THE AUTOMATIC RECOGNITION OF INDIVIDUAL TREES IN AERIAL IMAGES OF FORESTS BASED ON A SYNTHETIC TREE CROWN IMAGE MODEL

By

Richard James Pollock

B. L. A. University of Toronto

M. Comp. Sci. Concordia University

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

COMPUTER SCIENCE

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 1996

Department of COMPUTER SCIENCE

The University of British Columbia
Vancouver, Canada

Date 26 JUNE 1996

Abstract

The thesis of this work is that individual tree crowns can be automatically recognized in monocular high spatial resolution optical images of scenes containing boreal or cool temperate forests in a leaved state. The thesis was advanced by developing and testing an automatic tree crown recognition procedure that is based on a model of the image formation process at the scale of an individual tree. This model provides a means of applying specific scene and image formation knowledge to the recognition task. The procedure associates instances of a three-dimensional shape description with locations in a scene image such that the descriptions estimate the visible scene extent of tree crowns that existed at the corresponding scene locations when the image was acquired. This provides an estimate of the average horizontal diameter of the vertical projection of individual recognized tree crowns, and a basis for species classification.

This work makes a contribution to the overall effort to increase the level of automation in forest type mapping. This work also introduces and demonstrates the use of a pre-defined image model to support the manual acquisition of a sample of unmodelled tree crown image properties, and the use of constraints related to the spatial relationships among multiple neighbouring candidate recognition instances to resolve image interpretation conflicts.

The procedure was tested with a scene of mixed uneven-aged forests in which the trees represent a wide variety of species, size, and growing conditions. The results were assessed on the basis of ground reference data and compared to those produced by human interpreters. The scene represented a greater level of difficulty than that which has been addressed by previous attempts at automating the tree crown recognition task. The test results show that the procedure was able to largely accommodate the variation represented by the test scene, but that human interpreters were better able to accommodate irregularities in tree crown form and irradiance that were caused by tight vertical and horizontal spacing of the crowns.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I acknowledge the guidance provided by the members of my supervisory committee: Drs. Don Leckie, Jim Little, David Lowe, and Bob Woodham. Bob Woodham was my primary source of scholarly advice and encouragement. I consider myself extremely fortunate to have had Bob as my Research Supervisor. Don Leckie, Project Leader at the Petawawa National Forestry Institute (PNFI), provided access to the MEIS II images, reference data, and associated documentation that were used in this research. Don was a source of many valuable suggestions, and he shared with me his detailed knowledge of forest type mapping procedures and his ideas for their improvement. David Lowe provided an implementation of the VSM learning procedure and of the nearest neighbours interpolation procedure that I adapted to this research. Jim Little's contagious bouyancy helped me maintain the sense of self-confidence that is so necessary in research.

I thank my University Examiners Drs. Alan Mackworth and Peter Murtha, and External Examiner Dr. John Miller of York University for their generous contribution of time and insight to my oral examination.

I also thank my soon-to-be boss Dr. Jan Mulder of the Alberta Research Council for his patient tolerance of several changes to my completion date.

Marcia Wilson produced the line drawings for many of the figures. Steve Joyce and Ed Cloney, both formerly at PNFI, answered questions about the experimental image and reference data. Art Pope answered questions on the use of the *Vista* software, which I used extensively in my research (Art and David Lowe are the principal developers of *Vista*).

The Laboratory for Computational Intelligence (LCI) is an immensely supportive place in which to work. LCI Lab Manager Dan Razzell does a magnificent job of keeping the computational resources up and running smoothly. The work, ideas, tenacity, and advice of other students in the LCI, particularly Jeffrey Beis, Jiping Lu, Dan McReynolds, Jane Mulligan, Art Pope, and Michael Sahota were constantly inspiring. I give special thanks to LCI Secretary Valerie McRae for being a good friend.

Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada in the form of a postgraduate scholarship to the author and research grants to R. J. Woodham, and by the Institute for Robotics and Intelligent Systems, one of the Canadian Networks of Centres of Excellence.

Ultimately, I could not have done this without the companionship, understanding, and love of my wife Marcia.

Chapter 1

Introduction

The thesis of this work is that individual tree crowns can be automatically recognized in high spatial resolution optical images of scenes containing boreal or cool temperate forests in a leaved state. The images are assumed to be monocular and vertical[1], and acquired under clear sky conditions. We categorize electro-optical sensor images that are produced with a scene element[2] dimension no greater than one metre as having high spatial resolution (this is explained in Section 2.3.1). In the case of stratified forests, the scope of the thesis is limited to the crowns of emergent trees and dominant or codominant trees belonging to the upper continuous stratum. A secondary assertion included in the thesis is that constraining the recognition task on the basis of knowledge of the sensing system, the sensing situation, and the physical characteristics of trees can be the basis of an effective strategy. This work is intended to contribute to the increased automation of forest type mapping procedures.

The thesis was advanced by developing and testing an automatic tree crown recognition procedure that is based on a model of the image formation process at the scale of an individual tree. The procedure associates instances of a three-dimensional shape description with locations in a scene image such that the descriptions estimate the visible scene extent of tree crowns that existed at the corresponding scene locations when the image was acquired (Figure 1.1).

The procedure was tested on multispectral aerial digital images of scenes within the research forest of the Petawawa National Forestry Institute (PNFI) near Chalk River, Ontario,

[1] acquired with a nominally vertical optical axis

[2] region of flat and level terrain spanned by the sensor instantaneous field of view

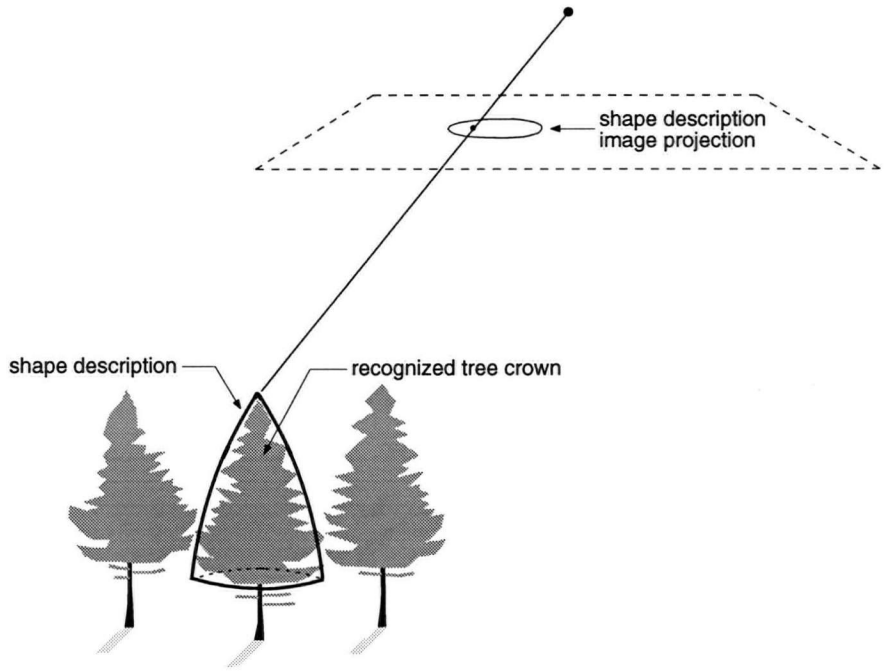


Figure 1.1: Tree crown recognition scenario.

Canada. These scenes contain mixed forest and represent many different stand types, including mixed hardwood and softwood stands, and a wide range of individual tree size. The images were acquired with the second generation of the multi-detector electro-optical imaging scanner (MEIS II), a pushbroom scanner (McColl, Neville, and Till 1983) and they have a 36 cm ground projected pixel dimension. Figure 1.2 displays some of the test image data, scene variation, and results that were produced by the recognition procedure.

The test results were evaluated on the basis of reference data that were collected by PNFI personnel on the ground within the imaged scenes, and were compared to individual tree recognition results that were produced by human interpreters from the same image data.

The remainder of this dissertation is organized as follows:

Chapter 2 provides relevant background information on forest type mapping and describes and motivates a framework for automation in forest type mapping that incorporates the individual tree crown recognition task.

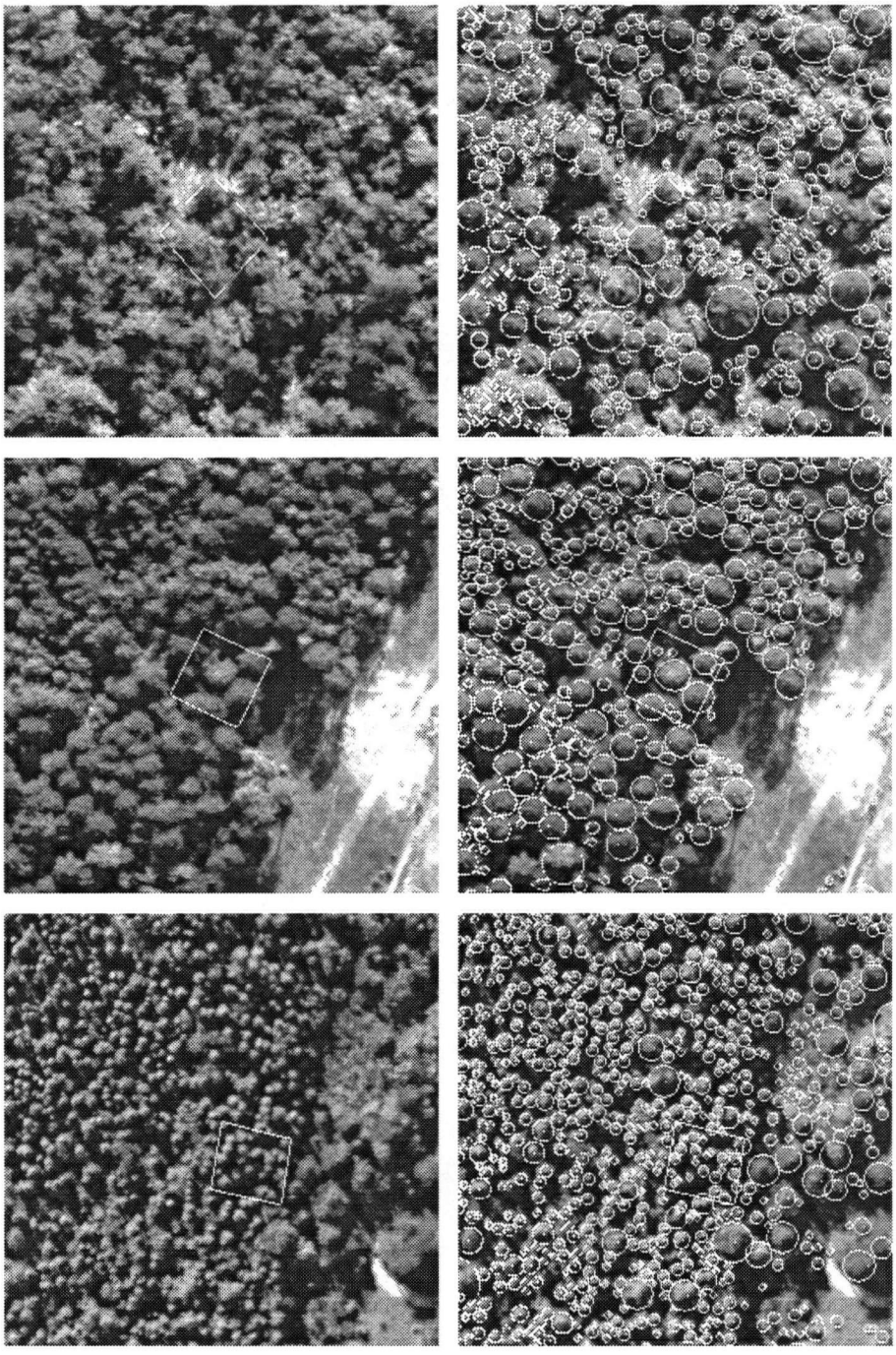


Figure 1.2: Subimages from the test image data and automatic individual tree crown recognition results for 1 hectare subscenes. The subimages were extracted from the visible green wavelength channel of the test image data. The image projection of the shape description boundary for each of the recognized crowns is superimposed in white on the subimages in the right column. The boundary of a 20-by-20 metre ground sample plot is superimposed in white on all of the subimages.

Chapter 3 discusses the individual tree crown recognition task in detail, and motivates and describes a new automation strategy for achieving the goals of the task.

Chapters 4 and 5 respectively describe a tree crown image model and procedures based on the model that implement the proposed automation strategy.

Chapter 6 describes the experiment that was performed in order to test the proposed automation strategy and presents an evaluation of the results.

Chapter 7 summarizes the dissertation, discusses what the experimental results reveal about the validity of the thesis, and suggests further related research directions.

Chapter 2

Background

The work that is described in this dissertation is intended to contribute to the automation of forest type mapping procedures, particularly those that are carried out in Canada. These procedures and the motivation for increased automation are outlined in Section 2.1. A framework for automation that incorporates the individual tree crown recognition task, which is the focus of this work, is described in Section 2.2, and technical factors that motivate and support this particular framework are described in Section 2.3.

2.1 Forest Type Mapping

A *forest type map* stratifies an area of forested land in order to enhance the efficiency of ground sampling for forest inventory. A stratum is composed of forest stands of a particular type. Normally, the purpose of a forest inventory is to provide an estimate the spatial distribution of wood volume (Gillis and Leckie 1993; Gillis and Edwards 1988). Forest inventories have also been adapted to estimate the spatial distribution of wildlife habitats, and there is increasing pressure to make new inventories even more suited to this purpose (O'Brien and Chojnacky 1992; Schroer 1992).

In Canada, forest type mapping normally requires the evaluation of species composition, density, tree height, tree age class, and site class attributes (Gillis and Leckie 1993). Regardless of the exact definition of these attributes and the classification systems associated with them, both of which vary among provinces, the manual interpretation of medium scale vertical aerial

photographs in stereo is the standard method of evaluating the first three (Ahern and Leckie 1987). Site class (an indication of the potential forest productivity of the site) and tree age class are estimated from the values of the other attributes, contextual information, and existing records.

The photo-interpretation procedure produces a set of stand boundaries and, for each stand, a set of stand attribute values. A stand attribute value is the average of a sample of local values for that attribute within the stand. The local attribute values may pertain to individual trees (*e.g.,* height) or to clusters of trees (*e.g.,* species composition, density). Stand boundary delineation does not involve the explicit measurement of local attribute values, but instead relies on visual estimates of regions where attribute values appear relatively homogeneous and distinct from those of adjacent regions. Even when based on considerable experience, these estimates may vary significantly from one interpreter to another (Biging, Congalton, and Murphy 1991). Limits on the minimum allowable stand area (which may vary according to stand type) and continuity with earlier-defined boundaries for adjacent land areas are also observed (Brickell 1992).

Local attributes are evaluated on a sample basis with the stand boundaries using simple tools such as a parallax bar (for measuring height), wedge or dot type crown diameter scales, crown closure example cards (for estimating the fraction of ground area that coincides with vertical tree crown projections) and reference stereograms (which provide examples of species identification, species composition estimation, and stand delineation) (Hershey and Befort 1995; Howard 1991). The local attributes are normally evaluated only for the emergent trees and the dominant and codominant trees of the upper continuous canopy stratum, since these are the trees that are visible in aerial images. These are also the trees that are most significant with respect to merchantable wood volume. Understory vegetation is significant with respect to wildlife habitat and stand development but is not generally visible in aerial images.

Interpreters who are highly experienced, familiar with photographs of the given scale, and familiar with the type of forest that is pictured often estimate the values of the local attributes

or even the stand attributes by eye (pers. comm. Dr. Donald Leckie[1]).

The photo-interpretation results are usually incorporated into a computerized geographic information system (GIS). The usual procedure for doing so consists of manually transcribing the stand boundaries onto a base map, manually digitizing the stand boundaries from the base map, and manually entering stand attribute values into a database management system that is a part of the GIS. Each of these steps is subject to quality control. This procedure is typically even more expensive than the photo-interpretation (Leckie 1994).

An estimate of the amount of wood volume per unit land area in each stand can be derived from an estimate of stand tree density (numbers of stems per hectare) combined with observations that are related to individual tree wood volume. Counting trees on medium scale aerial photographs or high spatial resolution digital images is a tedious and slow manual task. Therefore, tree density is usually estimated by combining an estimate of crown closure with an estimate of average crown area within the same land region (Howard 1991, pp. 305, 306).

On a species-by-species basis, the average horizontal diameter or area of the vertical projection of individual crowns onto a horizontal surface (referred to in the rest of this dissertation as crown diameter and crown area, respectively) can be a useful predictor of individual tree wood volume when combined with height and supported by ground plot sampling (Howard 1991, pp. 300, 301, 310, 311). Crown area or diameter may also be useful in estimating the amount of non-timber biomass, which is important in wildlife habitat inventories (Warbington and Levitan 1992), as well as in estimating fire hazard, site productivity, and industrially useful non-stem biomass (Schreuder, Gregoire, and Wood 1993).

Accuracy standards for local attribute evaluation from aerial images are not well established or documented. This reflects normal practice in Canadian forest type mapping, which is to directly assess the accuracy of the stand attribute evaluation (Gillis and Leckie 1993). This may involve aerial overflights or ground plot surveys within the stands. The placement of boundaries on the photos may also be checked for obvious inconsistencies. The photo-interpreter may elect,

[1]Project Leader, Petawawa National Forestry Institute, Forestry Canada

or may be required, to perform a certain number of site visits during the interpretation process in order to calibrate the interpretation or resolve ambiguities.

Precision standards for stand attributes values are readily available since they correspond to the associated class intervals. These class intervals are normally described in the legends of the completed forest type maps, and they dictate the coarsest allowable precision for the evaluation of local attributes.

In Canadian forest type mapping, tree height classes are based on intervals ranging in width from 1 m (calibrated with field surveys) to 10 m, and crown coverage classes are based on intervals ranging in width from 10 % to 30 % (Gillis and Leckie 1993). Crown diameter and stem density do not appear in the legends of Provincial forest type maps, but may be involved in the wood volume density estimates that appear on some maps. In British Columbia, species composition for young stands is based on the number of stems per hectare for different species. Howard states that, for North American forests, crown diameter manually evaluated from 1:8,000 to 1:20,000 scale photographs can often be reliably assigned to classes that represent 0.5 m to 1.5 m wide diameter intervals (Howard 1991, p. 299).

There is a continual need for up-to-date forest type maps. Each year in Canada, approximately 27 million hectares of land are mapped in Provincial forest management inventory programs (Leckie 1993) in order to continually re-map a total of 344 million hectares on a 10 to 20 year cycle (Gillis and Leckie 1993). At the same time, inventory managers are having difficulty finding competent photo-interpreters to replace the ones that are retiring (Leckie 1993). Evidently, few forestry programme graduates are attracted to forest type mapping work (pers. comm. Dr. Peter Murtha[2]). This, together with the general attraction of cost reduction, motivates the development of automation in the forest type mapping process.

[2]Professor, Faculty of Forestry, University of British Columbia

2.2 A Framework for Automation in Forest Type Mapping

The work that is described in this dissertation is intended to contribute to the development of a system for automatically extracting information relevant to forest type mapping from digital optical images that are vertical, monoscopic, and have high spatial resolution.

The envisaged image interpretation approach is based on the automatic recognition of individual tree crowns that are visible in the image data, which is the focus of the current work. The recognition results would include estimates of the image region, diameter, and scene location of individual crowns.

Species classification would be based on features derived from individual crown image regions, but may also be constrained by general knowledge of the tree species expected in the scene and their commonly occurring associations. Knowledge of species site preferences might also be used in the species classification procedure should data on site conditions within the scene be available.

Laser range data acquired concurrently with the image data would be registered to the image data and analyzed in combination with the recognition results in order to obtain a sample of individual tree height estimates (this is discussed further in Section 2.3.5).

The most direct application of the information provided by the proposed image interpretation approach would be the evaluation of attributes for predefined stand regions. This information would represent a denser sample of local attribute values than is typically obtained in the manual photo-interpretation process. A more advanced application would be the automatic definition of stand regions through the grouping of neighbouring trees such that the variance of the properties of individual trees and tree clusters would be within pre-defined limits within each stand region.

The proposed image interpretation approach would produce basic information that could be stored as an intermediate product and adapted to a variety of forest type classification systems. This is an important issue because of the increasing demands on the versatility

of forest management inventories. In contrast, mapping procedures based on manual photo-interpretation do not produce an application-independent intermediate product.

2.3 Factors that Motivate the Proposed Framework

The following are technical factors that motivate the proposed framework for automation in the forest type mapping process (each factor is discussed in further detail in the indicated section):

- Airborne electro-optical sensors can produce images in which individual tree crowns can be resolved (Section 2.3.1).
- Individual crown recognition is required for general forest type mapping from optical images (Section 2.3.2).
- Data from airborne or satellite-borne synthetic aperture radar (SAR) sensors are not a complete substitute for high spatial resolution optical image data in forest type mapping (Section 2.3.3).
- With respect to tree species identification, the high spectral resolution of image data from modern electro-optical sensors offsets their low spatial resolution relative to that of medium scale aerial photographs (Section 2.3.4).
- Co-registered laser range data and monoscopic vertical aerial image data provide an effective means of measuring tree heights that could be further automated through the automatic recognition of individual tree crowns in the image data (Section 2.3.5).
- Inertial navigation system (INS) and global positioning system (GPS) technology has made it possible to georeference images acquired from aircraft, including line scanner images, on the basis of few if any ground control points (Section 2.3.6).

