatial resolutions will be achieved by having the sensors operating in monochrome or low spectral dimensionality modes. This means that a greater emphasis must be placed on exploiting spatial and contextual attributes than has been the case with, for example, Landsat TM images. Second, because of the wealth of intra-object detail in this new imagery, higher level image primitives (e.g. lines, regions, etc.) will be more useful in interpretation than pixels and these will be best exploited within the frameworks of goal-directed, image understanding (IU) recognition strategies.

For the past 3 years, a program has been underway to develop and assess automated recognition methods for high resolution data (Guindon, 2000). Our approach involves the creation and processing of segmented renditions of monochrome images. For each segment, both spectral and spatial attributes are derived as well as a spatial organization table which links each segment to its adjacent (first order) neighbours. A goal-directed strategy is used to identify and label a subset of segments belonging to object classes of interest (e.g. roads, buildings, etc.). This involves the application of recognition modules consisting of combinations of attribute tests couched in the form of rules. Each module is tailored to a specific object class and therefore will only encompass those attributes most suitable for distinguishing segment members of that class from all other segments/classes. Each attribute within a module will have an associated evidential weight that quantifies its relative discrimination effectiveness. When a segment is examined, each attribute condition which is met adds additional evidence to support the hypothesis of class membership. If the accumulated evidence exceeds a threshold, the segment is assigned the relevant class label.

From a procedural point of view, goal-directed labelling differs significantly from conventional per-pixel classifiers such as maximum likelihood (hereafter referred to as MLHD). MLHD is a single pass operation where the probabilities of membership of all possible classes are computed for each pixel and where a decision regarding the class membership of each pixel is reached independent of all other pixels (i.e. in a context-free manner). The major features of a goal-directed approach, on the other hand, are quite different and involve sequential and iterative execution of a series of class-specific recognition modules leading to the gradual creation of a scene interpretation. In addition, since only a few object classes may be of interest, a 'sparse' interpretation as opposed to a 'wall-to-wall' classification may be sought. The need for iterative processing arises from the use of semantic context (e.g. a segment may only be recognizable as being part of a road once nearby buildings have been identified). The other major use of context is to direct searches for objects in restricted regions of an image thereby limiting testing to only a subset of the full segment population.

The objective of this paper is to summarize some of the principal issues and underlying assumptions related to the exploitation of combinations of diverse attributes within the framework of a goal-driven segment labelling strategy. To aid in the discussion, we compare and contrast our evidence accumulation labelling (EAL) approach (see Guindon, 2000) with MLHD. The issues considered include attribute (band) selection and combination, training and decision rule definition.

ATTRIBUTE SELECTION AND COMBINATION

(a) MLHD

In conventional MLHD classification of multi-spectral images the principal attributes of a pixel are its radiometric responses (grey levels) in various spectral bands. These 'spectral attributes' form the components of an overall spectral vector. Implicit to this assumption that the spectral axes are orthogonal and that these bands have some measure of equivalence. If spatial (texture) information is to be exploited as well, then it is most convenient to encode texture measures in the form of 'pseudo-bands' that can form additional dimensions in an overall spectral-spatial space (e.g. Haralick, 1979). Once again one must assume that there is dimensional equivalence of all attributes and that compliance with an analytical form of the probability density function (PDF), typically a multi-variate normality condition, still applies. The author is unaware of any studies that have addressed the validity of these assumptions.

Another key feature of MLHD classification is the requirement that the same attribute set (i.e. vector space) be used to evaluate each class membership probability. This presents a number of challenges when selecting a suitable attribute set. In the interest of computational efficiency and improved classification performance, it is desireable to limit vector dimensionality in order to eliminate redundant attributes and attributes with low class discrimination power that may lead to enhanced inter-class confusion. The selection criteria are generally based on maximization of inter-class divergence measures as viewed in this vector space. These measures take into account inter-class grey level differences and intra-class attribute covariance.

(b) EAL

In a diverse attribute regime involving spatial and contextual measures that cannot be expressed in a ‘pseudo-band’ form, there are no rational arguments to support an a priori analytical PDF form for individual attributes let alone for ones in combination. In addition, many attributes will only have a meaningful truncated range of values for a given class and hence the concept of a continuum of probability of occurrence over a broad range is inappropriate. For example, roads may exhibit a range in widths and therefore one can compute some dispersion measure to characterize that range (e.g. standard deviation) but using this in a continuum sense to compute probabilities of finding 1 km wide roads is meaningless. The acceptable value range of an attribute such as width then must be set by physical arguments or by the observed range of values of training example, i.e. in a manner independent of other object classes in the scene. This differs from MHLD where decision boundaries are governed by relative inter-class probability levels.

A second important fact is that not all attributes are applicable to all classes of objects. For example, ‘width’ may have relevance to roads but will be of limited use in identifying agricultural fields and be of no use in recognizing forested areas. This implies that the recognition module of each object class should consist of a unique set of meaningful attributes.

