COMPUTER-BASED IMAGE UNDERSTANDING AT UBC

Alan K. Mackworth

Department of Computer Science
University of British Columbia
Vancouver, B.C., V6T 1W5

Abstract

An outline of some of the activities in computer-based image understanding in the Department of Computer Science at the University of British Columbia is given. The introduction focusses on paradigms for image analysis and discusses the development of image understanding within artificial intelligence. Techniques from that area have been combined with modifications of traditional pattern recognition techniques to enhance the spectral, spatial and semantic sensitivity of Landsat image segmentation programs. Two programs, one using interpretation-guided region merging and the other using image hierarchies, are briefly presented. Further work on understanding sketch maps and the like shows how to make explicit and then exploit, the cartographic and geographic semantics of simple maps, and, moreover, contributes directly to the new computational theory of perception.

Résumé

Un exposé des activités en compréhension des images par ordinateur du département de Computer Science à l'université de Colombie Britannique est présenté. L'introduction se concentre sur les paradigmes pour l'analyse des images et considère le développement de ce sujet au sein de l'Intelligence Artificielle. Des méthodes de cette discipline sont combinées avec certaines modifications de techniques employées dans la reconnaissance des formes pour augmenter la sensitivité spectrale, spatiale et sémantique des programmes de segmentation des images de Landsat. Deux programmes, l'un pour la fusion de régions guidé par interprétation et le deuxième se servant d'hiérarchies d'images, sont brièvement présentés. Des travaux supplémentaires sur la compréhension des cartes-croquis montrent comment rendre explicite, et ensuite exploiter, la sémantique cartographique et géographique de simples cartes, et contribuent, de plus, directement à la nouvelle théorie computationelle de la perception.

Introduction

The purpose of this paper is to provide a brief overview of some of the activities of the image understanding research group in the Department of Computer Science at the University of British Columbia. Particular emphasis is placed on our work in the understanding of Landsat images.

Computer-based Image Analysis

The science of image analysis by computer is, at the moment, in that state of turbulent, almost violent, flux characteristic of a science in the process of growing beyond the metaphors of its ruling paradigm. The predominant paradigm for image analysis really has two aspects: image processing and pattern recognition. Image processing is concerned with the image *qua* image: to eliminate noise or distortion introduced by the image capture process, to code images for efficient transmission, to protect coded images against noise, or to enhance elements of the image such as the edges or the overall contrast. Pattern recognition, on the other hand, is concerned with the classification of objects in the image into one of a finite number of categories. Pattern recognition often assumes that the objects are presegmented from the rest of the image; furthermore, it is assumed that objects are adequately described by a list of values for some fixed set of attributes that can easily be extracted from the image. Classification of the object is done by partitioning the n-dimensional feature space or traversing a sequential decision tree.

Image Understanding

Over the last decade or so, there has developed within Artificial Intelligence, an alternative approach to the machine interpretation of visual data. Usually known as scene analysis or, more broadly, *image understanding*, this approach is characterised by computer programs that are expert in interpreting pictures of a particular domain. A prerequisite to writing such programs is a careful analysis of the structure of the knowledge that we have about images in the domain, the knowledge we have about the objects depicted, the scene, and, most importantly, how structures in the image depict objects in the scene (1). There is now emerging a paradigm for perceptual processes captured by the *cycle of perception* (2). Image understanding is a delicate intermingling of low-level segmentation techniques and high-level interpretation. The cycle alternates segmentation and interpretation although, in a real sense, segmentation *is* interpretation and *vice versa*. In more detail, context sensitive cues segmented from the image invoke models for parts of the scene. These models must be tested; if they are established then certain consequences follow, including the instigation of a search at the lower levels for new cues which can invoke new models....

All of our work is based on this paradigm. We are concerned both with developing a new working theory of perception and with applying the theory to significant image domains. The two domains we focus on are Landsat images and freehand sketches.

Understanding Landsat Images

The goal of interpreting Landsat image data automatically by computer programs has been the focus of much research effort in recent years. Achieving such a goal would have enormous economic and social benefits; however, many of the results to date are not too encouraging if judged by the accuracy of the pixel classification. Any attempt to classify pixels which relies solely on the spectral signature of each individual pixel is, in most applications, doomed to mediocre performance. The spectral evidence alone is not enough. Most Landsat classification systems, whether they be supervised or non-supervised, maximum-likelihood or minimum distance, classifying single pixels or samples of pixels, operate within the pattern recognition classification paradigm.