2.3.1 Resolving Individual Tree Crowns in Aerial Images

Leckie used a series of MEIS II images acquired in a visible red wavelength region and at different altitudes over the PNFI research forest to illustrate the relationship between the ground projected pixel dimension (GPPD) and the ability to manually resolve individual tree crowns in the image (Leckie 1993). At a GPPD of 0.75, 1.5, and 3.0 m most individual crowns are respectively clearly resolved, inconsistently resolved, and consistently not resolved (presumably the flight conditions were such that the scene elements were practically square). The corresponding scene element dimensions are 0.56, 1.12, and 2.25 m, because the MEIS II along-scanline scene element dimension is 75 % of the along-scanline GPPD due to resampling of the detector array output as part of a registration procedure that removes optical system distortions and prevents aliasing (McColl, Neville, and Till 1983). This suggests that automated individual tree crown recognition may be possible in electro-optical sensor images of boreal and cool temperate forests with a scene element dimension no greater than approximately 1.0 m. For convenience, we categorize such images as having high spatial resolution. Digital images with comparable spatial resolution may also be obtained by digitizing 1:10,000 to 1:20,000 scale (medium scale) aerial photographs or the analogue signal from a video camera. In contrast, image data acquired with a scene element that is large enough to encompass multiple trees (for example, image data from currently operating civilian satellite sensors) is categorized as having low spatial resolution in this dissertation.

2.3.2 The Necessity of Individual Tree Crown Recognition

Experiments in the analysis of low spatial resolution optical image data strongly suggest that individual tree crown recognition is necessary for extracting information from optical sensor images at the level of detail required for compiling type maps of all but the simplest forests. At low spatial resolutions, radiance from multiple trees and their surroundings contributes to a single pixel value. Therefore the information on tree species composition, size, and density that is required for forest type mapping is not directly accessible from low spatial resolution

image data. The possibility of inverting models of scene element reflectance in order to extract information on the sub-element components contributing to the radiance has been a subject of considerable research. Li and Strahler had reasonable success in computing the average size and spacing of trees corresponding to individual pixels of simulated SPOT HRV image data (20 m GPPD) (Li and Strahler 1985). The scene contained single storey stands of one or two coniferous species on a uniform contrasting background (species composition was not computed), a conical crown form was assumed, and models of the density, height distribution, and radiometric properties of the trees and the background provided computational constraints. However, it is not clear that techniques such as this can be extended to deriving species composition, tree density, and tree size information for scenes as complex as a typical cool temperate mixed or uneven-aged forest (Ahern and Leckie 1987; Leckie 1990a).

2.3.3 Comparison of SAR Sensing and Optical Sensing

Airborne and satellite-borne SAR sensors are providing an increasingly available alternative to optical sensing, and the haze and cloud penetrating capability of SAR sensing is attractive in forestry applications. The recently deployed Canadian RADARSAT satellite carries a C-band HH polarized SAR sensor. However, experiments with image data from both airborne and satellite-borne SAR sensors indicate that RADARSAT data will not provide a significant improvement over satellite-borne optical sensors in detailed species composition mapping, and that the significant role in forestry of RADARSAT data will instead be in mapping logged or burned areas (Ahern, Leckie, and Werle 1993).

The spatial resolution of SAR data is more constrained than that of electro-optical data since it depends on both the radar pulse bandwidth and the azimuth beam width (Fitch 1988). Operational airborne SAR sensors produce data with GPPDs on the order of several metres. In an experiment involving medium spatial resolution image data from airborne SAR and electro-optical sensors, Leckie found that the SAR data were better for distinguishing different softwood species stands than the electro-optical data, but that the electro-optical data was better for

distinguishing hardwood and softwood stands, and open and forested areas (Leckie 1990b). The SAR data were acquired in different combinations of sending and receiving polarization in both the X and C bands with GPPDs of 3.0 m (range) and 4.0 m (azimuth). The electro-optical data were acquired in visible and near-infrared bands with a GPPD of 4.1 m, and both data sets were resampled to the same 10 m grid before analysis. These results indicate that although airborne SAR image data can be a useful complement to medium spatial resolution electro-optical image data in forestry applications, it cannot replace medium or high spatial resolution electro-optical image data. The degree to which airborne SAR image data can complement high spatial resolution electro-optical image data in tree species discrimination is still unknown.

2.3.4 The Advantage of Multispectral Optical Sensing

High spatial resolution images from operational electro-optical sensors generally have a significantly coarser spatial resolution, with respect to manual forest cover interpretation, than do aerial photographs at a scale of 1:15,000 or larger (Leckie 1990a). This means that crown margin shape and crown texture are less useful as features for species classification when interpreting the digital images than when interpreting photographs. However, in experiments reported in Leckie (1993) manual tree species identification from colour composite displays of multi-channel high spatial resolution MEIS II image data was more accurate than manual species identification from 1:9,000 scale colour infrared photographs of the same scene. The higher spectral resolution of the digital images relative to that of the photographs appeared to more than compensate for their lower spatial resolution.

Experiments in the automatic supervised classification of various tree species that are based on multi-channel pixel values from manually delineated tree crown regions in high spatial resolution digital images have produced results in which 50 % to 90 % of the test trees were correctly classified, depending on species (Gougeon 1995; Gougeon 1993; Hughes, Evans, and Burns 1986; Thomasson, Bennet, Jackson, and Mailander 1994). This accuracy range may overlap those obtained through operational manual photo-interpretation procedures: in an experiment

involving experienced photo-interpreters and 1:12,000 scale aerial photographs, between 68 % and 88 % of a sample of coniferous stands in Northern California were correctly classified for species (Biging, Congalton, and Murphy 1991). The automatic classification experiments did not involve texture or shape features, but image values were selected from portions of tree crown image regions that were estimated to correspond to directly irradiated crown surfaces.

2.3.5 Laser Ranging for Tree Height and its Synergy with Individual Tree Crown Recognition

Tree height can only rarely be directly evaluated from monoscopic optical images. Furthermore, it is doubtful that tree height can be reliably evaluated automatically from stereo high spatial resolution images. Monoscopic tree height estimation requires measurement of the length of the shadows cast on level open ground by the trees. Such shadows exist only in very open stands or at the edge of clearings. Stereo tree height estimation requires the discrimination of both the base and apex of the tree, and usually very few trees in the scene have a base that is visible in an aerial image.

Airborne laser profiling and "light detecting and ranging" (LIDAR) systems provide an alternative means of estimating individual tree height or average stand height to a useful degree of accuracy. These systems generate vertical range data along a narrow ground trace by measuring the time of travel of reflected laser pulses. A laser profiling system records the distance to parts of tree crowns, and to the ground where there are gaps in the canopy. In order to transform the profile data into tree height data, the samples that correspond to ground reflections must be recognized and a ground profile must be derived from them. This may be difficult to do where the forest canopy is dense and the topography is varied.

A LIDAR system records the amplitude of each received pulse as a function of time. The part of an amplitude record that corresponds to tree crown reflections can usually be distinguished from the part that corresponds to the ground reflection. A system implemented by Aldred and Bonner was able to automatically measure average stand height to ± 4.1 m at a 95 % confidence

level or, equivalently, with a 2.1 m root-mean-square (rms) error, from LIDAR data acquired at a height of approximately 400 m above ground-level (Aldred and Bonner 1985) (this height was chosen for compatibility with the limit of another instrument and does not represent an upper operational height limit for the LIDAR instrument). This result was considered to be at least as good as the results obtained from parallax difference measurements made with medium scale aerial photographs. Scanning LIDAR systems can be used to acquire range data samples over two dimensions, but the accuracy with which tree height can be estimated using the off-nadir samples has not been determined.

The range data that closely correspond to the centres of the vertical projections of tree crowns must be identified in order to obtain a sample of individual tree heights (Jacobs, Evans, and Ritchie 1993). In some sensing systems, it is possible to synchronize the acquisition of range data and image data, and to automatically determine the image coordinates of the range data samples. Co-registered range data and high spatial resolution digital image data in which individual trees have been recognized would be complementary in that the localization results could be used to determine which range samples are useful for tree height estimation.

2.3.6 Advances in Georeferencing Images Acquired from Aircraft

Image interpretation results must be referenced to a standard geographic coordinate system (georeferenced) in order for them to be useful in any type of mapping application. This means that the image coordinate frame must be georeferenced. Geometric distortion in the images that is caused by platform instability greatly complicates this task. High spatial resolution optical image data for civilian use are currently only acquired from aircraft, and the various types of line scanner sensors provide the best spectral resolution. Because undesired aircraft motion can occur between each scan when acquiring line scanner images, the geometric distortions that occur in such images are more complex than those that occur in film and digital frame camera images. However, good progress is being made in the development of techniques for georeferencing line scanner images from aircraft on the basis of few if any ground control points

(GCPs).

Gibson achieved rms errors of sub-pixel and single-pixel sizes, respectively, for horizontal and vertical positions when correcting and georeferencing MEIS II images with a 0.5 m GPPD (Gibson 1994). This experiment involved a fore-, aft-, and nadir-view image acquired for each of five approximately parallel flightlines, and three additional flightlines oriented approximately orthogonal to the first five. An INS mounted on the sensor was used to track sensor velocity, orientation, and (relatively inaccurately) position. A total of five GCPs and 3375 conjugate points were manually located in the images (Gibson states that this number of conjugate points is unnecessarily high). The correction and georeferencing was performed with a least-squares bundle adjustment.

Cosandier and his colleagues achieved average rms errors of single pixel size for horizontal positions in correcting and georeferencing 7.5 m GPPD images acquired with the compact airborne spectrographic imager (CASI) (Cosandier, Ivanco, and Mah 1992). No GCPs or conjugate points were used. A vertical gyroscope integrated with the CASI provided roll and pitch data and a pair of GPS receivers (one in the aircraft and one at a ground station) provided position and heading data.

Chapter 3

The Recognition Task

Section 3.1 defines precisely what is meant by the term "object recognition" in this dissertation. This definition provides a basis for the statement of the overall goal of the individual tree crown recognition task, which is given in Section 3.2. Strategies that have been developed for achieving related goals are discussed in Sections 3.3 and 3.4. The alternative strategy for achieving the stated goal that was adopted in this work is described and motivated in Section 3.5.

3.1 Object Recognition

We consider the process of object recognition from a monoscopic optical image to have two components: *detection* and *localization*.

Detection is the process of deciding whether or not a particular region of the image depicts the existence of a particular object (or instance of a particular type of object) in the scene at the time that the image was acquired.

Localization can be further divided into *image-space* and *scene-space* components. Image-space localization is the process of defining the set of two-dimensional image-space coordinates to which the visible spatial extent of the detected object, or specified visible parts of the detected object, was projected. Scene-space localization is the process of defining the set of three-dimensional scene-space coordinates that coincided with the detected object, or specified parts of the detected object, at the time that the image was acquired.

Image-space localization has two broad purposes: to support further analysis and interpretation of the image, and to support scene-space localization. The extraction of further information about the type, three-dimensional shape, or condition of the object, and the recognition of further objects that are composed of the already recognized objects are examples of image analysis or interpretation tasks that may be supported by image-space localization.

Scene-space localization can be formulated as two tasks: determine the coordinates of the points-of-interest on or within the object with respect to an object-centred coordinate frame, and determine the position and orientation of the object-centred coordinate frame with respect to the scene coordinate frame. The image-space localization results are necessary but not sufficient to perform these tasks because the back-projection of an arbitrary point in the image-space into the scene-space is not unique. The further constraints that are required to perform these tasks can sometimes be derived from some combination of external measurement, (*e.g.,* measurement of the position and orientation of the sensor within the scene with a radar altimeter and an inertial navigation system), image analysis (*e.g.,* estimating surface shape from shading patterns in the image), and reference to a pre-defined geometric model of the detected object (which may have been used in the detection process). The techniques that are applied will depend on the assumptions that can be made about the scene and the image formation process, the external measurements that can be made, and the required precision and accuracy of the localization results.

3.2 Goal of Individual Tree Crown Recognition

A procedure that performs individual tree crown recognition from monoscopic optical images and that fits within the framework described in Section 2.2 must perform image-space localization that supports further processing for species identification, and perform scene-space localization that provides an estimate of the crown diameter of the detected trees. Limitations on the applicability of the procedure with respect to terrain properties, forest stand properties (including tree size and density), and the location of tree crown images within the overall image

extent should be minimized.

As discussed in Section 2.1 there are no established accuracy standards for individual tree crown detection or diameter estimation that can be used to measure the success of an automated procedure. Human performance is a meaningful alternative standard since manual evaluation of local attribute values is the current basis for stand attribute evaluation (see Section 2.1).

3.3 Token-based Individual Tree Crown Recognition

Procedures to automatically recognize individual tree crowns in aerial images on the basis of the detection of image tokens that were manually observed to be associated with, or incorporated within, examples of tree crown images have been implemented by Gougeon (Gougeon 1995; Gougeon 1994) and Pinz (Pinz 1991). These two procedures emphasize image-space localization for the purpose of counting tree crowns and for providing a basis for species identification; crown diameter estimation is not explicitly addressed. They involve different image tokens, but they have common limitations: both procedures rely on the general assumption of a high level of image contrast between a tree crown image and the immediately surrounding region, which implies certain scene irradiance and tree crowding conditions. They are also limited in the range of tree size that can be accommodated. These limitations will be discussed in greater detail after the procedures and the results that have been produced by them are outlined.

Pinz's procedure is based on the assumptions that the tree crown image regions have near-circular boundaries and that they generally have higher pixel values than the surrounding image regions. The procedure smooths the image with a lowpass filter of a size that is related to the expected size range of the tree crown image regions, and locates local maxima in the smoothed image. Each local maximum is considered to be the centre of a tree crown image region. The average of the smoothed image pixel values in concentric circles of increasing radius (starting and ending at user-specified radii) centred at each local maxima is computed. For a given local maximum, the first circle whose average is a user-specified fraction of the smoothed image value at the local maximum is selected as the boundary of the crown image region. The image tokens

in this procedure are local maxima, and circular image trajectories across which the average image pixel value gradient is high.

Pinz's procedure was tested on portions of large and medium scale aerial photographs that were digitized to have GPPDs of 0.40 m and 0.80 m. The scenes portrayed in the test subimages contained well-spaced coniferous trees. The shadows that were cast by the trees when the images were acquired were long enough to cause the between-crown image regions to be much darker than the images of the directly irradiated portions of the crown images. Also, the subimages exhibit no noticeable height displacement so the tree crown image regions have a generally near-circular shape. Some results are illustrated in (Pinz 1991), but their accuracy is not assessed. Many tree crown image regions of widely varying sizes appear to be correctly delineated. However, the results also contain many cases where the same image region is interpreted both as a single large crown and as multiple small crowns.

Gougeon's procedure is based on the assumption that image regions corresponding to individual tree crowns are separated by narrow interstices of relatively low image pixel values. The procedure smooths the image with a 3-by-3 averaging filter, finds local minima within a 3-by-3 neighbourhood, joins minima that occupy the same 5-by-5 neighbourhood, and applies a hierarchically organized system of heuristically defined procedures to join sets of linked minima into possible crown outlines. The image tokens in this procedure are sequences of local minima that define closed boundaries of image regions. The permissible spatial relationships between adjacent minima in a boundary sequence were derived from inspection of example crown images and were manually encoded into the procedure.

Gougeon's procedure was tested on a near infrared wavelength channel of a MEIS II subimage of a scene that contained several tree plantation compartments. The subimage has a GPPD of 0.31 m and individual tree images do not exhibit noticeable height displacement. The solar zenith angle at the time of image acquisition was 67°. A subset of the compartments contained well-spaced trees with crown diameters that varied over a small range. The spacing of the trees combined with the high solar zenith angle caused the image regions between the images of

the directly irradiated portions of the tree crowns to be dark and well-defined. The procedure localized 84% of all the images of actual tree crowns in these compartments, and the number of delineated crown images was slightly closer to the actual number of trees than a manual tree count that was performed with the same image data. Most of the errors occurred where tree crowns were unusually close to each other and the divisions between crown images were relatively indistinct. In such cases, the images of multiple (usually two) crowns were recognized as a single crown image. One compartment in the scene contained a row of trees that had grown such that in the image most of them appear to be in physical contact with an adjacent tree. The procedure localized 35% of all the images of actual tree crowns in this compartment. Images of relatively large tree crowns in another compartment were separated into multiple regions. These tree crown images were correctly delineated when the procedure was run on the image data that had been resampled to a 0.70 m GPPD.

The demonstrated scope and effectiveness of the procedures described above point out the following sources of difficulty in the individual tree crown recognition task:

- The boundary of the image of an individual tree crown may be weakly defined, or even only implicit, for much of its length. This may be the case when the tree crown is in physical contact with other crowns, or is viewed against a background with which it has a low degree of image contrast. Such a background may consist of understory trees or shrubs, or even other trees of the same canopy layer in parts of the image where there is a large amount of height displacement. Wide crown spacing combined with short crown shadows may prevent the understory or ground from appearing much darker than tree crowns in vertical images. Acquisition of image data when the solar zenith angle is large can result in dark between-crown image regions. However, a large solar zenith angle may cause tree crowns to cast shadows on each other in places where trees are crowded, and this would add variation the appearance of the tree crown images.

- The size of visible tree crowns can vary widely over a small area of land in uneven-aged and mixed stands. A recognition procedure that is effective only when the size of a crown

image is within certain fixed limits must be applied to resampled versions of the image in order to accommodate relatively large crown image size variation (*e.g.*, Gougeon's procedure). Alternatively, the procedure can be designed to search a single version of the image for crown images of size that varies continuously within a wide range (*e.g.*, Pinz's procedure). In either case, conflicting interpretations of the same image region (*i.e.*, one large crown image versus multiple small crown images) may arise.

3.4 Recognition with an Example-based Comprehensive Image Model

An alternative to the token-based approach to object recognition is based on searching for a match between image regions and a model of the full spatial extent of the object image (a *comprehensive* image model). This approach may be effective when there are no reliable tokens for the object images. However, tokens provide a model of an object image that is relatively selective and therefore may also be less sensitive to unimportant object image variation.

Traditional *template matching* (Pratt 1991, § 20.1) is a simple version of recognition with a comprehensive image model. A measure of the strength of the match between a reference image array, or template, and the input image array is computed at different relative offsets of the two arrays. Image-space object location is derived from the relative offset, and the object is considered to be detected at an input image location where the match-strength is locally maximum and above a specified threshold. The n coincident pixel values of the two arrays at a given relative offset may be treated as corresponding elements of two n-dimensional coordinate vectors, and the match-strength defined such that it is inversely-related to the value of a between-point distance measure (*e.g.*, Euclidean distance).

Components of a comprehensive image model may be based on representative examples of object images (example-based), or on knowledge of the physical properties of the object of interest, the sensing system, and the sensing situation (synthesized). The example-based

approach to defining a comprehensive image model is attractive because its implementation is potentially much simpler than the implementation of the synthetic approach: the need for a detailed understanding of the process through which the objects of interest in the three-dimensional world become two-dimensional images, and for the parameters of this process, is avoided. In a particular task domain, the size of an adequately representative set of examples would be directly related to the degree of variation in the spatial extent of the object images (which is related to variations in the shape, size, and orientation of the object in scene-space and to variations in local imaging geometry) as well as to the degree of pattern variation within the object image extent (*e.g.,* two tree crown images may occupy almost the same spatial extent within two uniformly-dimensioned image arrays, but may represent considerable pattern variation because the imaged tree crowns may vary in three-dimensional shape and foliage density). If this variation is naturally large and little control can be imposed on it then the manual effort that is required to collect a representative set of examples may be impractical.

Turk and Pentland's work in detecting human faces in monochromatic images (Turk and Pentland 1991) provides an example of the purely example-based approach to the definition of a comprehensive image model that illustrates the points raised above. A sample of m face images was acquired in which each image array has the same row and column dimensions and each face image occupies approximately the same region within its array. The concatenated rows of each image array were considered as the coordinate vector for a single point in an n-dimensional space. Principal components analysis was used to compute the orthonormal unit vectors in the n-dimensional space that coincide with the first $m-1$ principal axes of the set of m sample points. Turk and Pentland called these vectors *eigenfaces.* The eigenfaces associated with the principal axes along which most of the sample point variance occurs were retained as a compact representation of the m face images. These eigenfaces represent considerable variation in full-frontal face images that all occupy approximately the same region within the image arrays and are all superimposed on a relatively featureless background. An image submitted for interpretation was 'centred' by subtracting the mean of the m sample images from it, and

a face was considered to be detected if the squared Euclidean distance between the centered image (as a vector) and its projection onto the space spanned by the eigenfaces was smaller than a specified threshold.

In a preprocessing step, the m sample images were partitioned into subsets each of which contained images of the same individual with varying facial expression and obtained under different lighting conditions. The components of the projection of the centred version of each of the sample images onto the space spanned by the eigenfaces (*i.e.*, its *principal components*) were computed and stored in a vector. The average principal component vector for each sample image subset was then computed. A submitted image that was interpreted as a face was further indentifed as portraying a particular individual by computing the Euclidean distance between the principal component vector of the submitted image and the average principal component vector for each sample image subset. The assigned identity was associated with the smallest distance.

Turk and Pentland's eigenface technique is attractive in a 'mug-shot' face-detection task domain because the object image spatial extent variation is controlled and the technique automatically derives a representation of most of the pattern variation within an object image sample that is much more compact than the sample itself. An adaptation of this technique to individual tree crown recognition in the case of limited crown variation is described in Murgu (1996). In the more general individual tree crown recognition task domain involving uneven-aged or mixed-species forest stands, the spatial extent variation is relatively uncontrolled. A rectangular image window of certain dimensions may closely approximate the spatial extent of some individual tree crown images within the overall image array, but in other locations it may encompass the images of multiple relatively small tree crowns or only part of the image of a single relatively large tree crown. Also, height displacement variation may cause the spatial extent of images of uniformly sized and shaped tree crowns to vary in some high spatial resolution aerial images. This means that multiple instances of a eigenface-type image model would be required in this task domain, and a representative tree crown image sample would have to be

manually collected for each of the different window sizes. The precision of the image-space localization of tree crown images would be limited to that which is provided by a bounding rectangle. Experiments in the species classification of tree crown image regions have involved a greater image space localization precision than this (*cf.* Section 2.3.4), and some have even found that classification is significantly improved when it is based on pixel values from only the directly irradiated portion of the crown surface. A modification of the eigenface technique to allow non-rectangular windows is possible, but a large amount of manual effort would be required to estimate the boundary position of each sample tree crown image.