Third, one must differentiate between two types of attributes in a recognition module, namely, those will be characteristic of every segment member (hereafter referred to as requisite attributes) and those which are characteristic of the class but will not necessarily be exhibited by every member (corroborative attributes). Taking the examples of roads again, width and measures of logical connectivity within an overall network may be considered requisites for a new road segment candidate while radiometric similarity to nearby road segments may be considered supporting or corroborative evidence. It is only necessary to estimate the evidential weight of the latter attribute category since requisite attributes, by definition, have a statistical certainty of 1 and act as a screen to limit the candidate segment pool which must be examined. EAL is based on the premise that the more corroboration found the higher the confidence in the class label hypothesis.

Finally, one has to deal with inter-attribute correlation or covariance. Most evidential reasoning models assume statistical independence (e.g. Haralick and Shapiro, 1992). Because EAL sums evidence from attributes in a scalar sense rather than a vector one, it is difficult to utilize correlation. On the other hand, it can be estimated using statistical binomial theory (Guindon, 2000).

TRAINING

(a) MHLD

There are two key features of training per-pixel classifiers such as MHLD. First, it is most convenient to acquire training samples in sets of contiguous blocks rather than in a truly random way. Implicit in this approach is that inter-sample spatial auto-correlation is minimal. Second, since typical images contain hundreds of thousands or millions of pixels, it is feasible to capture thousands of training samples at least for the major thematic classes of interest. This means that the statistical uncertainty in derived parametric measures will be small.

(b) EAL

Segment-based labelling faces a number of different challenges. First, spatial/contextual attributes are useful because they exploit an underlying spatial organization of features in a scene (e.g. road segments are not randomly distributed but form a linked network). This organization cannot be captured through random sampling. Instead, 'wall-to-wall' training acquisition must be used in restricted training sub-scenes. Second, the number of segments in a scene will be much smaller than the number of pixels. In our studies (Guindon, 2000), a typical residential sub-scene at 2 metre resolution covering 256X256 pixels will typically be partitioned into about 2000 segments of which only 20-30 may be road segments. This implies that the number of training segments that can be practically captured will be small and hence the limitations of small number statistics must be taken into account when computing evidential weights and inter-attribute correlation.

DECISION RULE DEFINITION

(a) MHLD

The training data define the mean class vectors and the covariance matrices for each class. From this information, one can calculate, for any pixel vector, the probability of membership in each class and therefore select the most probable class. The key points are that the class probabilities can be assessed simultaneously and need only be computed once since a decision regarding class membership for a pixel is independent of its surroundings.

(b) EAL

Evidential accumulation relies on a completely different approach.

(1) Because the processing is goal-driven, one addresses the probability of membership in one class only at a time, i.e. the question being dealt with is of the form 'is there sufficient evidence to support the hypothesis that segment A is a member of object class C?". In this sense all non-class C segments are grouped together in a single population. Thus the evidential weight assigned to an attribute must be related to the ratio of the probability that a class C segment meets the attribute test condition divided by the probability that a segment, not of class C meets the attribute test condition. It should be noted that this approach differs from probabilistic inexact reasoning methods such as Dempster-Shafer and diagnostic expert systems such as MYCIN (see, for example, reviews in Haralick and Shapiro, 1992 and Rich, 1983). In these cases the numerical measure of belief (evidence) is associated with the absolute probability that a hypothesis is true given that the attribute condition is

observed. In our case, evidential weight measures the relative merit of an attribute within a specified attribute group.

(2) Since most recognition modules will contain a number of corroborative attribute tests, a final class membership decision must be based on an evidence accumulation threshold. This evidence threshold can be estimated by summing the ratios described above, each weighted by the absolute probability that any segment of class C will meet the attribute test condition.

(3) In the above sense, evidential weights can be considered measures of belief since they are pro-active measures of identification. On the other hand, EAL is an iterative procedure and one must allow for the possibility that a label assigned to a segment in an early iteration may be found to be incorrect in later stages when a more complete (contextual) interpretation become available. Continual consistency checking forms an integral part of EAL. Fortunately the concept of measures of disbelief, expressed as negative evidential weights, is readily supported and can be used to eliminate commission errors. A ‘stable’ interpretation is reached when both measures of belief and disbelief do not change the label status of any segment during an iteration.

CONCLUSIONS

The low spectral dimensionality of imagery acquired by very high resolution satellites will require that a greater reliance be placed on spatial/contextual attributes to achieve an optimum scene interpretation. The ‘vector’ attribute representation, currently in use in per-pixel classification is ill-suited to diverse sets of attributes which go beyond spectral and pseudo-spectral features. An alternate approach is to use goal-directed evidential reasoning. The underlying issues and assumptions of one such approach, EAL, based on segment labelling, have been discussed through a comparison with conventional maximum likelihood classification. Although EAL is more complex, involving iterative processing and on-going consistency checking, evidential measures including inter-attribute correlation can be derived from properly acquired training data.

REFERENCES

Guindon, B., 2000, “’A Framework for the Development and Assessment of Object Recognition Modules for High Resolution Satellite Images’, *Canadian Journal of Remote Sensing*, (in press).

Haralick, R.M., 1979, ‘Statistical and Structural Approaches to Texture’, *Proceedings of the IEEE*, Vol. 67, pp. 786-804.

Haralick, R.M. and L.G. Shapiro, *Compter and Robot Vision*, Addison-Wesley Publishing Company, Chapter 19.

Rich, E., 1983, *Artificial Intelligence*, McGraw-Hill Book Company, Chapter 6.