The attraction of the classification approach is its computational simplicity - it would indeed be nice if the world were so structured as to make it work; however, the world is more subtle than that. The fundamental assumption is that a pixel's interpretation depends only on its spectral attributes not, for example, on its location in the picture or on the interpretation of neighbouring pixels. If one randomly permuted the pixels of a Landsat image such a classifier would not even notice, let alone change its interpretation of any pixel! Clearly the paradigm is ignoring the crucial fact that all meaningful images are subject to spatial organization which must be heavily exploited in the interpretation process. Moreover, it is further assumed that there will, in fact, be good separation between the classes in the spectral space. Ideally, each class would be separated from all the rest by a set of hyperplanes in the spectral space. In reality this hoped-for separation does not occur; there is often enormous overlap of the distributions for each class. When the classifier is asked to pronounce on a pixel whose signature falls within such overlaps it can only do so by making unreliable guesses. As a consequence of the fact that the two essential assumptions do not hold, the performance of such classifiers is often poor, except, of course, in the rare cases where pixels can be classified context-free and where clear separation does occur: when distinguishing water from forest, for example.

There have been several attempts to use spatial information in the interpretation of remotely sensed images. These often involve adding additional features, for example local texture, to the four intensity band values of a pixel. Another approach classifies local blocks of pixels that are reasonably homogeneous. These do, in fact, increase the accuracy of the classification somewhat; however they are still open to most of the same criticism.

Interpretation-guided Region Merging

The implications of the image understanding paradigm for the Landsat task are that we should attempt to find cues that can be sensibly interpreted. These can then suggest interpretations for new areas of the picture. To be specific, we can introduce the spatial sensitivity we are after by dealing with regions of connected pixels in the picture, not individual, isolated pixels. Region merging techniques have been developed in scene analysis that show, starting with atomic regions of pixels with identical intensities, how one can merge regions with similar intensities to produce a region segmentation for subsequent interpretation. Moreover, one can go further and allow the interpretation of the initial regions to control the segmentation or merging process itself (3).

The strategy is an emerging principle in artificial intelligence. It is coming to be known as best-first or island-driving: in this case, use the interpretations of those segmented partial regions we are most confident of as a guide and context for further region segmentation and interpretation.

In adapting these ideas to this task we combined the best features of the pattern recognition approach with these scene analysis techniques. As a front-end we used a supervised maximum likelihood classifier that is traditional in all but one important respect: if a pixel is highly ambiguous, that is, if no one class is overwhelmingly likely, the classifier is simply to report the two most likely classes. Pixels are either unambiguously put in a class or ambiguously related to two classes. The subsequent region merging process uses this information to form regions of connected pixels for each of these unambiguous (strong) and ambiguous categories. It then should sensibly merge regions: the strong regions grow by devouring, in amoeba-like fashion, the ambiguous regions according to a systematic set of rules. The interpretation of the spatial context of a pixel can then occasionally suggest that its most likely interpretation should be discarded in favour of its second most likely interpretation.

To test this method we used ground truth data for a forested area of Vancouver Island and a Landsat image of the same area. The task was to classify the ground cover into regions of old growth, second growth, recent clear cut and water. A conventional maximum likelihood classifier misclassified 30% of the pixels while the interpretation-guided region merger misclassified only 21% of the pixels. Moreover, the number of regions in the final classification was cut in half. (The number of regions vitally affects the "readability" of the final product.) However, these results were achieved at the cost of an increase in the CPU time by a factor of 3.5. Full details of the algorithms and the results are in reference (4).

Pyramids and Forests: Using Image Hierarchies

Recently we have completed a study on the use of image hierarchies in understanding Landsat images (5). Several workers in scene analysis (6,7) have advocated a pyramid-like structure representation of the image. This consists of a series of arrays. At the lowest level the entire image is represented as a square array of pixels. At the next level up an array element is formed by averaging the intensity values of a 2x2 cell of pixels at the level below. This process is continued until at the highest level the image is represented by a single pixel whose intensity values are the average of all the original pixels. The pyramid is a useful structure for allowing image segmentation routines to work at varying levels of detail in the image. A segmentation at one level in the pyramid, for example, may serve as a plan for the segmentation at the level below. We have studied ways to use these image hierarchies to implement the cycle of perception for Landsat images. In particular, in reference (5), we show how, with the same data as we used previously, one can achieve remarkable improvements in the readability of the segmented image (reduction in the number of regions) together with slight improvements in accuracy. The crucial point is that this can be achieved without any sacrifice of efficiency; indeed, these techniques can be more efficient than traditional techniques. Using modifications of Levine and Lemeet's approach (7), we show how to classify certain areas of the image at higher levels in the pyramid thereby vastly reducing the number of calls to the classification routines and ensuring greater homogeneity in the final segmentation. The time saving is somewhat offset by the cost of building and maintaining the image hierarchy. Our conclusion is that these techniques allow some of the spatial sensitivity required for Landsat image interpretation and obtain some of the effects of our region merging techniques without their high cost.