3.5 The Adopted Processing Strategy: Recognition based on a Synthetic Comprehensive Image Model

The processing strategy that was adopted in this work is based on the use of a synthetic comprehensive image model. This image model facilitates a manual training set selection procedure in which unmodelled tree crown image properties are sampled, and the recognition procedure uses the image model that has been extended with example-based knowledge derived from the training set. In contrast to the existing token-based individual tree crown recognition procedures reviewed in Section 3.3, the adopted strategy does not rely on the existence of explicit tree crown image region boundaries or a large amount of contrast between the tree crown image region and the surrounding image region. In contrast to procedures that are based on purely example-based comprehensive image models (see Section 3.4) the adopted strategy permits readily acquired knowledge of the sensing system, the sensing situation, and physical properties of individual tree crowns in the scene (the image formation knowledge) to be used to constrain the task of collecting example-based data on tree crown images, and to constrain the recognition task. Image formation knowledge is directly encoded as value intervals for the parameters of the synthetic component of the image model. The involvement of image formation knowledge in the recognition procedure is also intended to make the procedure less reliant on specific scene

and sensing conditions than other tree crown recognition procedures.

We consider that the definition of a synthetic tree crown image model is generally practical only for a single image channel and only up to, but not including, the application of a uniform gain and offset (this is discussed further in Chapter 4). Further constraints that are related to absolute apparent radiance in one or more wavelength regions and to the gains and offsets that were applied independently to one or more image channels are derived from the training set.

The synthetic component of the image model includes a parameterized three-dimensional description of general crown envelope size and shape. The objective of the recognition procedure is to associate instances of the image model with locations in the image such that the crown envelope descriptions provide approximations of the visible scene extent of tree crowns that existed at the corresponding scene locations when the image was acquired. Each such association of a model instance and an image location is called a *recognition instance.* The projection of the spatial extents of the estimated crown envelopes onto the image provides image-space localization of tree crown image extent. Crown diameter estimates are derived from the model instance parameter values.

The overall strategy is to compute a set of initial recognition instances that satisfy inexpensively applied constraints derived from the image formation knowledge, but that may also contain a large proportion of false instances, and then to filter the recognition instances in further processing steps that involve constraints that are applied at greater expense per recognition instance. The strategy is described in further detail below and illustrated in Figure 3.3.

The training set selection procedure and the recognition procedure are preceded by the same step. In this initial step the image extent and the parameter space of the synthetic component of the image model are searched for locations where a measure of the strength of the match between the synthetic model instance and the local pattern of image values is at a local maximum and is no lower than a liberally-specified threshold. These matches constitute the initial recognition instances. The search region of the model parameter space is defined by the user-specified model parameter value intervals. In the training set selection procedure,

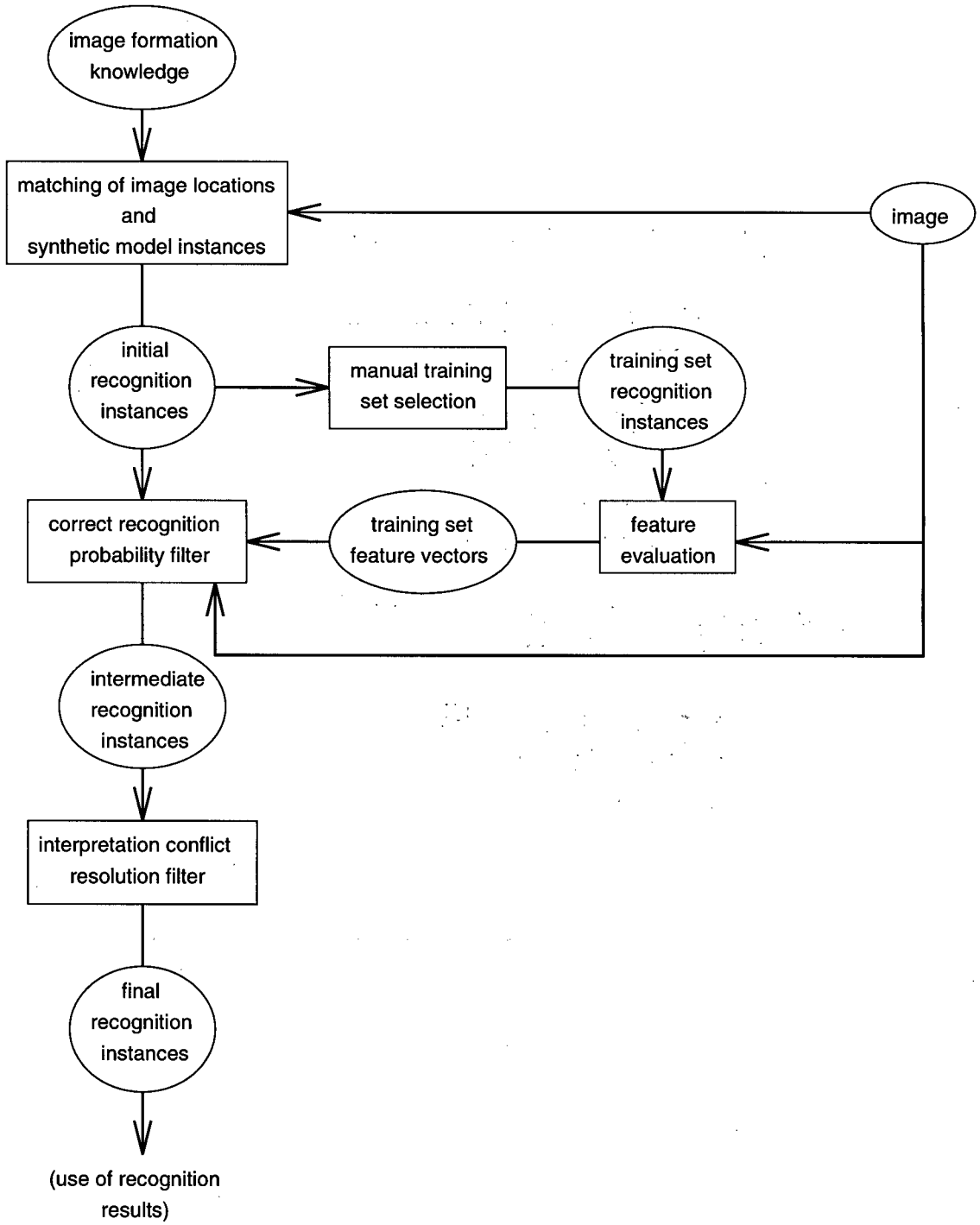


Figure 3.3: Processing strategy (ellipses label data, boxes label operations).

image regions that correspond to a sample of the full range of crown and stand types in the overall scene are manually defined. Examples of correct and incorrect recognition instances are then manually selected from the initial recognition instances within these image regions, and features involving the image values within the corresponding tree crown image regions are evaluated for these examples. The user is not required to manually delineate or window individual tree crown image regions.

The example-based component of the image model is applied in the second step of the recognition procedure. For each of the initial recognition instances, the features are evaluated and the probability that the instance is correct (the *correct recognition probability*) is estimated on the basis of the distribution of the elements of training set in the feature space. Initial recognition instances with a correct recognition probability value under 0.5 are rejected.

The results of the second step of the recognition procedure (the intermediate recognition instances) are expected to sometimes contain conflicting interpretations of the same image region (*i.e.*, one large crown image versus multiple small crown images), since each intermediate recognition instance is established independently of any other. The goal of the third and final step of the recognition procedure is to resolve conflicts due to physically implausible spatial relationships among the intermediate recognition instances while retaining a good overall interpretation of parts of the image that portray forest cover. The physical constraint that within forest stands the volume of the crown envelope intersections is generally less than the remaining crown envelope volume is applied in this step. This is possible because each recognition instance is associated with a crown envelope boundary defined in both the scene-space and the image-space.

A synthetic comprehensive tree crown image model and procedures that implement the adopted strategy are described in Chapters 4 and 5, respectively.

Chapter 4

Synthetic Tree Crown Image Model

The main simplifying assumptions incorporated into the synthetic tree crown image model (referred to simply as the 'image model' in the remainder of this chapter) are the following:

1. a tree crown is rotationally symmetric about a vertical axis in both overall shape and composition

2. the sensor produces single-channel image values that are linearly related to sensor irradiance

3. the images are acquired under clear sky conditions.

Assumption (1) arises from the observation that tree crowns tend to be compact objects with a large degree of rotational symmetry about a vertical axis.

Assumption (2) is reasonable because modern electro-optical sensors such as the MEIS II are capable of producing images in which the pixel values for a scene object have a highly linear relationship to the object's apparent radiance pattern at the sensor. Furthermore, the variation in the length and composition of the atmospheric path between a point on a single tree crown and an airborne sensor at typical flying heights is, as a fraction of the average, minute over the visible extent of the tree crown. Therefore, when sensing narrow wavelength band intervals that are well within the capability of many electro-optical sensors ($< 0.15\,\mu$m), it is reasonable to assume that the atmospheric effect at the sensor on the actual radiance of a single tree crown is equivalent to the application of a uniform gain (due to the atmospheric path absorption and

scattering) and a uniform offset (due to the atmospheric path radiance) (Horvath, Braithwaite, and Polcyn 1970).

Assumption (3) is reasonable because it corresponds to a restriction that normally applies to the acquisition of aerial photographs for forest type mapping.

Electro-optical sensors designed for remote sensing typically generate multiple image channels that correspond to sensor irradiance within different wavelength intervals or bands, and the multispectral radiometric properties of individual elements of trees (individual leaves, bark, and cones) are well-studied, as are the integrated radiometric properties of collections of tree crowns (see Guyot *et al.* (1989) for a survey). This motivates the inclusion of interband radiance relationships in the image model. However, different gains and offsets are normally applied to each image channel to optimize the use of the image value representation and this causes the relationship of the apparent radiance in the different sensing bands to be lost. In principle, this information can be recovered from knowledge of the sensing system, but the atmospheric effect on the relationship of the actual radiance in the different bands may also be significant.

In consideration of the above, the image model is intended to be valid for a single image channel, and only up to, but not including, the application of a uniform gain and offset. The procedure that uses the model (described in Chapter 5) includes a step in which examples of correct and incorrect recognition results that are based entirely on the model are manually selected. The uses of these examples include the derivation of constraints that are related to interchannel image value relationships.

The model components that are concerned with overall tree crown shape and the geometric characteristics of the sensing system are discussed in Section 4.1, and those that are concerned with the scene irradiance, its interaction with a tree crown, and the resulting sensor irradiance are discussed in Section 4.2. The evaluation of a model instance is described in Section 4.3. Notation pertaining to the use of model instances in the training and recognition procedures and ways in which differences between model instances may be measured are described in Sections 4.4 to 4.6.

4.1 Geometric Aspects

4.1.1 Crown Envelope Shape

The crown of a non-suppressed tree can be divided at approximately the widest horizontal cross-section into an upper "sun crown" and a lower "shade crown" (Hakkila 1989). Only sun crowns are modelled since in aerial images of forest stands the sun crown is the only visible part of a tree, unless the view direction is off-nadir and the tree is open-grown, in a low-density stand, or at the edge of a clearing.

We define a tree crown Cartesian coordinate frame $\{C\}$ by the unit vectors $\{\mathbf{x}_c, \mathbf{y}_c, \mathbf{z}_c\}$ such that the system is right-handed, $\mathbf{z}_c$ points vertically upwards, and $\mathbf{x}_c$ is directed towards true north (azimuth = 0°).

The botanist Henry Horn proposed the following equation as a general model for the two-dimensional vertical profile of a crown envelope (Horn 1971, p. 106), defined here in the first quadrant of the yz-plane of $\{C\}$:

$$\frac{z^n}{a^n} + \frac{y^n}{b^n} = 1 \tag{4.1}$$

This is actually the equation for a *generalized ellipse*, due to Piet Hein (Gardiner 1965). In Equation 4.1, a and b are positive numbers that represent the intersection of the curve with the z- and y-axes, respectively, and n is positive and non-zero. When $n = 1$ the curve is a straight line. As n decreases past 1 to 0 the curve becomes increasingly upwards concave, and approaches the z- and y-axes. As n increases from 1 the curve becomes increasingly downwards concave, and approaches the lines $z = a$, $y = b$. When $n = 2$ the curve is part of an ellipse. To obtain the full crown profile, the curve is reflected in the z-axis. Horn proposed the use of one set of parameters to model the sun crown profile, and another to model the shade crown profile, with the same value for the horizontal radius, b, in both sets.

We extend Horn's idea into three dimensions and model a crown envelope with a *generalized ellipsoid of revolution* (GER):

$$\frac{z^n}{a^n} + \frac{(x^2 + y^2)^{\frac{n}{2}}}{b^n} = 1. \tag{4.2}$$

The modelled crown volume is bound by this surface and the xy-plane of $\{C\}$. This model can be made to closely fit general graphical descriptions of crown form, such as those presented in (Hosie 1969). Figure 4.4 illustrates several crown envelopes defined by Equation 4.2.

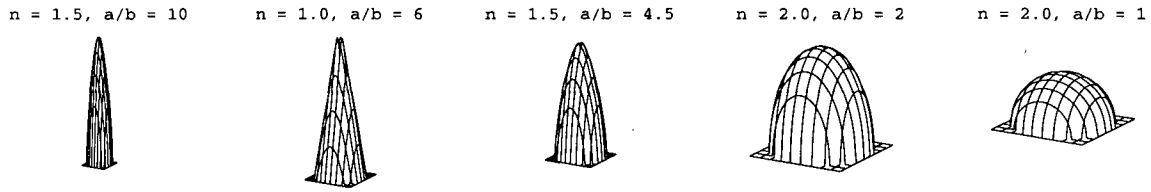


Figure 4.4: Crown envelope examples.

4.1.2 Sensing Geometry

We define a sensor Cartesian coordinate frame $\{S\}$ by the unit vectors $\{\mathbf{x}_s, \mathbf{y}_s, \mathbf{z}_s\}$ such that the system is right-handed, $\mathbf{x}_s$ points in fore direction of the platform and $\mathbf{y}_s$ points in the port direction. The orientation of $\{S\}$ relative to $\{C\}$ is represented by the rotation matrix $\mathbf{R}$, and the position of the origin of $\{S\}$ with respect to $\{C\}$ is represented by the translation vector $\mathbf{t}$. Both $\mathbf{R}$ and $\mathbf{t}$ are a function of time. The *projection vector* $\mathbf{p}$ is the vector joining the apex of a tree crown and the origin of $\{S\}$, so

$$\mathbf{p} = \mathbf{t} - h\,\mathbf{z}_c$$

where h is the z_c coordinate of the crown apex.

The projection of a particular tree crown onto the image plane is modelled with a scaled orthographic projection of the tree crown in the direction of $\mathbf{p}$ onto a plane perpendicular to $\mathbf{p}$. The scale of the image is computed as f/H, where H is distance between the optical centre and the scene along the optical axis (this is the flying altitude above the scene when the optical axis is vertical and the ground is flat and level). This projection simplifies the sensor irradiance model (this will be made clear in Section 4.2.3), and it approximates the perspective projection geometry of both cameras and pushbroom scanners that have an optical centre located at the origin of $\{S\}$, an optical axis aligned with $\mathbf{z}_s$, and an image plane parallel to the $x_s y_s$-plane and

intersecting the z_s axis at f. The quality of the approximation is related to the angle between $\mathbf{p}$ and $\mathbf{z}_s$, as will be described later in this section.

Relief displacement is the distance between the images of two separate points in the scene that lie on the same vertical axis. In a vertical perspective projection of a scene, a given scene object with a non-zero vertical dimension will exhibit increasing relief displacement with increasing distance from the intersection of the optical axis and the image plane (the *optical axis image point*). Relief displacement also increases with an increase in the ratio of object height to flying height. In the large to medium scale vertical aerial photographs commonly used in forest type mapping, radial relief displacement has a significant effect on the appearance of individual tree crowns.

In aerial image acquisition with a pushbroom scanner the *scanline* is parallel to the detector array axis and the *flightline* is parallel to the flight velocity. We assume that the detector array axis is parallel to $\mathbf{y}_s$, and we consider the scanline and flightline to map to the along-row and along-column dimensions of the image, respectively (in this discussion, we assume that the image has not been subjected to a geometric transformation, including any transformation intended to compensate for platform instability). A pushbroom scanner image is a perspective projection of the scene only in the along-row dimension. When the platform velocity is constant and perpendicular to the scanline, the image is a scaled orthographic projection of the scene in the along-column dimension. Therefore, a pushbroom scanner image that has been acquired with a vertical optical axis exhibits relief displacement only in the along-row dimension.

Figure 4.5 shows that along-row relief displacement has a noticeable effect on the appearance of individual tree crowns in the MEIS II image data that were used in the experiment (the sensing situation and the image data are described in Chapter 6).

Figure 4.6a represents the vertical perspective projection of a vertical scene object. In this figure f and O are, respectively, the lens focal length and optical center, A and A' are, respectively, the top and bottom of a vertical object, a and a' are, respectively, the projections of A and A' onto a horizontal image plane, h and H are, respectively, the height of the object

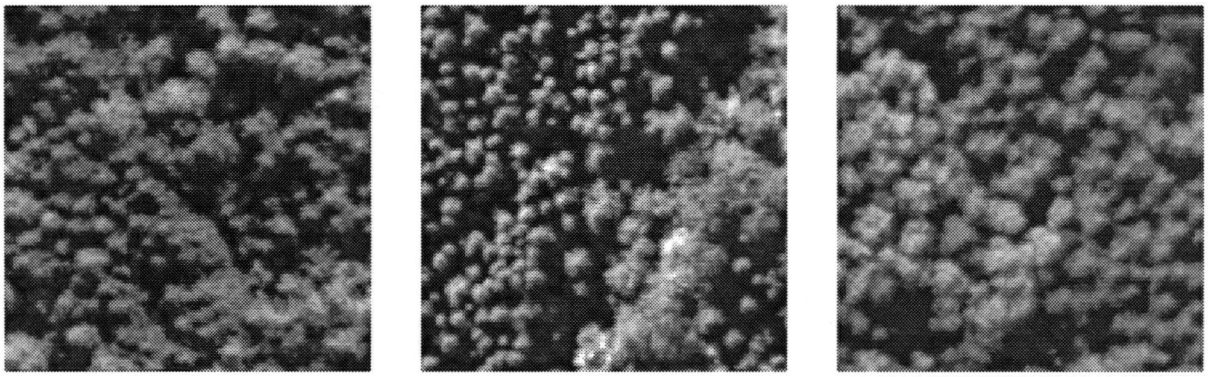


Figure 4.5: MEIS II subimages that show tree crown relief displacement. The subimages were extracted from a visible green wavelength channel, each subimage has 125 rows and columns, and the rows are aligned with the page width dimension. The middle subimage intersects the middle column of the flightline image, and the left and right subimages are parts of the flightline image near the opposite ends of the image rows. Because of relief displacement, the trees in the left and right subimages appear to "tilt" respectively to the left and right, and the dark between-tree gaps are less visible in these subimages than in the middle subimage.

and the height of the lens optical centre, and r and R are, respectively, the distance between a' and the image plane nadir and A and the ground nadir. The relief displacement d is the distance between a and a'.

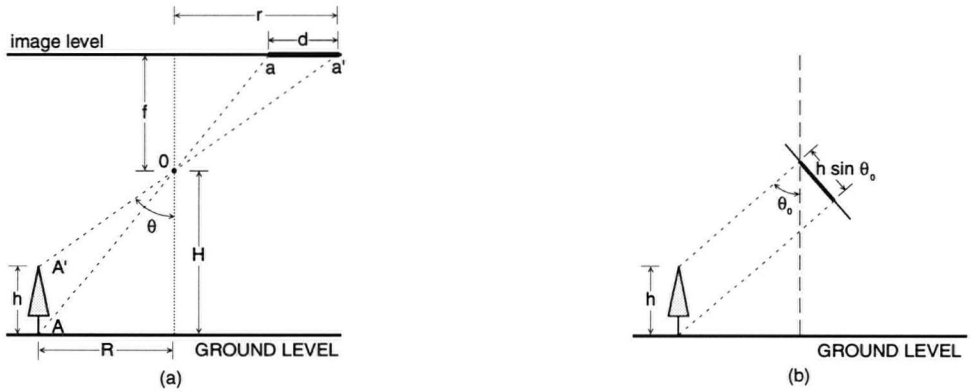


Figure 4.6: Vertical perspective projection (a) and orthographic projection (b).

By similar triangles, $r(H-h) = Rf$ and $H(r-d) = Rf$. Equating the left hand sides of these two equations gives us the well-known formula for relief displacement in vertical aerial photographs:

$$d = \frac{rh}{H}.$$

Alternatively,

$$d = Sh\frac{r}{f}$$

where S is the scale factor f/H.

Let θ be the angle between the line $\overline{A'O}$ and the optical axis. Then,

$$d = Sh\tan\theta.$$

In a scaled orthographic projection of a vertical object where θ_o is the angle between the view axis to the object and the vertical axis (Figure 4.6b) the image distance between the top and bottom of the object, d_o, is given by

$$d_o = Sh\sin\theta_o$$

Therefore, when θ is small, the projection of the object onto the sensor's image plane may be closely approximated by a scaled orthographic projection where $\theta_o = \theta$ (and $d_o/d = \cos\theta$). For MEIS II along-scanline relief displacement the maximum value of θ is 19.57° (McColl, Neville, and Till 1983) and at this angle $d_o/d = 0.94$. The approximation may be improved by scaling the object projection in the direction of the relief displacement by a factor of $1/\cos\theta$.