Understanding Sketches

Another aspect of our work is concerned with understanding freehand sketches drawn on a graphical data tablet. MAPSEE (8) interprets sketch maps depicting roads, towns, rivers, mountains, bridges, land, lakes and ocean. Our aims are to provide, eventually, useful sketch understanding programs, to investigate algorithms required to implement forms of the cycle of perception, and to explore how to represent geographic and cartographic semantics so that we can use those representations in our Landsat work.

A current issue within the cycle of perception paradigm is the use of cue-model hierarchies wherein an instantiated model may serve as an internal cue to a higher level of interpretation. Havens (9), in his LISP-based language MAYA, has provided mechanisms to allow this and to allow an intermingling of top-down model-driven and bottom-up, cue-driven techniques. Mulder (10) has elaborated MAPSEE to allow many levels of interpretation for a three-dimensional world of architectural sketches.

The Interdisciplinary Programme in Remote Sensing

In carrying out our Landsat investigations we received excellent cooperation and vital ground truth data from Dr. Peter Murtha of the Faculty of Forestry at UBC. This cooperation has grown into an Interdisciplinary Graduate Programme in Remote Sensing. This Programme involves Geography, Soil Sciences, Civil and Electrical Engineering and Astronomy as well as Forestry and Computer Science. It is supported by the Province of British Columbia. This support is allowing us to develop the facilities and personnel necessary to enhance the research and teaching environment in an area of central concern to the economies of British Columbia and Canada.

Conclusion

In our Landsat work, we have so far demonstrated how to bring certain spectral, spatial and semantic constraints to bear on the segmentation task. Further progress will require the continued development of computational mechanisms whereby we express and use some of the wide varieties of knowledge that a skilled photointerpreter brings to the task. We need to develop programs that use, in addition to the knowledge of colour, adjacency and meaning we have used already, knowledge or theories of shape, texture, lighting, cloud cover, geography, cartography, *a priori* information on a ground area, the purpose of logging roads, the effect of steep slopes on logging machinery, and, crucially, the effect of climate, soil, time of year and altitude on the growth of trees!

Acknowledgements

The research reported here was supported by grants from the National Research Council of Canada, the Province of British Columbia and the University of British Columbia.

References

(1) A. K. Mackworth, "How to see a simple world", in *Machine Intelligence 8*, E. W. Elcock and D. Michie (eds.), N. Y.: Wiley, 1977, pp. 510-537.

(2) A. K. Mackworth, "Vision research strategy: black magic, metaphors, miniworlds and maps." Proc. Workshop on Computer Vision Systems, Amherst, Mass., (in press).

(3) M. Tenebaum and H. G. Barrow, "IGS: a paradigm for integrating image segmentation and interpretation" in *Pattern Recognition and Artificial Intelligence*, N. Y.: Academic Press, 1976.

(4) D. W. Starr and A. K. Mackworth, "Exploiting spectral, spatial and semantic constraints in the segmentation of Landsat images", TR78-1, Dept. of Computer Science, Univ. of British Columbia, Vancouver, B. C., 1978, and *Canadian Journal of Remote Sensing* (in press).

(5) E. Catanzariti and A. K. Mackworth, "Forests and pyramids: using image hierarchies to understand Landsat images", TR78-2, Dept. of Computer Science, Univ. of B. C., Vancouver, B. C., 1978.

(6) S. L. Tanimoto and T. Pavlidis, "A hierarchial data structure for picture processing", *Computer Graphics and Image Processing*, 4, 2, 1975, 104-119.

(7) M. D. Levine and J. Lemeet, "A method for non-purposive picture segmentation", *Third Int. Joint Conference on Pattern Recognition*, Coronado, Ca., 1976, 494-498.

(8) A. K. Mackworth, "On reading sketch maps", *Fifth Int. Joint Conference on Artificial Intelligence*, Cambridge, Ma.: MIT, 1977, pp. 598-606.

(9) W. S. Havens, "A procedural model of recognition", *Fifth Int. Joint Conference on Artificial Intelligence*, Cambridge, Ma.: MIT, 1977, p.264.

(10) J. A. Mulder, "Using multi-level semantics to understand sketches of houses and other polyhedral objects", Dept. of Computer Science, Univ. of B. C., Vancouver, B. C., 1978.