The previous discussion assumes that the optical axis is vertical and that the scene terrain is flat and level. The relationship between a scaled local orthographic projection of a scene object and its perspective projection will not be as described above if these conditions are not met. In aerial image acquisition with a pushbroom scanner, a deviation of the optical axis from the vertical will cause a variation of the along-scanline dimension of the scene elements, a deviation due to aircraft pitch will result in a relief displacement component orthogonal to the scanline, and a flight velocity component that is not orthogonal to the scanline will skew the image.

4.2 Radiometric Aspects

4.2.1 Scene Irradiance

In the simplest scenario for natural scene irradiance, an irradiated scene object is on flat and level terrain, light reflected from the adjacent terrain and the object contributes insignificantly to the sky radiance, and the sky is cloud-free and clear (*i.e.,* with little haze). Under these conditions the *total irradiance* of a horizontal surface (E_{tot}) is the the sum of the irradiance due to direct sunlight (the *direct irradiance*, E_{dir}), and of the irradiance due to the scattering of direct sunlight by the atmosphere (the *diffuse irradiance*, E_{dif}). Although the atmosphere also emits radiation, this is negligible when sensing radiation with a wavelength shorter than about $3\,\mu$m (Silva 1978).

One component of the image model is a model of the scene irradiance under the conditions described above. The parameters of this model are the components of the unit vector $\mathbf{s}$ that points from the origin of $\{C\}$ towards the centre of the sun and the diffuse fraction of the total horizontal irradiance $K = E_{\mathrm{dif}}/E_{\mathrm{tot}}$. The components of $\mathbf{s}$ are derived from the value for the angle θ_{s} between $\mathbf{s}$ and $\mathbf{z}_{\mathrm{c}}$ (the *solar zenith angle*), and the clockwise (as seen from above) angle ϕ_{s} between the projection of $\mathbf{s}$ on the $x_{\mathrm{c}}y_{\mathrm{c}}$-plane and x_{c} (the *solar azimuth*).

The output of the scene irradiance model is a set of collimated beams. Each beam has a direction and a perpendicular irradiance value that is normalized with respect to E_{tot}. One of the beams represents the direct irradiance. The direction of this beam is $(\theta_{\mathrm{s}}, \phi_{\mathrm{s}})$ and its normalized perpendicular irradiance value is $(1 - K)/\cos\theta_s$. The remaining beams represent the diffuse irradiance.

The method of discretizing the distribution of the diffuse atmospheric radiation is taken from Woodham and Gray (1987). The sky hemisphere is discretized into 109 approximately circular, equal-area patches by mapping a 109-tile uniform hexagonal tessellation of the $x_{\mathrm{c}}y_{\mathrm{c}}$-plane that is centred at the origin of $\{C\}$ onto the hemisphere with the azimuthal equal-area projection. Each diffuse irradiance beam is associated with one patch. The direction of diffuse

irradiance beam i is the direction (θ_i, ϕ_i) to the centre of the corresponding patch.

Let $L_{\text{dif}}(\theta_i, \phi_i, \theta_s, \phi_s)$ be the incident power per unit area per unit solid angle from an infinitesimal patch of the sky hemisphere centred along direction (θ_i, ϕ_i), when the sun is in position (θ_s, ϕ_s). Various models have been developed to provide estimates of $L_{\text{dif}}(\theta_i, \phi_i, \theta_s, \phi_s)/E_{\text{dif}}$, (denoted by $N(\theta_i, \phi_i, \theta_s, \phi_s)$) for shortwave radiation from clear skies (see Siala *et. al.* (1990) for a survey). Any such model can be used to evaluate the normalized perpendicular irradiance value, $P(\theta_i, \phi_i)$, associated with diffuse irradiance beam i.

We have not found an existing model of the hemispherical distribution of sky radiation that is wavelength-specific, or existing calibration data that makes a model wavelength-specific. However, the relative hemispherical distribution of clear sky radiation follows the same general pattern across the visible wavelength range: greatest in a circumsolar region, least in the quadrant opposite the sun, and increasing towards the horizon. Furthermore, clear sky radiation at wavelengths longer than the visible red and shorter than $0.3\,\mu$m is negligible since (1) 46 % of the total energy from solar irradiance above the earth's atmosphere falls between $0.40\,\mu$m and $0.76\,\mu$m, (2) Rayleigh scatter, which causes the great majority of clear sky diffuse radiation, is inversely proportional to the fourth power of wavelength, and (3) ozone absorption causes a cutoff in solar irradiation of the earth's surface at $0.3\,\mu$m (Fraser, Gaut, Reifenstein, and Sievering 1975).

If ω_p is the solid angle subtended at the origin by each sky hemisphere patch, then we have

$$P(\theta_i, \phi_i) = \omega_p K N(\theta_i, \phi_i) \tag{4.3}$$

(θ_s and ϕ_s are omitted from the notation for clarity). If we assume that the patches exactly cover the hemisphere, then

$$E_{\text{dif}} = \omega_p \sum_{j=1}^{109} L_{\text{dif}}(\theta_j, \phi_j) \cos\theta_j \tag{4.4}$$

Since $L_{\text{dif}}(\theta_j, \phi_j) = E_{\text{dif}} N(\theta_j, \phi_j)$, we have

$$E_{\text{dif}} = \omega_p E_{\text{dif}} \sum_{j=1}^{109} N(\theta_j, \phi_j) \cos\theta_j \tag{4.5}$$

and

$$\omega_{\mathrm{p}} = \frac{1}{\sum_{j=1}^{109} N(\theta_j, \phi_j) \cos\theta_j}. \tag{4.6}$$

The substitution of the right hand size of Equation 4.6 into Equation 4.3 yields

$$P(\theta_i, \phi_i) = \frac{KN(\theta_i, \phi_i)}{\sum_{j=1}^{109} N(\theta_j, \phi_j) \cos\theta_j}. \tag{4.7}$$

The effect of opaque obstacles on scene irradiance can be considered by setting N to zero for those beams that are intercepted by those obstacles (obstacles would also have an effect on the value of K).

The three-component continuous distribution (TCCD) model (Hooper, Brunger, and Chan 1987) is used to estimate $N(\theta_i, \phi_i, \theta_{\mathrm{s}}, \phi_{\mathrm{s}})$ and Equation 4.7 is used to evaluate $P(\theta_i, \phi_i)$. Relative to other models, the TCCD model is easy to use, accurate, and has been calibrated (for shortwave radiation) at a location close to PNFI.

4.2.2 Scene Irradiance and Tree Crown Interaction

The interaction of scene irradiance with a tree crown depends on both the shape and the physical composition of the crown. The simplest composition model is based on the assumption of an opaque surface. This assumption may be reasonable for young fairly open growing needle-leaved trees belonging to species with characteristically dense foliage (Albers, Strahler, Li, Liang, and Clarke 1990), or for broad-leaved trees with leaves that are concentrated near the crown envelope and are oriented to form a fairly continuous surface. However, an opaque surface assumption is difficult to justify in general, and in particular for mature trees and trees growing under competition. In the context of manual air-photo interpretation, Sayn-Wittgenstein describes variation between different crowns in the amount of photographic contrast between the part of a crown that is directly irradiated by the sun and the part that is not, and credits this variation mainly to differences in foliage density; he even suggests using this effect as a cue for distinguishing species with characteristically different foliage densities (for example, separating jack pine from spruce or balsam fir) (Sayn-Wittgenstein 1978).

These considerations have led to the development and examination of two models of the interaction of scene irradiance with a tree crown: an *opaque surface model* and a more complex *permeable volume model*. In both models, the crown volume is approximated by a subset of a three-dimensional *volume array* that consists of the array cells with centres inside a crown envelope defined by a generalized ellipsoid of revolution and and the $x_c y_c$-plane (the *crown cells*). A volume array cell has uniform side dimensions. The position and orientation of the volume array depends on the relative position and orientation of $\{S\}$ and $\{C\}$. This will be discussed further in Section 4.2.3.

A *surface cell* is a crown cell that is 6-connected to a non-crown cell, and whose centre is closer to the generalized ellipsoid of revolution part of the envelope than to the $x_c y_c$-plane. A surface cell is associated with a *surface point*, which is the point on the surface of the crown envelope that is closest to the centre of the surface cell. For each surface cell, the components of the outward unit normal vector at the corresponding surface point are determined.

In the opaque surface model, the contribution of an incident collimated beam of light to the radiance leaving a surface cell in a particular view direction is a function of the angle between the beam direction and the surface cell normal vector (the *incident angle*) and of the angle between the view direction and the normal vector (the *exitant angle*). The *Minnaert reflectance function* (Woodham and Gray 1987) is used:

$$L_r = \rho \frac{E_o(k+1)}{2\pi} \cos^k(i) \cos^{k-1}(e). \tag{4.8}$$

In this equation E_o is the perpendicular irradiance of a collimated beam of light with incident angle i, L_r is the observed radiance at exitant angle e, ρ is assigned a value in (0,1] that can be used to account for absorption, and k is assigned a value related to the intrinsic properties of the surface. Equation 4.8 is a physically plausible model of the reflectance of an opaque surface for $0 \leq k \leq 1$. This can be argued by considering the signs of the partial derivatives of L_r with respect to i and e and by reasoning that (1) with increasing i a given beam of light illuminates a larger surface area, so one should not observe an increase in brightness though a given aperture, and (2) with increasing e one sees more surface area through a given aperture, so should not

observe a decrease in brightness. Both i and e are restricted to $[0,\frac{\pi}{2})$. If i is outside this interval then the beam does not irradiate the surface point, and if e is outside this interval then the surface point is invisible. Under the Minnaert reflectance function, $L_r/(E_o\cos(i))$ is unchanged when the values of i and e are interchanged. This is a basic physical constraint for opaque surfaces called the *Helmholtz reciprocity condition* (Minnaert 1941).

The reflectance of many natural diffusely reflecting opaque surfaces can be accurately modelled with the Minnaert reflectance function. Values of k in the neighbourhood of 0.7 have been associated with loose porous surfaces, while values approaching 1 are characteristic of more compact surfaces (Young and Collins 1971). When $k = 1$ Equation 4.8 becomes the expression for the *Lambertian reflectance function.* In the opaque surface model implementation, the same user-specified value of k applies to all surface cells and ρ is arbitrarily set to 1 for all surface cells.

The simulated radiance of a surface cell in the opaque surface model is the sum of the contributions of all of the beams that comprise the simulated scene irradiance.

In the permeable volume model a tree crown is considered to be a volume of light-scattering elements. These elements are assumed to be randomly oriented and positioned, small relative to the size of the cells, and to have uniform size, shape, and radiometric characteristics. Within a crown cell there is a uniform density of light-scattering elements. Each crown cell is assigned a projected one-sided light-scattering element area density D according to a specified distribution (this is discussed in detail below). The value of D for a particular cell is assumed to be constant over all projection directions.

A light ray that is incident to a cell is either intercepted by a light-scattering element or passes through unintercepted. The assumption of small randomly oriented and positioned light-scattering elements leads to the use of Bouguer's law to compute the probability G that ray r will pass unintercepted through a cell when the path length of r through the cell is s (the *gap probability*):

$$G = \exp(-D\,s).$$

A beam of light impinging on a cell from direction (θ, ϕ) with normalized perpendicular irradiance P is represented by a pencil of regularly spaced parallel rays directed along (θ, ϕ). Each ray is the central axis of a portion of the beam. Each portion has the same cross-sectional area A. The normalized radiant flux Q from the portion of the beam represented by ray r that is intercepted by the cell is evaluated according to

$$Q = A\,P(1 - \exp(-D\,s)).$$

The constant A is arbitrarily set to 1 in the model implementation. The normalized perpendicular irradiance P' of the portion of the beam represented by r upon leaving the cell is evaluated according to

$$P' = P\,\exp(-D\,s).$$

Some of the non-intercepted light that reaches a crown cell may first pass through other crown cells. If a particular ray has to travel through cells $1, \ldots, n$ with density values $D_1, \ldots, D_n$ and individual cell path lengths $s_1, \ldots, s_n$, and the normalized perpendicular irradiance associated with the ray before entering cell 1 is P, then the normalized irradiance P' that is associated with the ray on leaving cell n is evaluated according to

$$P' = P\exp(-\sum_{i=1}^{n} D_i\,s_i).$$

The total simulated normalized radiant flux intercepted by a cell is computed as the sum of the contributions from all the beams that comprise the simulated scene irradiance, with each beam ray passing through the centre of the cell.

The permeable volume model is limited to single-scattering, in other words re-radiated flux from one cell that is intercepted and scattered by other cells is ignored. The unabsorbed intercepted flux for a cell (the flux that is transmitted and reflected by the light-scattering elements in the cell) is assumed to leave the cell as isotropically distributed flux originating from the centre of the cell. Therefore, the radiant intensity (power per unit solid angle) leaving a cell in the direction towards the viewer is directly proportional to the unabsorbed intercepted flux for the cell.

We assume that the significant light-scattering elements in a fully leaved tree crown are the leaves. As the size of a tree crown increases, the distribution of foliage within the crown generally becomes less uniform because the light that reaches an inner region of the crown falls below the compensation point of the leaves, which die and are not replaced (Jack and Long 1992). In response to this, the value of D for a crown cell is a function of the distance x from the centre of the cell to the closest point on the surface of the crown envelope, as defined by

$$D(x) = a_1 x^{a_2}(1-x)^{a_3}. \tag{4.9}$$

This function can provide a wide range of distributions, from nearly uniform to sharply peaked, depending on the values of a_1, a_2, and a_3. In Wang *et al.* (1990) a close fit to the measured radial horizontal and vertical leaf area density distributions for 26 Radiata pine trees with this function is demonstrated.

The values of a_1, a_2, and a_3 are derived from (1) the depth beyond which the projected one-sided light-scattering element area density is zero (X_{end}), (2) the depth of the maximum density (X_{max}), (3) the maximum density (D_{max}), and (4) the average density (D_{avg}). With a_2 and a_3 greater than zero, $D(x)$ in Equation 4.9 has a single maximum. Let b_1 and b_2 be, respectively, the value of this maximum and the corresponding value of x. Let b_3 to denote the value of the second derivative of $D(x)$ at b_2. It is possible to show that

$$\begin{aligned} a_2 &= \frac{-b_3 b_2^2(1-b_2)}{b_1} \\ a_3 &= \frac{-b_3 b_2(1-b_2)^2}{b_1} \\ a_1 &= \frac{b_1}{b_2^{a_2}(1-b_2)^{a_3}}. \end{aligned}$$

The interval $[0, X_{\text{end}}]$ is scaled to $[0,1]$, so that $b_2 = X_{\text{max}}/X_{\text{end}}$. Also, $b_1 = D_{\text{max}}$. The value of b_3 corresponding to D_{avg} is estimated with a root-finding procedure, which is possible because for fixed positive values of b_1 and b_2 the area under the curve defined by Equation 4.9 strictly decreases with an increase in the magnitude of b_3. Figure 4.7 illustrates some distributions produced by Equation 4.9.

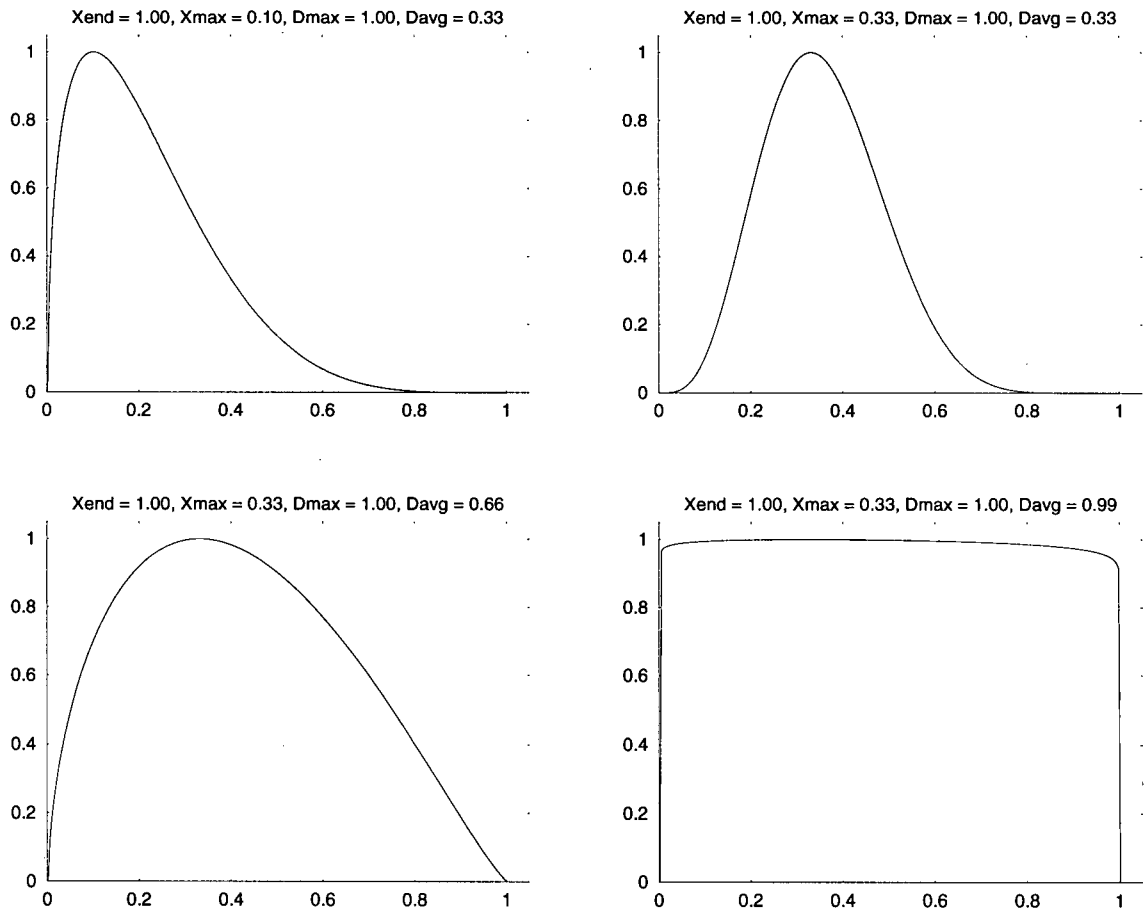


Figure 4.7: Light-scattering element area density distribution examples (X: horizontal axis, D: vertical axis)

Individual leaves generally have greater reflectance and transmittance in the near-infrared wavelength region (over 40 % for both) than in the visible region (under 20 % for both) (Gausman, Allen, Wiegand, Escobar, and Rodriguez 1971; Williams 1991). This suggests that both the opaque surface model and the permeable volume model should be more appropriate for visible wavelength sensing than for near-infrared wavelength sensing.

4.2.3 Sensor Irradiance

A sensor is modelled such that its optical centre is at the origin of $\{S\}$, its view axis coincides with the z_s-axis, and its image plane is parallel to the $x_s y_s$-plane and intersects the z_s-axis at the sensor focal length f. An image plane array is defined such that the along-column axis is parallel to the x_s-axis and the along-row axis is parallel to y_s-axis. The image plane array elements are square with side dimension p.

The offset of the optical axis image point from the nearest array element centre depends on the sensor that is being modelled. The MEIS II has 1024 image plane array row elements, so the optical axis image point lies on the boundary between column 511 and column 512, where the first column has index 0. A pushbroom sensor such as the MEIS II acquires an image one row at a time and is therefore more accurately described as having an 'image line' rather than an 'image plane.' However, we assume that the sensor moves in the $\mathbf{x}_s$ direction at a constant speed such that the spacing of the rows is the same as the spacing of the elements along the rows. This permits the definition of an image plane array with an indefinite number of rows.

The axes of the three-dimensional volume array that contains the crown cells are labelled a, b, and c. The cell dimension is $p\,H/f$, where H is the vertical distance between the origin of $\{S\}$ and the ground. This is the dimension of the projection of an image plane array element onto flat and level ground when the $x_s y_s$-plane is horizontal. The orientation of the volume array when $\mathbf{p}$ is oriented such that the axis through $\mathbf{p}$ intersects the point $(x_s = d_x, y_s = d_y, and z_s = f)$ on the image plane is defined by the following sequence:

1. Orient the volume array such that the a, b, and c-axes are parallel to $\mathbf{x}_s$, $\mathbf{y}_s$, and $\mathbf{z}_s$,

respectively, and define a rotation axis that intersects the origin of $\{S\}$, is in the $x_s y_s$-plane, and is perpendicular to the line joining $(0, 0, f)$ and (d_x, d_y, f) (see Figure 4.8a).

2. Rotate the volume array about the rotation axis until the c-axis is parallel to $\mathbf{p}$ (see Figure 4.8b).

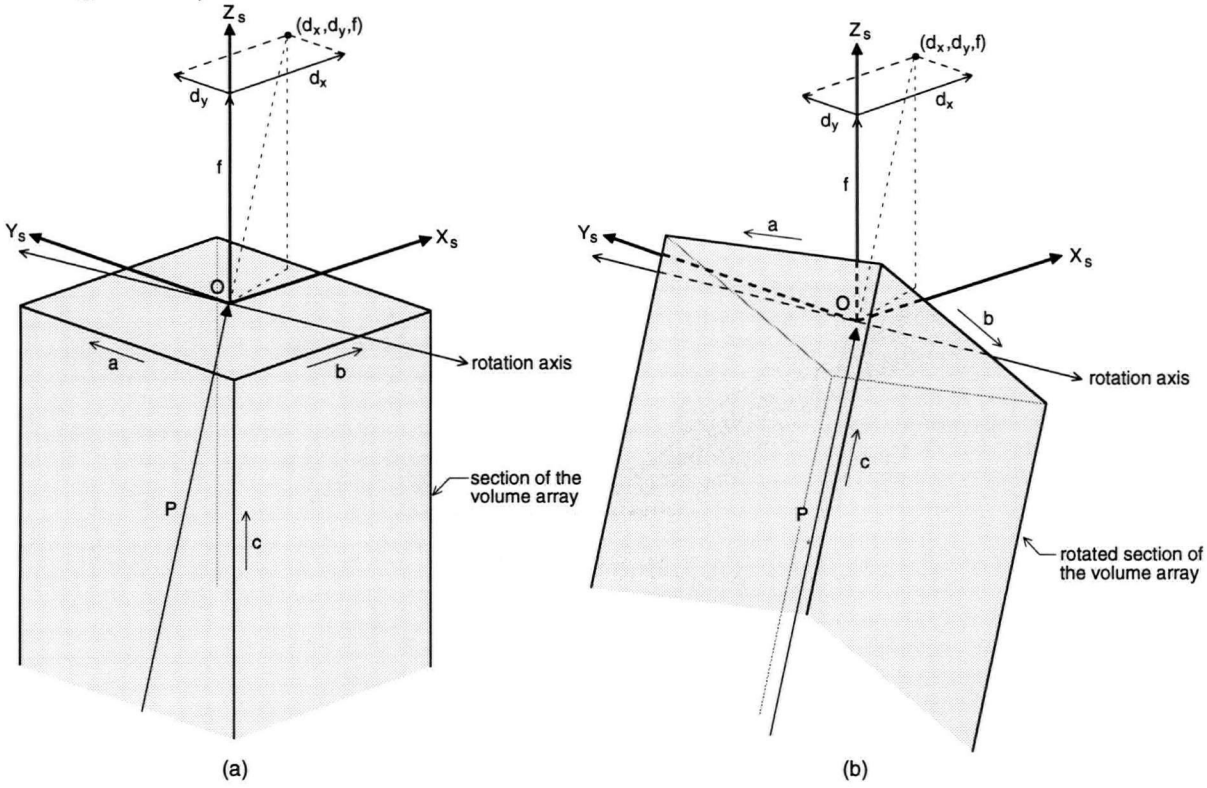


Figure 4.8: Volume array orientation (see text for explanation).

Additionally, the volume array is positioned such that in step (1) the a- and b-axis cell intervals coincide with intervals of the projection of the image plane array onto a plane that is perpendicular to the z_s-axis and intersects the z_s-axis at $-H$ (this is the ground plane when the $x_s y_s$-plane is horizontal and the ground is flat and level). An arbitrary 'slice' of the volume array (defined by a constant c-coordinate) between the scene and $\{S\}$ represents a sensor detector array. A detector array element corresponds to the image plane array element that projects to the detector array's ab-position at step (1). The mapping between the detector array elements and the image plane array elements is not changed by the rotation in step (2). This causes

the relationship of sensor irradiance to the modelled image of a tree crown to be based on the scaled local orthographic projection described in Section 4.1.2. For a pushbroom sensor, the d_x component of the offset is always zero.

In the opaque surface model, the exitant angle for a surface cell is the angle between the cell's normal vector and $\mathbf{p}$. The magnitude of the simulated irradiance of the corresponding detector element is set to the magnitude of the simulated surface cell radiance.

Let Q be the total intercepted flux for a crown cell in the permeable volume model and let α be the associated absorptance (the fraction of the intercepted flux that is absorbed by the light-scattering elements in the cell). If there are n crown cells between this cell and the detector element array with density values $D_1, \ldots, D_n$ and each with a side dimension c, then the contribution of the cell to the irradiance of the corresponding detector element is evaluated according to

$$V = (1 - \alpha)Q \exp(-c \sum_{i=1}^{n} D_i).$$

In the current model implementation, α is the same for all cells, so the factor $(1-\alpha)$ is arbitrarily set to 1.

In the permeable volume model, a background is represented by a two-dimensional array of volume cells occupying an *ab*-plane of the volume array. The background is positioned such that all of the crown volume cells are between it and the detector element array. Background radiance is simulated by assigning an intercepted flux value to background cells and including these cells in the computation of the sensor irradiance in the same manner as the crown cells. Only the background cells that have the same (a, b) coordinate of at least one crown cell are given an intercepted flux value, and the value is the same for all such cells. The intercepted flux value is a specified multiple B of the average of the detector irradiance values for the crown with no background radiance. Only the detectors that have the same (a, b) coordinate of at least one crown cell are involved in the computation of the background intercepted flux value (the irradiance value is undefined for the other detectors).

4.3 Model Instantiation

A pixel value in an image model instance is obtained by applying a gain and an offset to the magnitude of the simulated irradiance of the corresponding detector element. The gain and offset are uniform over the image and are evaluated so that the pixel values in a single model instance span the full value range of the specified pixel value representation. An unsigned byte pixel value representation was used in the main experiment.

The elements of the rotation matrix $\mathbf{R}$ that represents the orientation of $\{S\}$ with respect to $\{C\}$ (see Section 4.1.2) may be derived from many different specifications. One convenient approach that we have implemented is to use a *roll angle* (γ), a *pitch angle* (β), and a *flightline azimuth* (ϕ_f). An orientation that is specified with these angles is equivalent to the result of first aligning the axes of $\{S\}$ with the corresponding axes of of $\{C\}$ and then applying the following sequence of rotations to $\{S\}$: (1) about $\mathbf{x}_c$ by γ, (2) about $\mathbf{y}_c$ by β, and (3) about $\mathbf{z}_c$ by $-\phi_f$ (azimuth is a negative rotation angle in a right-handed coordinate frame).

The orientation of the projection vector $\mathbf{p}$ is a function of the image offset of the crown apex from the optical axis image point and of the sensor orientation. Pushbroom scanner image skew (due to aircraft yaw) is not modelled.

The parameters of the complete model are summarized as follows:

(1) sensing geometry:

- (1.1) f: lens focal length (metres)
- (1.2) p: image plane array element dimension (metres)
- (1.3) H: vertical height above ground level (metres)
- (1.4) γ: roll angle (degrees)
- (1.5) β: pitch angle (degrees)
- (1.6) ϕ_f: flightline azimuth (degrees)
- (1.7) d_r: image row offset of the crown apex from the optical axis image point (pixels)

(1.8) d_c: image column offset of the crown apex from the optical axis image point (pixels)

(2) scene irradiance[1]

(2.1) θ_s: solar zenith angle (degrees)

(2.2) ϕ_s: solar azimuth (degrees)

(2.3) K: diffuse fraction of the total horizontal irradiance (dimensionless)

(3) tree crown shape and size:

(3.1) a: generalized ellipsoid of revolution (GER) vertical radius (metres)

(3.2) b: GER horizontal radius (metres)

(3.3) n: GER exponent (dimensionless)

(4a) Minnaert reflectance function k (opaque surface model only) (dimensionless)

(4b) projected one-sided light-scattering element area density distribution (permeable volume model only):

(4.1) X_{end}: depth beyond which the density is zero (metres)

(4.2) X_{max}: depth of maximum density (metres)

(4.3) D_{max}: maximum density (metres^2/metres^3)

(4.4) D_{avg}: average density (metres^2/metres^3)

(5) B: factor for computing the simulated background radiant intensity (permeable volume model only) (dimensionless)

Values for f and p are derived from the sensor specifications, and are fixed. Values for H, γ, β, and α must be obtained during image acquisition. Ideally, these are constant over a single image. Values for θ_s and ϕ_s are readily calculated from the time and location; these values

[1]This list excludes calibration data for the adopted model of the hemispherical distribution of diffuse sky radiation.

vary continuously during a mission but the significance of the variation depends on the length and geographic extent of the mission. The value of K for a scene depends on the wavelength, relative optical path length, atmospheric conditions, and reflectance properties of the terrain (Iqbal 1983) and may be readily measured during image acquisition. Measurements of the value of K under various conditions are also available from the literature (*e.g.,* Oliver *et al.* (1975) and McDowell (1974)).

Knowledge of the scene and of the tree species expected to exist in the scene is required to evaluate the parameters related to tree crown shape, size, and composition. This is discussed further in Chapter 6.

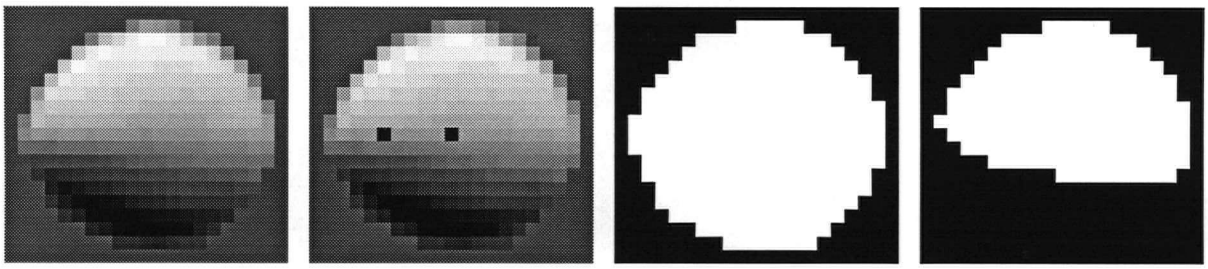
An instance of the image model is contained within a rectangular array of pixels. The origin pixel of a model instance array, as well as any scene image array, has coordinates $(0, 0)$. The following ancillary data is associated with the model instance array:

1. the values of the model instance parameters
2. the array coordinates of the pixel to which the crown apex projects (the *apex location*)
3. the array coordinates of the pixels that comprise the synthetic tree crown image (the *extent locations*)
4. the array coordinates of the pixel within which the centroid of the crown image extent locations lies (the *extent centroid location*)
5. the array coordinates of the pixels that indicate the visible extent of the directly irradiated portion of the surface of the crown model instance (the *direct irradiation locations*).

The crown apex and the extent centroid are *point features.*

Figure 4.9 illustrates an example of an image model instance and the associated point features locations. The apex location is to the left of the extent centroid location in the illustrated model instance because the projection vector is not vertical.

The direct irradiation locations in a model instance array correspond to visible directly irradiated surface cells. If the angle between the surface cell normal vector and the vector



(a) (b) (c) (d)

Figure 4.9: Tree crown image model instance (a), superimposed apex and extent centroid locations (b), extent locations (c), and direct irradiation locations (d).

pointing in the direction (θ_s, ϕ_s) from the origin of $\{C\}$ is less than $\frac{\pi}{2}$ then the corresponding surface cell is classified as directly irradiated.

4.4 Notation

The following notation will be used in the remainder of this dissertation:

- M_i: image model instance i, consisting of an image array and the ancillary data listed in Section 4.3
- I_i: the image array for M_i; the array has R_i rows and C_i columns
- $(r_{a,i}, c_{a,i})$: the (row, column) coordinates of the apex location for M_i
- $(r_{c,i}, c_{c,i})$: the (row, column) coordinates of the extent centroid location for M_i
- $\mathcal{L}_i$: a list of the (row, column) coordinates $(r_1, c_1), (r_2, c_2), \ldots, (r_{l_i}, c_{l_i})$ of l_i distinguished pixel locations in I_i; the locations are listed in ascending order primarily on the value of r and secondarily on the value of c so that $r_{j+1} \geq r_j$ and $c_{j+1} > c_j$ if $r_{j+1} = r_j$
- $\mathcal{L}_{e,i}$: list of the $l_{e,i}$ extent locations in I_i
- $\mathcal{L}_{d,i}$: list of the $l_{d,i}$ direct irradiation locations in I_i

- $\mathbf{e}^{(i)}$: vector of dimension $l_{\mathrm{e},i}$ evaluated such that $\mathbf{e}_j^{(i)} = I_i(r_j, c_j)$ where (r_j, c_j) is item j of $\mathcal{L}_{\mathrm{e},i}$.
- $\mathbf{d}^{(i)}$: vector of dimension $l_{\mathrm{d},i}$ evaluated such that $\mathbf{d}_j^{(i)} = I_i(r_j, c_j)$ where (r_j, c_j) is item j of $\mathcal{L}_{\mathrm{d},i}$.

The model instance index (i in the above list) is omitted where it is clearly implied by the context of discussion.

A model instance may be positioned with respect to a scene image array by specifying the scene image (row, column) array location that coincides with the location of a designated point feature (called the *registration feature*) in the model instance. The row and column axes of the model instance are assumed to always be aligned with, respectively, the row and column axes of the scene image array. Let (x, y) denote a scene image array (row, column) location and reg denote a registration feature (reg = a when the registration feature is the apex and reg = c when the registration feature is the extent centroid). The vectors of image values from scene image array S that coincide with the extent locations and the direct irradiation locations of M_i when this model instance is registered to (x, y) by reg are denoted by $\mathbf{s}^{(\mathrm{e}^{(i)},reg,x,y)}$ and $\mathbf{s}^{(\mathrm{d}^{(i)},reg,x,y)}$, respectively (items may be omitted from the superscripts where they are clearly implied by the context of the discussion). Vector $\mathbf{s}^{(\mathrm{e}^{(i)},reg,x,y)}$ has $l_{\mathrm{e},i}$ elements and vector $\mathbf{s}^{(\mathrm{d}^{(i)},reg,x,y)}$ has $l_{\mathrm{d},i}$ elements. The elements of these vectors are evaluated according to

$$\mathbf{s}_j^{(\mathrm{e}^{(i)},reg,x,y)} = S(x + r_j - r_{reg,i},\ y + c_j - c_{reg,i})$$

where (r_j, c_j) is item j of $\mathcal{L}_{\mathrm{e},i}$, and

$$\mathbf{s}_j^{(\mathrm{d}^{(i)},reg,x,y)} = S(x + r_j - r_{reg,i},\ y + c_j - c_{reg,i})$$

where (r_j, c_j) is item j of $\mathcal{L}_{\mathrm{d},i}$.

Subsequent discussions will also refer to *normalized* vectors that have an element mean value of zero and unit length. If $\mathbf{v}$ is a d-dimensional vector, then the normalized version, denoted $\hat{\mathbf{v}}$,

is computed according to

$$\mu = \frac{1}{d}\sum_{i=1}^{d} \mathbf{v}_i$$

and

$$\hat{\mathbf{v}} = \frac{\mathbf{v} - \mu\,\mathbf{1}}{\| \mathbf{v} - \mu\,\mathbf{1} \|}.$$

4.5 Extent Similarity

Consider a set of model instance image arrays $\mathrm{M} = \{M_1, \ldots, M_n\}$ in which all the elements have the same row and column dimensions and the same location for a designated point feature. The corresponding model instances are *aligned* on the designated point feature, which is referred to as the *alignment registration feature* (or simply as the *registration feature* when the context is clear). Let $\mathcal{L}_\cap$ be the list of $l_\cap$ locations that occur in all of $\mathcal{L}_{\mathrm{e},1}, \ldots, \mathcal{L}_{\mathrm{e},n}$, and let $\mathcal{L}_\cup$ be the list of $l_\cup$ locations that occur in at least one of $\mathcal{L}_{\mathrm{e},1}, \ldots, \mathcal{L}_{\mathrm{e},n}$. The *extent similarity* η_M for M is the ratio $l_\cap / l_\cup$.

Extent similarity is a value in $[0, 1]$; the value is 1 if and only if for all aligned image arrays each extent location is an extent location in all of the others, and is 0 if and only if no extent location in one of the aligned image arrays is an extent location in any other.

Two model instances, b and c, exhibit an *extent difference* if and only if their extent similarity $\eta_{\{b,c\}}$ is less than 1. Two model instances that are generated with different values for one or more of the parameters that define shape, size, and sensing geometry may exhibit an extent difference. If the sensing geometry parameter values for the different parameter sets are such that the $x_\mathrm{s}y_\mathrm{s}$-plane is horizontal and the projection vector is vertical, then two model instances that are generated from different values for a, n, or ϕ_f will not exhibit an extent difference. Also, model instance extents vary in discrete steps, so a parameter value change can be too small to cause an extent difference.

4.6 Pattern Correlation

Consider model instances a and b with an extent similarity of 1 and with image value vectors $\mathbf{e}^{(a)}$ and $\mathbf{e}^{(b)}$. The correlation coefficient for corresponding elements of $\mathbf{e}^{(a)}$ and $\mathbf{e}^{(b)}$ is the *pattern correlation* for the model instances. The correlation coefficient is a dimensionless value that ranges between -1 and 1 and measures the degree to which two variables are linearly related (Hogg and Craig 1978, § 2.3). Perfect negative and positive linear relationships are represented by correlation coefficients of -1 and 1, respectively.

The normalized vector $\hat{\mathbf{e}}^{(i)}$ is the *pattern* for model instance i. The pattern correlation for the two model instances a and b, denoted $\rho_{a,b}$, is the inner product of the corresponding patterns:

$$\rho_{a,b} = \hat{\mathbf{e}}^{(a)\mathrm{T}}\,\hat{\mathbf{e}}^{(b)}. \tag{4.10}$$

Pattern correlation is undefined for two model instances that exhibit an extent difference.

Two model instances exhibit a *pattern difference* if and only if their pattern correlation is less than 1. Two model instances that are generated from different composition models (opaque surface or permeable volume), or with different values for one or more of the parameters that define the scene irradiance and the light-scattering element area density (permeable volume model only), k (opaque surface model only), or B (permeable volume model only) may exhibit a pattern difference. If the sensing geometry parameter values for the different parameter sets are such that the $x_{\mathrm{s}}y_{\mathrm{s}}$-plane is horizontal and the projection vector is vertical, then two model instances that are generated from different values for a, n, or ϕ_{f} will not exhibit an extent difference and may exhibit a pattern difference. The model instance image values vary in discrete steps, so a parameter value change can be too small to cause a pattern difference.

Chapter 5

Procedures

Let M_i be an instance of the synthetic tree crown image model[1], reg be a specified registration feature (*i.e.*, the crown apex or the extent centroid), S be a scene image array that represents a single specified channel of the scene image data, and (x, y) be a location in S. The correlation coefficient

$$\rho_{M_i,S,reg,x,y} = \hat{\mathbf{e}}^{(i)\mathrm{T}}\,\hat{\mathbf{s}}^{(\mathbf{e}^{(i)},reg,x,y)}$$

is used as a measure of the strength of the match between M_i and the scene image in the shared initial matching step of the training and recognition procedures when M_i is positioned at (x, y) (the notation used here is described in Section 4.4). This is appropriate because the model is intended to be valid only up to, but not including, the application of a uniform gain and offset to the scene image, and variations in uniform gain and offset do not affect $\rho_{M_i,S,reg,x,y}$. Components of the subscript for ρ will be omitted where they are clearly understood from context or are irrelevant to the discussion.

The initial matching step consists of searching the scene image spatial extent and the model parameter space for locations where ρ is at a local maximum and above a liberally-specified threshold. The search region of the model parameter space is defined by the user-specified parameter value intervals. The amount of computation involved in the generation of a model instance has motivated the use of precomputed model instances in the initial matching procedure. The organization and generation of these model instances is described in Section 5.1.

[1]The synthetic tree crown image model will be referred to simply as the 'model' in the remainder of this chapter where the context excludes example-based model components from the discussion.

The representation of the recognition instances that are produced by the initial matching procedure and filtered by the remaining procedures is described in Section 5.2. The procedures for the initial matching step, the evaluation of correct recognition probability, and the resolution of interpretation conflicts are respectively described in Sections 5.3, 5.4, and 5.5.

5.1 Model Instance Organization and Generation

The scene image extent is partitioned into p *recognition regions*

$$\mathrm{RECREG}_1, \ldots, \mathrm{RECREG}_p$$

within which the sensing geometry, scene irradiance, and constraints on the various tree properties that correspond to the model parameters are considered to be practically constant. The recognition region RECREG_i is associated with the *model instance set* MISET_i. Each model instance in MISET_i incorporates the values for the sensing geometry and scene irradiance parameters that are associated with RECREG_i, and only model instances from MISET_i are involved in the final tree crown recognition instances within RECREG_i.

The model instance set MISET_i is partitioned into q_i *uniform-extent subsets*:

$$\mathrm{MISET}_i = \mathrm{UESSET}_{i,1} \cup \cdots \cup \mathrm{UESSET}_{i,q_i}.$$

The uniform-extent subset $\mathrm{UESSET}_{i,j}$ consists of $r_{i,j}$ model instances:

$$\mathrm{UESSET}_{i,j} = \{M_{i,j,1}, \ldots, M_{i,j,r_{i,j}}\}.$$

The elements of a uniform-extent subset are such that when they are aligned on a specified registration feature (*i.e.*, the crown apex or the extent centroid) the subset has an extent similarity of 1 (see Section 4.5 for the definition of extent similarity).

The different uniform-extent subsets represent the modelled tree crown spatial extent variation within the image-space: any two elements drawn from different uniform-extent subsets exhibit an extent difference when they are aligned on the specified registration feature (see

Section 4.5 for a definition of extent difference). The different model instances within the same uniform-extent subset represent the modelled pattern variation associated with a fixed image-space tree crown spatial extent: any two elements of the same uniform-extent subset exhibit a pattern difference (see Section 4.6 for the definition of pattern difference).

The procedure that we use to generate a model instance set is described in Figure 5.10. The procedure generates random samples of model instances that respect the specified constraints on the parameter values, and constructs the model instance set from relatively small subsets of the sample sets. The model instance set is representative of the sample sets in the sense that all elements of the sample sets are similar to one or more elements of model instance set with respect to extent and pattern, where the required degree of similarity is expressed by the values of $\eta_1, \ldots, \eta_s$ and ρ_{diff}. The evaluation of the inputs to the procedure is discussed in Section 5.1.1, and an algorithm that accomplishes steps (3.2) and (3.4.4) of the procedure is described in Section 5.1.2.

5.1.1 Evaluation of Inputs to Model Instance Set Generation

The evaluation of the inputs to the model instance set generation procedure involves a certain amount of experimentation and adaptation to the scene. Therefore, details of the process are best described with an example, one of which is provided in Section 6.3. In this section we provide an outline of the process. The basic steps, which are discussed in further detail in the remainder of this section, are the following:

1. Define constraints on the locally variable parameters for the overall scene.
2. Evaluate $b_1, \ldots, b_s$, and $\eta_1, \ldots, \eta_s$.
3. Evaluate ρ_{diff}, SS_{E} and SS_{M}.
4. Determine a partitioning of the scene image extent into initial recognition regions.
5. If warranted by the scene knowledge, specialize the constraints on the locally variable parameters to different parts of the scene and subdivide the recognition regions accordingly.

Input:

- fixed values for the sensing geometry and scene irradiance parameters for RECREG_i
- constraints on the values for the locally variable parameters RECREG_i
- a set of discrete values for b (the GER horizontal radius) that respect the specified constraints and are denoted $b_1, \ldots, b_s$, and corresponding extent similarity values $\eta_1, \ldots, \eta_s$
- SS_{E},
- SS_{M}, ρ_{diff}

Output: MISET_i

(2) Let $\text{MISET}_i = \emptyset$ and j = 1

(3) For $g = 1, \ldots, s$:

(3.1) Generate a set E containing a random sample of SS_{E} model instance extents within the constraints that have been defined for $(b_g, \frac{a}{b}, n)$ (a is the GER vertical radius and n is the GER exponent).

(3.2) Find an approximately minimal subset E′ of E such that for any element e of E there is an element e' of E′ such that $\eta_{\{e,e'\}} \geq \eta_g$.

(3.3) Let $t = |\text{E}'|$ and denote the elements of E′ by $e'_1, \ldots, e'_t$. Partition the elements of E into t disjoint subsets $\text{E}_1, \ldots, \text{E}_t$ such that for $h = 1, \ldots, t$, $e'_h \in \text{E}_h$ and $\eta_{\text{E}_h} = 1$.

(3.4) For $h = 1, \ldots, t$:

(3.4.1) Let $u = |\text{E}_h|$ and denote the elements of E_h by $e_1, \ldots, e_u$.

(3.4.2) Let $\text{M} = \emptyset$.

(3.4.3) For $k = 1, \ldots, u$:

(3.4.3.1) Generate a set M_k containing a random sample of $\lceil SS_{\text{M}}/u \rceil$ model instances within the constraints that have been defined for the locally-variable parameter values, and with the additional constraint that the value for $(b, \frac{a}{b}, n)$ is that of e_k.

(3.4.3.2) Let $\text{M} = \text{M} \cup \text{M}_k$.

(3.4.4) Find an approximately minimal subset M′ of M such that for any element M of M there is an element M' of M′ for which $\rho_{M,M'} \geq \rho_{\text{diff}}$.

(3.4.5) Let $\text{UESSET}_{i,j} = \text{M}'$, $\text{MISET}_i = \text{MISET}_i \cup \text{UESSET}_{i,j}$, and $j = j + 1$.

Figure 5.10: Procedure to generate a model instance set.

Within a particular recognition region, the model parameters related to sensing geometry and scene irradiance are considered to be fixed. The remaining parameters are *locally variable* since they are related to tree crown shape, size, and composition, and to background radiant intensity (*cf.* Section 4.3). The evaluation of the constraints that can be placed on the locally variable parameters will require knowledge of the scene and of the morphology of the tree species that are expected to occur in the scene. In Canada, information on expected species and stand types can be derived from knowledge of the one or more forest regions and sections (as described in Rowe (1972)) in which the scene is located. Useful published information on tree morphology for individual species is also available.

Since ρ is a function of the pixels comprising the extent of a model instance and the coinciding image region, we convert the desired precision for estimates of the radius of large crowns, σ, into a required minimum extent similarity η_{adj} of two model instances with $b = b_i$ and $b = b_{i+1}$. We make the assumption that the sensing geometry is such that most of the applicable model instances will have circular or near-circular extents, and evaluate η_{adj} to the extent similarity of two concentric circular regions with radii b_{max} and $b_{\text{max}} - \sigma$:

$$\eta_{\text{adj}} = (b_{\text{max}} - \sigma)^2 / b_{\text{max}}^2. \tag{5.11}$$

We then use η_{adj} to generate $b_1, b_2, \ldots, b_s$ according to the algorithm described in Figure 5.11. This algorithm was designed in recognition of the fact that for small i, $b_{i+1} - b_i$, where $b_{i+1} = \sqrt{b_i^2 / \eta_{\text{adj}}}$, may be too small relative to the GPPD to be useful. Therefore, the algorithm accepts a minimum b-value increment $(\triangle b)_{\text{min}}$.

We intend that a model instance for which $b = b_i$ will match images of tree crowns with radii approximately in the range $[b_i - x_1, b_i + x_2]$, where

$$x_1 = y_1 \, (\triangle b)_1, \; (\triangle b)_1 = b_i - b_{i-1}$$

$$x_2 = y_2 \, (\triangle b)_2, \; (\triangle b)_2 = b_{i+1} - b_i$$

Input: $b_{\min}$, $b_{\max}$, η_{adj}, $(\triangle b)_{\min}$

Output: $b_1, b_2, \ldots$

(1) Let $i = 1$, $b_i = b_{\min}$.

(2) If $b_i > b_{\max}$ then halt and return $b_1, \ldots, b_{i-1}$.

(3) Let $b_{i+1} = \sqrt{b_i^2/\eta_{\text{adj}}}$.

(4) If $b_{i+1} - b_i < (\triangle b)_{\min}$:

(4.1) Let $b_{i+1} = b_i + (\triangle b)_{\min}$.

(5) Let $i = i + 1$

(6) Go to (2).

Figure 5.11: Algorithm to compute a sequence of GER horizontal radius values.

and y_1, y_2 are evaluated so that x_1 and x_2 respectively define the extent difference midpoint for $[b_{i-1}, b_i]$ and $[b_i, b_{i+1}]$:

$$\frac{(b_i - y_1(\triangle b)_1)^2}{b_i^2} = \frac{b_{i-1}^2}{(b_i - y_1(\triangle b)_i)^2} = \frac{(b_i - (\triangle b)_1)^2}{b_i - (y_1(\triangle b)_1)^2} \longrightarrow y_1 = \frac{b_i - \sqrt{b_i^2 - b_i(\triangle b)_1}}{(\triangle b_1)}$$

$$\frac{b_i^2}{(b_i + y_2(\triangle b)_2)^2} = \frac{(b_i + y_2(\triangle b)_2)^2}{b_{i+1}^2} = \frac{(b_i + y_2(\triangle b)_2)^2}{(b_i + (\triangle b)_2)^2} \longrightarrow y_2 = \frac{-b_i + \sqrt{b_i^2 + b_i(\triangle b)_2}}{(\triangle b_2)}.$$

We then evaluate η_i according to

$$\eta_i = \begin{cases} \frac{b_i^2}{(b_i + x_2)^2} & \text{if } i = 1 \\ \frac{(b_i - x_1)^2}{b_i^2} & \text{if } i = s \\ \left[\frac{(b_i - x_1)^2}{b_i^2} + \frac{b_i^2}{(b_i + x_2)^2}\right]/2 & \text{otherwise} \end{cases}$$

We evaluate ρ_{diff} to an estimate of the value of ρ in cases where the pattern difference between a model instance and an image region is only due to unmodelled variation, or "noise." This is done by selecting a sample of tree crown images that have an exceptionally regular appearance (*e.g.*, relatively little within-crown image texture, and a relatively smooth image boundary), and, for each element of the sample, computing ρ with each element of a large sample

of model instances. The model instances are positioned by registering the model instance apex feature to the manually estimated apex locations in the tree crown images. The model instances are generated within the specified constraints on the locally-variable parameter values and with values for the sensing geometry and scene irradiance that are applicable to the tree crown image. The estimate for ρ_{diff} is set to a value that is no less than the maximum value of ρ obtained for each of the tree crown images.

The values SS_{E} and SS_{M} determine the size of the sample sets E and M_k computed in steps (3.1) and (3.4.3.1) of the procedure to generate a model instance set (Figure 5.10). Experimentally determined minimum values for SS_{E} and SS_{M} at which the sizes of the E′ and M′ subsets are maximum for the specified values of $\eta_1, \ldots, \eta_s$ and ρ_{diff} are used.

The initial recognition regions are intended to comprise a partitioning of the spatial extent of the image such that the sensing geometry and the scene irradiation can be regarded as practically constant over each region. The sensing geometry and scene irradiation parameter values that are associated with an initial recognition region are estimates of the averages for those values over the region. We define the recognition regions experimentally, using sensitivity tests to help judge whether or not the deviation of the sensing geometry and scene irradiance parameter values from the average values over the recognition regions represent model instance extent or pattern dissimilarities that exceed the limits expressed by $\eta_1, \ldots, \eta_s$ and ρ_{diff}.

A specialization of the constraints on the locally variable parameters to different parts of the scene, and a refinement of the recognition regions through subdivision, might be derived from further local knowledge of the terrain and of pre-existing stand types as recorded in old surveys. In many cases where knowledge of the scene is broad (at the level of knowing the forest region and section) and the overall spatial extent of the scene data is limited to the area covered by a single aerial image acquisition mission we can expect that there will be little grounds for varying the constraints over the scene.

5.1.2 Computing an Approximately Minimal Dominating Set

Steps (3.2) and (3.4.4) of the procedure to generate a model instance set both map to the problem of finding a minimal *dominating set* of a graph. A dominating set of a graph G that consists of the set of vertices V and the set of edges A is a subset V′ of V such that any vertex of G that is not in V′ is connected to one or more elements of V′ by a path of length 1. In the case of step (3.2) there is a one-to-one correspondence between elements of E and elements of V ($e_i \leftrightarrow v_i$), there is an edge joining v_i and v_j if and only if $\eta_{\{e_i,e_j\}} \geq \eta_{\text{diff}}$, and finding an approximately minimal E′ is equivalent to finding an approximately minimal V′. In the case of step (3.4.4) there is a one-to-one correspondence between elements of M and elements of V ($M_i \leftrightarrow v_i$), there is an edge joining v_i and v_j if and only if $\rho_{M_i,M_j} \geq \rho_{\text{diff}}$, and finding an approximately minimal M′ is equivalent to finding an approximately minimal V′.

The procedure to generate a model instance set involves the objectives of finding approximately minimal subsets E′ and M′ because the problem of finding a minimal dominating set of a graph is known to be NP-complete. The minimal dominating set problem is approximable within a factor of $1 + \ln |\text{V}|$ by means of a specialization of the greedy approximation algorithm for the minimal set covering problem described in Johnson (1974). This algorithm to find an approximately minimal dominating set of a graph is described in Figure 5.12, and is implementable in time $O(n^2)$ (or in time $O(n \log n)$ if step (2) is performed separately, as part of the process of constructing $\langle \text{V}, \text{A} \rangle$ out of E or M). This algorithm gives reasonable results in practice (*cf.* Section 6.3.6).

5.2 Representation of Recognition Instances

The location of a recognition instance in a scene image array is the scene image (row, column) array location that coincides with the specified registration feature of the model instance involved in the recognition instance.

At most, one element of a uniform-extent subset can be involved in a recognition instance

Input: $\langle \mathrm{V}, \mathrm{A} \rangle$

Output: $\mathrm{V}' : \forall v_i \in \mathrm{V},\ \exists v_j \in \mathrm{V}'$ such that $(v_i, v_j) \in \mathrm{A}$ and $|\mathrm{V}'|$ is within a factor of $1 + \ln |\mathrm{V}|$ of the minimum

(1) Let $\mathrm{V}' = \emptyset$, UNCOV = V, $n = |\mathrm{V}|$.

(2) For $i = 1, \ldots, n$:

  (2.1) Let $\mathrm{SET}[i] = \{v_i\}$.

    (2.1.1) For $j = 1, \ldots, n$:

      (2.1.1.1) If $(v_i, v_j) \in \mathrm{A}$ and $v_j \notin \mathrm{SET}[i]$ then let $\mathrm{SET}[i] = \mathrm{SET}[i] \cup \{v_j\}$.

(3) If UNCOV $= \emptyset$ then halt and return V':

(4) Choose j, $1 \leq j \leq n$, such that $|\mathrm{SET}[j]|$ is maximized.

(5) Let $\mathrm{V}' = \mathrm{V}' \cup \{v_j\}$, UNCOV = UNCOV $-$ SET$[j]$.

(6) For $i = 1, \ldots, n$:

  (6.1) Let $\mathrm{SET}[i] = \mathrm{SET}[i] - \mathrm{SET}[j]$.

(7) Go to (3).

Figure 5.12: Algorithm to find an approximately minimal dominating set of a graph (adapted from the algorithm to compute an approximately minimal set cover in Johnson (1974)).

at a particular location.

One *identifier map* is defined for each uniform-extent subset. The identifier maps and the uniform-extent subsets together represent the recognition instances at any stage of the recognition procedure.

The identifier map for a particular uniform-extent subset is only required to cover the associated recognition region. Since different recognition regions may have different numbers of associated uniform-extent subsets, they may also have different numbers of identifier maps.

In order to simplify the following discussions, we will assume that recognition regions are rectangular and that the diagonally opposite corner elements of the maps coincide with scene image locations $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$, where $x_{\max} \geq x_{\min}$ and $y_{\max} \geq y_{\min}$.

Let UESSET $= \{M_1, \ldots, M_r\}$ be a uniform-extent subset, ID be the associated identifier map, S be the scene image array, and (x, y) be a location within S that maps to a location within ID (in this discussion, components are omitted from subscripts for clarity). If some $M_i \in$ UESSET is involved in a recognition instance at (x, y) then $ID\,(x - x_{\min}, y - y_{\min}) = i$; otherwise, $ID\,(x - x_{\min}, y - y_{\min}) =$ null.

5.3 Matching Image Locations and Model Instances

The objective of the procedure for matching image locations and model instances is to produce for each recognition region RECREG_i, $i = 1, \ldots, p$, a set of identifier maps $\text{IDSET}_i = \{\text{ID}_{i,1}, \ldots, \text{ID}_{i,q_i}\}$ such that $\text{ID}_{i,j}$ represents initial recognition instances that involve model instances from $\text{UESSET}_{i,j}$.

An outline of the procedure that was used in the main experiment is presented in Figure 5.15. This procedure will be explained as a refinement of those outlined in Figures 5.13 and 5.14.

Figure 5.13 illustrates the basic approach. For each uniform-extent subset and each image location within the recognition region find a model instance that has the maximum correlation coefficient with the coincident image region when positioned at the image location. If that correlation coefficient is no less than a specified threshold ρ_{thr}, then evaluate the coinciding

location of the identifier map for the uniform-extent subset to the identifier *id* of the selected model instance; otherwise, evaluate it to null. The threshold ρ_{thr} is liberally specified (false recognition instances are eliminated in later processing steps) and confirmed through experimentation. The correlation coefficients associated with the selected model instances identified in $\text{ID}_{i,1}, \ldots, \text{ID}_{i,q_i}$ are retained in parallel maps $\text{RHO}_1, \ldots, \text{RHO}_{q_i}$. Locations in the RHO-maps that do not have locally maximum values are evaluated to null in the corresponding ID-maps. For a given RHO-map, the neighbourhood within which local maxima are searched for is defined by the extent of the associated uniform-extent subset. The procedure that implements this non-maximum suppression operation is outlined in Figure 5.16.

The basic procedure that is illustrated in Figure 5.13 is essentially an exhaustive search through the image-space and the uniform-extent subsets. We decided to investigate the possibility of reducing the amount of computation involved in the matching procedure by exploiting the fact that a local maximum in a correlation surface tends to be the peak of a 'hill' whose spatial extent is related to that associated with the the uniform-extent subset (model extents that are relatively large in a particular direction tend to produce hills that relatively wide in that direction) (*cf.* Section 6.4). The second version of the procedure, which is outlined in Figure 5.14, implements a refinement to support this investigation. This procedure has parameters $dfrac$ and ρ_{sthr}, in addition to those of the basic procedure. The image-space is searched for matches using two resolution levels. At the first resolution level, the x and y dimensions are searched in increments of d_x and d_y, respectively, where $d_x = \max(dfrac \times dim_x, 1)$ and $d_y = \max(dfrac \times dim_y, 1)$, and dim_x and dim_y are, respectively, the overall x- and y-dimensions of the extent associated with the current uniform-extent subset. If in this search a location $(x_\circ, y_\circ)$ is found such that an element of the current uniform-extent subset has a correlation coefficient with the coincident image region that is no less than ρ_{sthr}, then each location of a neighbourhood of $(x_\circ, y_\circ)$ is tested (this represents the second resolution level). The neighbourhood is defined by $x_\circ - d_x/2 \leq x \leq x_\circ + d_x/2$ and $y_\circ - d_y/2 \leq y \leq y_\circ + d_y/2$. An experimentally confirmed value is specified for ρ_{sthr}.

Input:

- S: single channel scene image
- MISET_i: model instance set such that $\text{MISET}_i = \text{UESSET}_{i,1} \cup \cdots \cup \text{UESSET}_{i,q_i}$ and $\text{UESSET}_{i,j} = \{M_{i,j,1}, \ldots M_{i,j,r_{i,j}}\}$
- $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$: S coordinates of diagonally opposite corners of RECREG_i, where $x_{\max} \geq x_{\min}$ and $y_{\max} \geq y_{\min}$
- ρ_{thr}: correlation coefficient minimum threshold

Output: IDSET_i: identifier map set such that $\text{IDSET}_i = \{\text{ID}_{i,1}, \ldots, \text{ID}_{i,q_i}\}$

Note: the registration feature used in computing correlation coefficients in this procedure is the same one that was used in computing the extent similarity values when MISET_i was generated (steps (3.2) and (3.3) in Figure 5.10).

(1) For $j = 1, \ldots, q_i$:

(1.1) Allocate $\text{ID}_{i,j}$ and RHO_j.

(1.2) For $x = x_{\min}, \ldots, x_{\max}$ and $y = y_{\min}, \ldots, y_{\max}$:

(1.2.1) Find id such that for $k = 1, \ldots, r_{i,j}$ $\rho_{M_{i,j,id},S,x,y} \geq \rho_{M_{i,j,k},S,x,y}$.

(1.2.2) If $\rho_{M_{i,j,id},S,x,y} \geq \rho_{\text{thr}}$ then
let $\text{ID}_{i,j}(x - x_{\min}, y - y\text{min}) = id$ and $\text{RHO}_j(x - x_{\min}, y - y\text{min}) = \rho_{M_{i,j,id},S,x,y}$;
otherwise,
let $\text{ID}_{i,j}(x - x_{\min}, y - y_{\min}) = \text{null}$ and $\text{RHO}_j(x - x_{\min}, y - y\text{min}) = \text{null}$.

(1.3) Perform non-maximum suppression on RHO_j and alter $\text{ID}_{i,j}$ accordingly.

Figure 5.13: Procedure to match image locations and model instances (version 1).

Input:

- S, MISET_i, $(x_{\text{min}}, y_{\text{min}})$, $(x_{\text{max}}, y_{\text{max}})$, ρ_{thr}
- $dfrac$: extent dimension fraction
- ρ_{sthr}: correlation coefficient minimum threshold for initiating full-resolution search

Output: IDSET_i: identifier map set such that $\text{IDSET}_i = \{\text{ID}_{i,1}, \ldots, \text{ID}_{i,q_i}\}$

(1) For $j = 1, \ldots, q_i$:

(1.1) Allocate $\text{ID}_{i,j}$ and RHO_j, and initialize $\text{ID}_{i,j}$ and RHO_j to null at all locations.

(1.2) Evaluate dim_x and dim_y respectively to the overall x-dimension and y-dimension of the $\text{UESSET}_{i,j}$ extent.

(1.3) Let $d_x = \max(dfrac \times dim_x, 1)$ and $d_y = \max(dfrac \times dim_y, 1)$.

(1.4) Let $x = x_{\text{min}}$.

(1.5) While $x \leq x_{\text{max}}$:

(1.5.1) Let $y = y_{\text{min}}$.

(1.5.2) While $y \leq y_{\text{max}}$:

(1.5.2.1) Find id such that for $k = 1, \ldots, r_{i,j}$ $\rho_{M_{i,j,id},S,x,y} \geq \rho_{M_{i,j,k},S,x,y}$.

(1.5.2.2) If $\rho_{M_{i,j,id},S,x,y} \geq \rho_{\text{sthr}}$ then perform full-resolution match search for $x = x - d_x/2, \ldots, x + d_x/2$ and $y = y - d_y/2, \ldots, y + d_y/2$ (see block (1.2) of Figure 5.13)

(1.5.2.3) Let $y = y + d_y$.

(1.5.3) Let $x = x + d_x$.

(1.6) Perform non-maximum suppression on RHO_j and alter $\text{ID}_{i,j}$ accordingly.

Figure 5.14: Procedure to match image locations and model instances (version 2).

A further refinement that was investigated was to search the uniform-extent subsets at two levels of resolution. This refinement is incorporated into the final version of the procedure, which is outlined in Figure 5.15. The uniform-extent subsets are grouped such that the members of each group represent extents that are approximately similar, within a specified tolerance. The groups are not necessarily disjoint sets, and one member of each group is the representative for that group (the definition of the groups will be described later). For each group, matching is first performed with the representative uniform-extent subset (the two-resolution search of the image-space, as described in Figure 5.14, is used). Tests for matches are performed with other members of the group only at those locations where a match was obtained using the representative. At each location where a match was obtained with an element of at least one member of the group only one match (one that is associated with the maximum corresponding correlation obtained over the group) is retained.

Let $\mathrm{E}_i = \{E_{i,1}, \ldots, E_{i,q_i}\}$ be the set of extents represented by $\{\mathrm{UESSET}_{i,1}, \cdots, \mathrm{UESSET}_{i,q_i}\}$. An approximately minimal subset E'_i of E_i is computed such that for any element e of E_i there is an element e' of E'_i such that $\eta_{e,e'} \geq \eta_{\mathrm{group}}$, where η_{group} is user specified (but generally greater than the value of η_{adj} that was involved in generating the model instance set (*cf.* Section 5.1.1)). The algorithm to find an approximately minimal dominating set of a graph (Figure 5.12) is applied to this task, in the same manner as it is to generate a model instance set (step (3.2) in Figure 5.10). The group representatives are those uniform-extent subsets whose extents are in E'_i, and the other members of a given group are those uniform-extents subsets with an extent similarity with the representative that is greater than or equal to η_{group}.

The registration feature that is used to compute correlation coefficients in the matching procedure and to compute the uniform-extent subset groups and representatives is assumed to be the same registration feature that in used in steps (3.2) and (3.3) of the procedure to model to generate a model instance set (Figure 5.10).

Input:

- S, MISET$_i$, $(x_{\min}, y_{\min})$, $(x_{\max}, y_{\max})$, ρ_{thr}, ρ_{sthr}, $dfrac$
- ng: number of UESSET groups
- rep: rep[m] is the representative of UESSET group m
- size: size[m] is the number of members of UESSET group m
- member: member[m, n] is member n of UESSET group m

Output: IDSET$_i$: identifier map set such that IDSET$_i$ = $\{\text{ID}_{i,1}, \ldots, \text{ID}_{i,q_i}\}$

(1) For $j = 1, \ldots, q_i$:

(1.1) Allocate ID$_{i,j}$ and RHO$_j$, and initialize ID$_{i,j}$ and RHO$_j$ to null at all locations.

(2) For $m = 1, \ldots, ng$:

(2.1) Let j = rep[m].

(2.2) Evaluate ID$_{i,j}$ using variable resolution search
(see steps (1.2) to (1.4) and block (1.5) in Figure 5.15).

(2.3) For $n = 1, \ldots,$ size[m]:

(2.3.1) Let k = member[m, n].

(2.3.2) If $k \neq j$ then check for match with UESSET$_{i,k}$ at those locations where ID$_{i,j}$ is non-null, and for each such location (x, y) let RHO$_{i,k}(x - x_{\max}, y - y_{\max}) = \rho_{M_{i,k,id},S,x,y}$.

(2.4) For $x = x_{\min}, \ldots, x_{\max}$ and $y = y_{\min}, \ldots, y_{\max}$:

(2.4.1) Search for id such that for $k \in \{\text{member}[m, 1], \ldots, \text{member}[m, \text{size}[m]]\}$
ID$_{i,id}(x - x_{\max}, y - y_{\max}) \neq$ null and
RHO$_{id}(x - x_{\max}, y - y_{\max}) \geq$ RHO$_k(x - x_{\max}, y - y_{\max})$.

(2.4.2) If id found then for $n = 1, \ldots,$ size[m]:

(2.4.2.1) Let k = member[m, n].

(2.4.2.2) If $k \neq id$ then let ID$_{i,k}(x - x_{\max}, y - y_{\max})$ = null.

(3) For $j = 1, \ldots, q_i$ perform non-maximum suppression on RHO$_j$ and alter ID$_{i,j}$ accordingly.

Figure 5.15: Procedure to match image locations and model instances (version 3).

Input:

- UESSET: uniform-extent subset
- ID: identifier map associated with UESSET
- RHO: ρ map associated with UESSET
- dim_x, dim_y: x and y dimensions, respectively, of the input maps

Output: ID (altered)

(1) Allocate dim_x-by-dim_y map PROP and initialize to null at all locations.

(2) For $x = 1, \ldots, dim_x - 1$ and $y = 1, \ldots, dim_y - 1$:

(2.1) If ID$(x, y) \neq$ null then:

(2.1.1) Register the extent of UESSET to PROP(x, y) using the centroid as the registration feature.

(2.1.2) For all x', y' such that (x', y') is a location in PROP that coincides with a location in the registered extent:

(2.1.2.1) If PROP$(x', y') =$ null or PROP$(x', y') <$ RHO(x, y) then let PROP$(x', y') =$ RHO(x, y).

(3) For $x = 1, \ldots, dim_x - 1$ and $y = 1, \ldots, dim_y - 1$:

(3.1) If ID$(x, y) \neq$ null and RHO$(x, y) <$ PROP(x, y) then let ID$(x, y) =$ null.

Figure 5.16: Procedure to perform non-maximum suppression.

5.4 Evaluation of Correct Recognition Probability

The probability that an initial recognition instance (or simply 'recognition instance' in the remainder of this section) is correct is evaluated on the basis of features that involve the scene image region that coincides with the extent locations or the direct irradiation locations (*cf.* Section 4.3) of the matching model instance.

Let there be d features. The feature values for any recognition instance are formatted into a d-dimensional feature vector $\mathbf{x}$. Correct and incorrect recognition represent two classes, to which we assign the indices 1 and 2, respectively. The correct recognition probability that is computed for $\mathbf{x}$, $P_{\mathbf{x}}(\text{class } 1)$ (or, equivalently, $1.0 - P_{\mathbf{x}}(\text{class } 2)$), is used to classify the associated recognition instance as correct (when $P_{\mathbf{x}}(\text{class } 1) \geq 0.5$) or incorrect (when $P_{\mathbf{x}}(\text{class } 1) < 0.5$). The incorrect recognition instances are then rejected, and the surviving instances (the intermediate recognition instances) are passed on to the conflict resolution filter, along with the associated correct recognition probability values.

Samples of correct and incorrect recognition instances with respectively e_1 and e_2 elements are manually selected from the initial recognition instances. An estimated probability of correct recognition is assigned to each element of each sample set. The feature vector for the j^{th} element of the i^{th} sample ($j \in \{1, \ldots, e_i\}$ and $i \in \{1, 2\}$) is $\mathbf{x}^{(i,j)}$. The estimated probability that $\mathbf{x}^{(i,j)}$ represents class k is $s_{i,j,k}$. Note that $s_{i,j,2} = 1.0 - s_{i,j,1}$.

The two sample sets may be considered to comprise a single *training set* of $e_1 + e_2$ *exemplars*. In this context the feature vectors are written with a single index. The estimated probability that $\mathbf{x}^{(i)}$ represents class j is $s_{i,j}$.

A nearest-neighbours interpolation procedure is used with the exemplars to estimate the probability that an input feature vector $\mathbf{x}$ represents correct recognition. This type of procedure was selected because it does not require prior knowledge of the manner in which members of the two classes are distributed in the feature space, and it readily supports the incremental expansion of the set of exemplars.

The nearest-neighbours interpolation procedure that we use is described in Section 5.4.1.

The selection and definition of features is discussed in Section 5.4.2.

5.4.1 Nearest-neighbours Interpolation

The performance of nearest-neighbours interpolation with a training set of finite size is sensitive to the weights assigned to the different features in the definition of the metric for the distance between an input vector and elements of the training set (the *similarity metric*). If features with relatively low class-discriminating power are assigned the same or higher weights as those with relatively high class-discriminating power, then a larger training set is required to obtain the same classification accuracy as is obtained when the low class-discriminating power features are assigned appropriate (relatively low) weights. A larger training set implies a greater data collection burden, greater memory requirements, and more computation in the classification.

The use of a fixed interpolation kernel in nearest-neighbours interpolation may result in the averaging of too few class probabilities to permit a good estimate of the local Bayes probability in parts of the feature space where the exemplars are sparsely distributed, and poor sensitivity to variations in the distribution in parts of the feature space where the exemplars are densely distributed. These problems are avoided through the use of a variable interpolation kernel that has a width that adapts to the density of exemplars in the feature space in the locality of the input vector. Such adaptation might be achieved by relating the kernel width to the k^{th} nearest-neighbour or to the average distance to the k nearest-neighbours.

Lowe's variable-kernel similarity metric (VSM) learning procedure (Lowe 1995) uses cross-validation optimization (based on the conjugate gradient optimization technique) to evaluate (1) the similarity metric weights and (2) a constant that relates the average distance to k nearest-neighbours to the width of a variable Gaussian interpolation kernel. The objective of cross-validation optimization is to minimize the error resulting when the class probabilities associated with each exemplar are predicted on the basis of the remaining exemplars. The VSM learning procedure and the associated nearest-neighbours interpolation procedure was used in

this work[2]

The nearest-neighbours interpolation procedure that is associated with VSM learning (Lowe 1995) has the following basis. Let N be the number of nearest neighbours that are included in the interpolation, and n_i be a weight that is associated with the i^{th} nearest-neighbour. A class probability for input vector $\mathbf{x}$ is computed according to

$$P_{\mathbf{x}}(\text{class } j) = \frac{\sum_{i=1}^{N} n_i s_{i,j}}{\sum_{i=1}^{N} n_i}.$$

The exemplar weight n_i is determined by a Gaussian kernel centred at $\mathbf{x}$ and is a function of the distance m_i of the neighbour from $\mathbf{x}$:

$$n_i = \exp(-m_i^2/2\sigma^2).$$

The similarity metric is a weighted Euclidean distance metric:

$$m_i^2 = \sum_{k=1}^{d} w_k^2(\mathbf{x}_k - \mathbf{x}_k^{(i)}).$$

Optimal values for the feature weights $w_1, \ldots, w_d$ are computed by the VSM learning procedure. The width of the Gaussian kernel is function of σ, which is evaluated to a multiple r of the average distance to the M nearest neighbours:

$$\sigma = \frac{r}{M} \sum_{i=1}^{M} m_i.$$

M is evaluated to some fraction of N. A value for r that is optimal with respect to the overall density of the exemplars within the feature space is computed by the VSM learning procedure.

In our 2-class application involving $e_1 + e_2$ exemplars, the error E that is minimized by the VSM learning procedure is computed according to:

$$E = \sum_{i=1}^{e_1+e_2} \sum_{j=1}^{2} (s_{i,j} - P_{\mathbf{x}^{(i)}}(\text{class } j))^2$$

[2]The implementation of the VSM learning procedure and of the nearest neighbours interpolation procedure (which includes an implementation of k-d tree construction and search procedures) that was used in this work was provided by Dr. David Lowe, Department of Computer Science, University of British Columbia.

where $P_{\mathbf{x}(i)}(\text{class } j)$ is the predicted probability that the i^{th} exemplar is in class j, based on that exemplar's nearest neighbours among all the other exemplars.

The N nearest neighbours of an input vector are found using an implementation of the k-d tree algorithm (Friedman, Bentley, and Finkel 1977). A k-d tree is essentially a multidimensional binary search tree. In this application, the number of search dimensions is equal to the number of features (d). Let T be the number of exemplars in the training set. Friedman *et. al.* (1977) show that construction of the k-d tree requires time proportional to $dT \log T$, and argue that the search for the N nearest neighbours of an input vector has expected time proportional to $\log T$. Sproull (1991) presents the results of experiments for different values of k and T, and for both uniformly distributed and clustered vectors. The search time was limited to being nearly proportional to $\log T$ and was nearly independent of N only for small d relative to T (*e.g.*, $d < 4$ and $T > 100$), and performance was better for clustered data than for uniformly distributed data. The k-d tree implementation provided reasonable performance in the main experiment (Section 6.7).

5.4.2 Feature Definitions

The set of features that is used in the nearest-neighbours interpolation is selected from a set of hypothetically useful features on the basis of feedback from the VSM learning procedure. This process is illustrated in Section 6.5.

The features that we hypothesize are useful in distinguishing correct and incorrect recognition instances fall into the following two classes:

- those that are based on the average value of the scene image pixels that coincide with the direct irradiation locations of the matching model instance (the *average direct irradiation pixel value*), for one or more scene image channels

- those that are based on the straight-line least squares regression of scene image pixel values on the values of the coincident matching model instance pixels (the *regression of the scene image on the model instance*), for one or more scene image channels

We denote the average direct irradiation pixel value for channel a by $\bar{D}_a$. The reflectance spectra of healthy green leaves (needle and broad) are characterized by local maxima in the visible green and near-infrared (nir) wavelength regions and local minima in the visible blue and visible red wavelength regions (Guyot, Guyon, and Riom 1989). This property allows the directly irradiated portion of healthy leaf-bearing trees to be manually distinguished from most other objects in colour-infrared photographs and colour-composite displays of multispectral digital images, on the basis of colour. Therefore, we hypothesize that $\bar{D}_{\text{blue}}$, $\bar{D}_{\text{green}}$, $\bar{D}_{\text{red}}$, $\bar{D}_{\text{nir}}$, and ratios $\bar{D}_{\text{maxima}}/\bar{D}_{\text{minima}}$, where 'maxima' and 'minima' denote wavelength regions centred respectively on green leaf reflectance maxima and minima, include useful features.

The regression of the scene image on the model instance incorporates the model instance and scene image pixel values as respectively the independent (X) and dependent (Y_a) variables in the regression, and the values of a coincident pair of pixels as one data point $(X_i, Y_{a,i})$ (the a component in the Y-subscript designates the scene image channel). This corresponds to the view that the model instance is an idealization of the image of a real crown up to, but not including, the application of a uniform gain and offset, and that the departures from the ideal due to such things as shape irregularities and gaps in the foliage comprise a form of noise (the possibility that these factors may provide useful information in the separate problem of species identification is acknowledged). Let n be the number of pixels in the model instance extent. A straight-line least squares regression produces values $b_{a,0}$ and $b_{a,1}$ such that

$$\hat{Y}_{a,i} = b_{a,0} + b_{a,1} X_i,$$

and such that the channel a residual sum of squares

$$SS_{\text{res},a} = \sum_{i=1}^{n} (Y_{a,i} - \hat{Y}_{a,i})^2$$

is minimized.

The channel a coefficient of determination R_a^2 is the ratio of the channel a regression sum of squares ($SS_{\text{reg},a}$) to the channel a total sum of squares corrected for the mean ($SS_{\text{tot},a}$):

$$R_a^2 = SS_{\text{reg},a}/SS_{\text{tot},a}$$

where

$$SS_{\text{tot},a} = \sum_{i=1}^{n} (Y_{a,i} - \bar{Y}_a),$$

$$SS_{\text{reg},a} = \sum_{i=1}^{n} (\hat{Y}_{a,i} - \bar{Y}_a),$$

and

$$\bar{Y}_a = \frac{1}{n} \sum_{i=1}^{n} Y_{a,i}.$$

R_a^2 is restricted to $[0, 1]$ and can be interpreted as the fraction of the total sum of squares corrected for the mean that is explained by the regression line; it is also the square of the channel a correlation coefficient for the recognition instance. Inspection of high spatial resolution vertical images of forests reveals that the amount of texture in a tree crown image can vary noticeably with species and crown size. R_a^2 should be sensitive to such variation, and we hypothesize that this would limit the usefulness of these values as features for distinguishing instances of the correct recognition of all sorts of tree crowns from instances of incorrect recognition.

The channel a pure error sum of squares ($SS_{\text{err},a}$) is the variation in the dependent variable that cannot be explained by the regression line (*i.e.*, the variation caused by multiple different values of Y_a at the same value of X). An upper bound on the value of R_a^2 ($R_{\text{max},a}^2$) can be defined using $SS_{\text{err},a}$ as follows:

$$R_{\text{max},a}^2 = \frac{(SS_{\text{tot},a} - SS_{\text{err},a})}{SS_{\text{tot},a}}.$$

If the increase in crown image texture is associated with an increase in $SS_{\text{err},a}$, then $R_a^2/R_{\text{max},a}^2$ (which is restricted to $[0, 1]$) should be less sensitive to species and crown size variation than R_a^2 is.

The pure error sum of squares can be estimated on the basis of *repeat runs* (*i.e.*, repeated measurements of Y_a for the same value of X). When there is an insufficient number of true repeat runs to support a reliable estimate of $SS_{\text{err},a}$, we can use a division of the X-value interval into a number of sub-intervals as the basis of an approximation of repeat runs (Draper and Smith 1981, pp. 35–42). Let m be the number of X-axis intervals, n_i be the number of

data points in the i^{th} interval, and $Y_{a,i,j}$ be the Y_a-value of the j^{th} data point in the i^{th} interval. Then,

$$\bar{Y}_{a,i} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{a,i,j}$$

and the estimated value for SS_{err} is

$$\widehat{SS}_{\text{err},a} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{a,i,j} - \bar{Y}_{a,i}).$$

When $m = 1$, the estimate is based on true repeat runs. When $n_i = 0$, $\bar{Y}_{a,i}$ is undefined and the i^{th} interval cannot be involved in the estimate.

Let X_{min} and X_{max} be, respectively, the minimum and maximum values of X_i, $i = 1, \ldots, n$. A measure of the degree of contrast within the channel a scene image region for a recognition instance is

$$C_a = (b_{a,0} + b_{a,1} X_{\text{max}}) - (b_{a,0} + b_{a,1} X_{\text{min}}) = b_{a,1} X_{\text{max}} - b_{a,1} X_{\text{min}}.$$

We hypothesize that this feature may be useful in distinguishing correct recognition instances from incorrect recognition instances that correspond to image regions that have a crown-like colour, but have shading that is too subtle to be crown-like (*e.g.*, part of a grassy field with local topographic shading variation that causes it to become a involved in an initial recognition instance). An alternate measure of contrast that does not involve the regression of the scene image on the model instance is

$$C'_a = \bar{D}_a - \bar{I}_a,$$

where $\bar{I}_a$ is the channel a average non-direct irradiation pixel value.

In summary, the hypothetically useful features are the following: $\bar{D}_{\text{a}}$, $\bar{D}_{\text{maxima}}/\bar{D}_{\text{minima}}$, R^2_a, $R^2_a/R^2_{\text{max},a}$, C_a, and C'_a.

5.5 Resolution of Interpretation Conflicts

We define the extent of a recognition instance to be the image extent of the positioned matching model instance. The initially implemented conflict resolution filter eliminated a recognition

instance if the centroid location of its extent was within that of a different recognition instance with a greater correct recognition probability value (*e.g.,* Figure 5.17a). This procedure was implemented as a version of the non-maximum suppression procedure outlined in Figure 5.13, with the modifications that it uses the correct recognition probability value associated with a recognition instance rather than the correlation coefficient value, and it operates over all of the indentifier maps. Within the context of the main experiment, this procedure produced reasonable results in some locations, but often failed. The following were the most commonly occurring failures:

- A conflict in which one or more recognition instances with relatively small extents and high correct recognition probabilities coincided with a single recognition instance with a relatively large extent and low correct recognition probability would not be resolved if the centroid of the large extent happened to fall outside of all the small extents (*e.g.,* Figure 5.17b).

- A recognition instance would be eliminated because its extent centroid was within the much smaller extent of a recognition instance with a slightly higher correct recognition probability. In such cases, the recognition instance with the larger extent was often the one that should have been retained. This failure often occurred because small-extent and large-extent model instances would respectively match the tip and the entirety of a conical tree crown (*e.g.,* Figure 5.17c).

These failures suggested the need for a broader notion of spatial interaction between recognition instances than whether the extent centroid for one recognition instance is inside or outside of the extent of a different recognition instance, and the need to link the quality of a recognition to more than just the correct recognition probability. This led to the idea of defining the conflict resolution filter to be an optimization procedure that is applied to the intermediate recognition instances on a window-by-window basis.

The windows are the rectangular cells of a uniform user-specified grid that is superimposed onto the image extent, and they are intended to be large enough to contain several of the largest

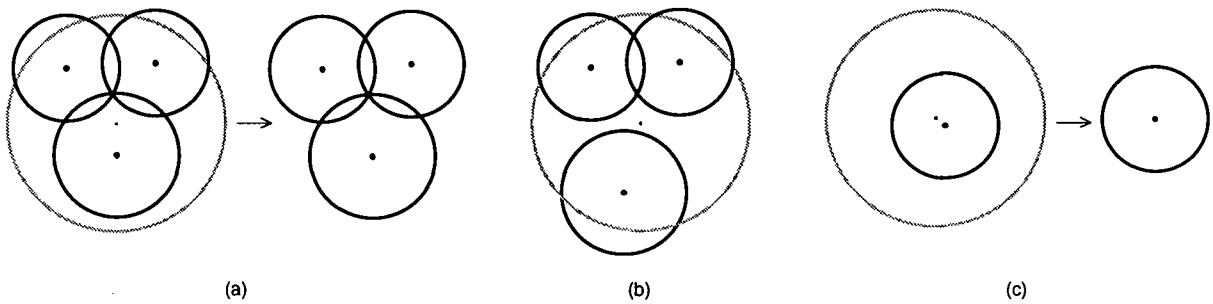


Figure 5.17: Schematic examples of conflict resolution based on non-maximum suppression. The dark and light boundaries represent the extents of recognition instances with, respectively, relatively high and low correct recognition probability values. Parts (a) and (c) illustrate the elimination of a recognition instance with a relatively large extent and low recognition probability value. Part (b) represents a small perturbation of the situation represented in Part (a) in which none of the recognition instances are eliminated.

tree crown images. Each window is also assigned a surrounding margin of a user-specified width that overlaps the adjacent windows (Figure 5.18). The optimization procedure is applied to each window and its surrounding margin (the *processing region*), but the results are retained only for the windows. Each processing region represents an instance of the optimization problem. The optimization is extended into the margins in order to reduce discontinuities in the results at window boundary locations.

The optimization procedure selects a subset of the intermediate recognition instances that are located within each processing region such that a certain objective function is maximized. Then, the results for each window are integrated into the final set of recognition instances by running them though the initially implemented conflict resolution filter that is described at the beginning of this section. This procedure attempts to remove conflicts that remain along the window boundaries. Although the procedure does not do this perfectly (examples of some failures are discussed and illustrated in Section 6.7.3), the results were found to be noticeably better when the procedure was applied than when it was not.

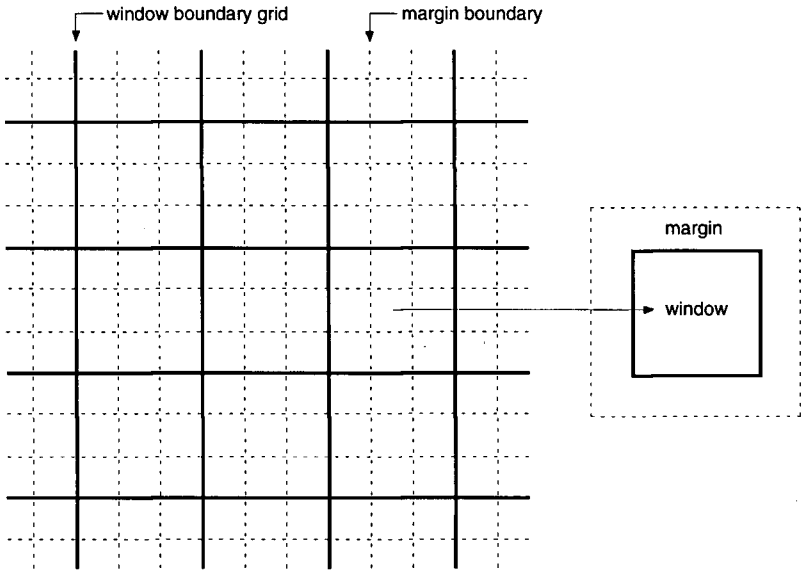


Figure 5.18: Processing windows and margins for interpretation conflict resolution.

The objective function F is a weighted combination of three components:

$$F = w_1 f_1 + w_2 f_2 + w_3 f_3$$

where the weights (w_1, w_2, and w_3) are user-specified and the components (f_1, f_2, and f_3) are related to the following aspects of the currently selected set of intermediate recognition instances:

f_1: the fraction of the processing region that does not coincide with the intersection locations of multiple positioned matching model instances

f_2: the average correct recognition probability value

f_3: the fraction of the processing region that coincides with locations within positioned matching model instances

The f_1 component is intended to control the degree of image-space overlap between positioned matching model instances and is related to the fact that within forest stands the crown

envelope intersection volume is generally less than the remaining crown envelope volume. The relationship is admittedly indirect since the intersections of the image-space projections of tree crowns in the scene do not exactly correspond to the intersections of the tree crown volumes in scene-space. This is particularly so in parts of the image where there is considerable height displacement and occlusion of some tree crown images by others.

The f_2 component is intended to reflect an overall preference for the recognition instances with higher correct recognition probability values, as evaluated by the correct recognition probability filter (*cf.* Section 5.4).

The f_3 component counterbalances the other two components: clearly, the rejection of all of the recognition instances would maximize f_1, and the retention of the one recognition instance with the highest correct recognition probability value would maximize f_2.

5.5.1 Definition of a Processing Region State and the Associated Objective Function Value

Let there be m locations, $(x_1, y_1), \ldots, (x_m, y_m)$, within the processing region at which the matching model instances for one or more recognition instances are positioned. We call these the *non-null* locations.

A *recognition instance selection* that may be made at non-null location (x_i, y_i) is

$$s_{x_i,y_i} \in \{\emptyset_{x_i,y_i}, r_{x_i,y_i,1}, \ldots, r_{x_i,y_i,n_i}\}$$

where $\{r_{x_i,y_i,1}, \ldots, r_{x_i,y_i,n_j}\}$ is the full set of n_j intermediate recognition instances positioned at (x_i, y_i), and $\emptyset_{x_i,y_i}$ denotes the selection of none of those recognition instances (the *null selection*).

A valid j^{th} state for a processing region during the optimization procedure is a set that represents one recognition instance selection at each of the non-null locations:

$$\mathrm{s}_j = \{s_{x_1,y_1,j}, \ldots, s_{x_m,y_m,j}\}$$

where $s_{x_i,y_i,j}$ is the selection at non-null location (x_i, y_i) that is associated with the state j.

The values of the objective function and its components for state j are denoted respectively F_j, and $f_{1,j}$, $f_{2,j}$, and $f_{3,j}$. The expressions for the components of the objective functions involve the following quantities:

N_j: the total number of non-null selections in s_j

$p(s_{x_i,y_i})$: the correct recognition probability associated with s_{x_i,y_i}
$(p(s_{x_i,y_i}) = 0$ if $s_{x_i,y_i} = \emptyset_{x_i,y_i})$

$p_{\max}$: $\max(p(r_{x_i,y_i,j}) : r_{x_i,y_i,j} \in \bigcup_{k=1}^{m}\{r_{x_k,y_k,1}, \ldots, r_{x_k,y_k,n_k}\})$

$p_{\min}$: $\min(p(r_{x_i,y_i,j}) : r_{x_i,y_i,j} \in \bigcup_{k=1}^{m}\{r_{x_k,y_k,1}, \ldots, r_{x_k,y_k,n_k}\})$

p_j^+: $\sum_{i=1}^{m} p(s_{x_i,y_i,j})$

c_j: the number of *singly or multiply covered* (or, simply, *covered*) locations in state j, which are the processing region locations that coincide with a location in at least one positioned matching model instance in state j

cm_j: the number of *multiply covered* locations in state j, which are the processing region locations that coincide with a location within the intersection of two or more positioned matching model instances in state j

c_{ub}: an upper bound on the number of covered locations in any state, defined by the number of covered locations when all of the recognition instances at all of the non-null locations are selected (this does not represent a valid state since in this case multiple selections may be made at a single non-null location)

The objective function components for state j are evaluated according to:

$$f_{1,j} = \frac{c_{\max} - cm_j}{c_{\text{ub}}}$$

$$f_{2,j} = \frac{(p_j^+/N_j) - p_{\min}}{p_{\max} - p_{\min}}$$

$$f_{3,j} = \frac{c_j}{c_{\text{ub}}}$$

The value for each component is restricted to $[0, 1]$ for any processing region. This helps ensure the applicability of the same set of weights to different processing regions.

The maximization of F for a particular processing region is an instance of a combinatorial optimization problem. The number of possible states grows exponentially with the number of non-null locations.

5.5.2 Initial Optimization Experiments

We initially implemented a backtracking procedure to perform the optimization, but found that within the the context of the main experiment, even small processing regions (*e.g.*, 25-by-25 pixels) are almost always associated with too many states to allow the procedure to terminate in a reasonable amount of time (the procedure performs an exhaustive search of the states). Processing regions that are small enough to allow a reasonable execution time (*e.g.*, 10-by-10 pixels) often cannot contain the spatial interactions between multiple large tree crown images, and lead to a significantly increased incidence of window-boundary discontinuities in the results.

In the backtracking procedure, each non-null location in a processing region corresponds to a different level in a state-space tree, and the different nodes at a level correspond to the different possible recognition instance selections at the corresponding location. A bounding function was added to the procedure in an attempt to improve the execution time. The bounding function is based on the assumptions that at all levels of the tree under the level of the current node there is a recognition instance selection with a recognition probability of 1.0, and that these selections together raise the coverage to c_{ub} without raising the multiple coverage. The bounding function returns the value of the objective function under these assumptions. If the bounding function value is less than that of the current largest objective function value then the descendants of the current node are not visited. The addition of this bounding function did not shorten the running time of the procedure to a reasonable level.

A greedy algorithm was also implemented. The procedure starts by selecting a recognition instance with the globally maximum correct recognition probability. It then searches for the remaining non-null location and selection selection that produces the maximum increase in the objective function value, and stops when an increase is no longer possible. Within the context of the main experiment the procedure consistently terminated with an obviously poor result. The procedure quickly became caught in locations of the state space with locally maximum objective function values.

5.5.3 Optimization by Similated Annealing

The optimization procedure that was used to produce the results reported in Section 6.7 is based on *simulated annealing*. The following definitions are fundamental to the discussion of our application of this technique:

S: the set of all states in a problem instance (a problem instance corresponds to a processing region)

ε: the state *score function*, which assigns a real number to a state; if $\varepsilon(s_i) < \varepsilon(s_j)$ then s_i is considered to be closer to optimum than s_j

μ: the *move set*; a relation over S ($\mu \subseteq S \times S$) that represents the pairs of elements of S that are connected by a permissable atomic state modification, or *move.*

μ^k: the pairs of elements of S that are connected by a sequence of k moves ($\mu = \mu^1$).

(S, μ): the *state-space*

α: the *acceptance function*, which assigns a real number in $(0, 1]$ to a pair of scores and a *control parameter* value; $\alpha(\varepsilon(s_i), \varepsilon(s_j), t)$ is the probability of accepting a move from state s_i to state s_j under control parameter value t

β: the *selection function*, which assigns a real number in $[0, 1]$ to a pair of states; $\beta(s_i, s_j)$ is the probability of randomly selecting a move that, if accepted, would cause a transition

from state s_i to state s_j

This notation and terminology is taken from Otten and van Ginneken (1989), which is also the main background reference on simulated annealing that was used in this work.

In discussions of simulated annealing it is customary to consider the problem to be one of minimization, in other words to find a state $s_i \in S$ such that

$$\forall s_j \in S \, [\varepsilon(s_i) \leq \varepsilon(s_j)].$$

Therefore, in our application we equate $\varepsilon(s_i)$ and $-F_i$.

A basic simulated annealing algorithm is outlined in Figure 5.19. The overall form of the algorithm is an inner loop (block (3.2)) nested within an outer loop (block (3)), with the number of iterations of the inner loop determined in the outer loop. Also, the inner loop contains an implementation of an acceptance function (line (3.2.2)) that involves a pseudo-randomly generated real number from the interval $[0, 1)$ (denoted *random*). The control parameter t for the acceptance function is evaluated in the outer loop.

The probability that s_j will be the next accepted state given that s_i is the current state (*i.e.*, the probability of the transition from s_i to s_j) is $\tau(s_i, s_j, t)$, where

$$\tau(s_i, s_j, t) = \alpha(\varepsilon(s_i), \varepsilon(s_j), t) \, \beta(s_i, s_j).$$

It is convenient to think of all the transition probabilities for a given value of t as elements of a |S|-by-|S| *transition matrix* $\mathbf{T}(t)$, where $\mathbf{T}(t)_{i,j} = \tau(s_i, s_j, t)$. Clearly, the entries of $\mathbf{T}(t)$ are non-negative and the row sums are equal to unity; therefore, $\mathbf{T}(t)$ is a *Markov matrix*, and the sequence of current states generated by the inner loop is a *Markov chain*. This sequence has the property that the next accepted state does not depend on the states that have preceded the current state. For a given value of t, the transition matrix does not change and that part of the sequence of current states is called a homogeneous Markov chain. The conditional probability that $s_{curr} = s_j$ given that n steps earlier in the same homogeneous Markov chain $s_{curr} = s_i$ is denoted by $\tau_n(s_i, s_j, t)$.

Input: (S, μ): the state-space

Output: s_{opt}: the estimated optimum state

(1) Initialize K and t.

(2) Let *stop* = FALSE.

(3) While *stop* $\neq$ TRUE:

- **(3.1)** Let $i = 1$.
- **(3.2)** While $i \leq K$:
  - **(3.2.1)** Apply a random move to s_{curr} to get s_{new} and let $\varepsilon_{new} = \varepsilon(s_{new})$.
  - **(3.2.2)** If *random* $< \exp((\varepsilon_{curr} - \varepsilon_{new})/t)$ then:
    - **(3.2.2.1)** Let $s_{curr} = s_{new}$ and $\varepsilon_{curr} = \varepsilon_{new}$.
  - **(3.2.3)** Let $i = i + 1$.
- **(3.3)** Update K, and t.
- **(3.4)** If the stop criteria are met then let *stop* = TRUE.

(4) Let $s_{opt} = s_{curr}$.

Figure 5.19: Simulated annealing algorithm.

Let the move set have the properties of *space connectivity* (*i.e.*, any state can be reached from any other state by a sequence of moves) and *reflexivity*:

$$\bigcup_{k=1}^{\infty} \mu^k = \mathrm{S} \times \mathrm{S}$$

and

$$\forall \mathrm{s}_i \in \mathrm{S}\, [(\mathrm{s}_i, \mathrm{s}_i) \in \mu]$$

Also, let β have the following properties:

$$\forall (\mathrm{s}_i, \mathrm{s}_j) \in \mu\, [\beta\, (\mathrm{s}_i, \mathrm{s}_j) \neq 0]$$

and

$$\forall (\mathrm{s}_i, \mathrm{s}_j)) \notin \mu\, [\beta\, (\mathrm{s}_i, \mathrm{s}_j) = 0].$$

Then for the homogenous Markov chain there exists a probability density function of states δ, called the *equilibrium density* of the chain such that

$$\delta\, (\mathrm{s}_j, t) = \lim_{k \to \infty} \tau_n\, (\mathrm{s}_i, \mathrm{s}_j, t)$$

and where the δ does not depend on s_i. This is the *chain limit theorem*, which is proved in Chapter 3 of Otten and van Ginneken (1989).

Let us add *symmetry* to the conditions on μ and β:

$$\forall \mathrm{s}_i \in \mathrm{S},\ \forall \mathrm{s}_j \in \mathrm{S}\, [(\mathrm{s}_i, \mathrm{s}_j) \in \mu \rightarrow (\mathrm{s}_j, \mathrm{s}_i) \in \mu]$$

and

$$\forall \mathrm{s}_i \in \mathrm{S},\ \forall \mathrm{s}_j \in \mathrm{S}\, [\beta\, (\mathrm{s}_i, \mathrm{s}_j) = \beta\, (\mathrm{s}_j, \mathrm{s}_i)],$$

and let us adopt the acceptance function encoded in Figure 5.19:

$$\alpha\, (\varepsilon\, (\mathrm{s}_i), \varepsilon\, (\mathrm{s}_j), t) = \min \{1, \exp\, ((\varepsilon\, (\mathrm{s}_i) - \varepsilon\, (\mathrm{s}_j))/t)\}.$$

Then, for a sufficiently high value of t, the equilibrium probability is practically the same for all states, and for a sufficiently low value of t the equilibrium probability is practically zero for all

states that do not have the globally minimum score (Otten and van Ginneken 1989, Chapter 3). We call a homogeneous Markov chain that satisfies above-specified requirements on μ, α, and β an *annealing chain*.

For annealing chains, fewer steps are required to get close to equilibrium (*i.e.*, reach *quasi-equilibrium*) with a relatively large value of t than with a relatively low value of t. Also, after quasi-equilibrium is obtained for one value of t, relatively few steps are required to re-establish quasi-equilibrium in a new chain for a slightly different value of t whose starting state is the old chain's ending state. The central idea of the annealing algorithm, then, is to quickly establish quasi-equilibrium for a high value of t and then re-establish quasi-equilibrium for decreasing values of t until the stopping criteria are met, which should occur when there is little or no decrease in the average score for successive chains. The procedures described in Otten and van Ginneken (1989) for initializing t and K (the number of steps to run the current annealing chain through), computing successive values of t and K, and determining when to stop were implemented in our application of simulated annealing. These procedures involve statistics compiled during the inner loop executions that are required to reach quasi-equilibrium for each value of t, and their implementation has been omitted from Figure 5.19.

For our application, we define a move as the selection of a single non-null location, followed by a recognition instance replacement at that location. A recognition instance replacement at a particular location is the removal of the previously selected recognition instance followed by the addition of a newly selected recognition instance at that location. In a random move, the location and new recognition instance selections are random. Clearly, the move set has the properties of space connectivity, reflexivity, and symmetry that are required for simulated annealing (space connectivity is satisfied because any final state can be reached from an initial state by transitioning to the state of all null selections, and then making the non-null selections required to establish the final state). The required conditions on β are trivially satisfied.

The evaluation of F_{j+1} to accompany the transition from S_j to S_{j+1} requires the evaluation of N_{j+1}, c_{j+1}, cm_{j+1}, and p^+_{j+1}. This procedure is outlined in Figure 5.20. The procedure is

supported by a *coverage count map* for the processing region, which is denoted by COVCNT, and the functions $\triangle c_r$, $\triangle cm_r$, $\triangle c_a$, $\triangle cm_a$.

Each location in a coverage count map is evaluated to the number of different positioned matching model instances that contain a coinciding location in the current state. The alteration of a coverage count map during a recognition instance replacement is illustrated in Figure 5.21.

The function $\triangle c_r$ returns the reduction in the number of covered locations in the processing region (a number that is less than or equal to zero) when the specified recognition instance is removed from the specified location. The function $\triangle c_a$ returns the increase in the number of covered locations in the processing region (a number that is greater than or equal to zero) when the specified recognition instance is added to the specified location. The functions $\triangle cm_r$ and $\triangle cm_a$ are similar to $\triangle c_r$ and $\triangle c_a$, except that they respectively return the reduction or increase in the number of multiply covered locations. Each of these functions also updates the coverage count map. Note that an implementation of any of these functions needs to scan and modify the coverage count map only within the extents of the removed or added recognition instances.

5.5.4 Summary

Interpretation conflict resolution was formulated as a window-based combinatorial optimization problem that involves the spatial interaction between recognition instances as well as the correct recognition probability values.

A backtracking optimization procedure that performs an exhaustive search of the states for a processing region (a window plus its surrounding margin) typically required an unreasonable amount of computation for windows that were large enough to contain the spatial interaction between multiple images or large tree crowns. The addition of a bounding function to the procedure did not decrease the required amount of computation to a reasonable level. A greedy optimization procedure usually became caught at local objective function maxima within the state-space. An optimization procedure that is based on simulated annealing had reasonable

Input: COVCNT, (x_i, y_i) $s_{x_i,y_i,j}$, $s_{x_i,y_i,j+1}$, N_j, p_j^+, c_j, cm_j

Output: N_{j+1}, p_{j+1}^+, c_{j+1}, cm_{j+1}

(1) If $s_{x_i,y_i,j} \neq \emptyset_{x_i,y_i}$ then let

$N_{\text{tmp}} = N_j - 1$,
$p_{\text{tmp}}^+ = p_j^+ - p\,(s_{x_i,y_i,j})$,
$c_{\text{tmp}} = c_j + \triangle c_r\,(\text{COVCNT}, (x_i, y_i), s_{x_i,y_i,j})$, and
$cm_{\text{tmp}} = cm_j + \triangle cm_r\,(\text{COVCNT}, (x_i, y_i), s_{x_i,y_i,j})$;

otherwise, let

$N_{\text{tmp}} = N_j$,
$p_{\text{tmp}}^+ = p_j^+$,
$c_{\text{tmp}} = c_j$, and
$cm_{\text{tmp}} = cm_j$.

(2) If $s_{x_i,y_i,j+1} \neq \emptyset_{x_i,y_i}$ then let

$N_{j+1} = N_{\text{tmp}} + 1$,
$p_{j+1}^+ = p_{\text{tmp}}^+ + p\,(s_{x_i,y_i,j+1})$,
$c_{j+1} = c_{\text{tmp}} + \triangle c_a\,(\text{COVCNT}, (x_i, y_i), s_{x_i,y_i,j+1})$, and
$cm_{j+1} = cm_{\text{tmp}} + \triangle cm_a\,(\text{COVCNT}, (x_i, y_i), s_{x_i,y_i,j+1})$;

otherwise, let

$N_{j+1} = N_{\text{tmp}}$,
$p_{j+1}^+ = p_{\text{tmp}}^+$,
$c_{j+1} = c_{\text{tmp}}$, and
$cm_{j+1} = cm_{\text{tmp}}$.

Figure 5.20: Procedure to update the objective function data to reflect the replacement of $s_{x_i,y_i,j}$ by $s_{x_i,y_i,j+1}$ at location (x_i, y_i).

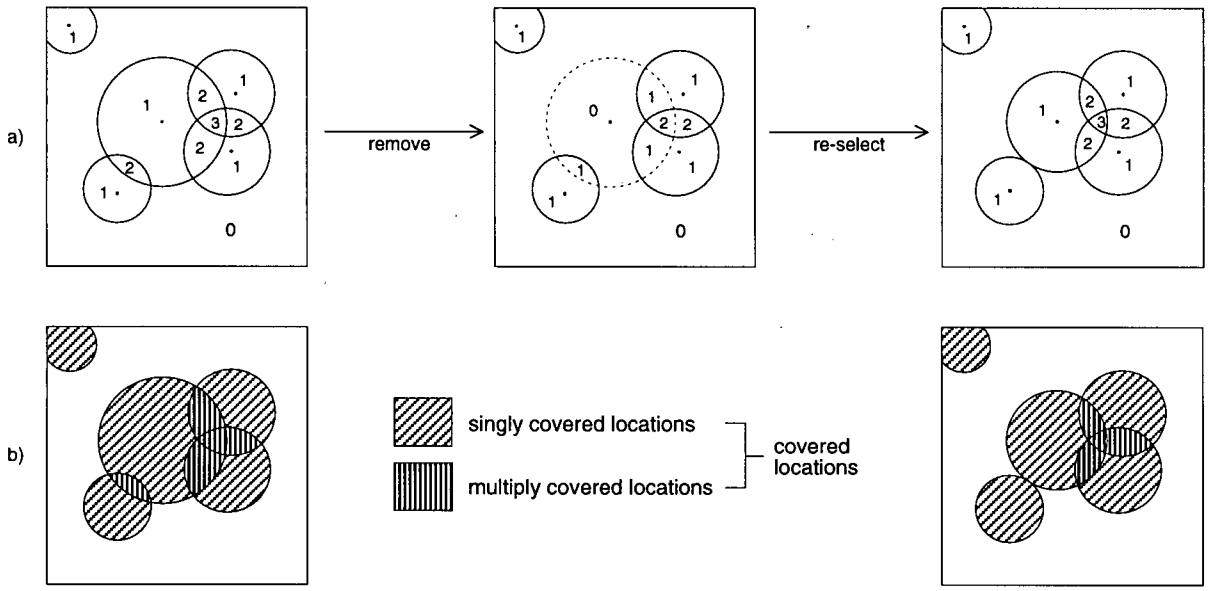


Figure 5.21: Alteration of a coverage count map during a recognition instance replacement: (a) states of the coverage count map, and (b) initial and final coverage classes.

execution times for adequately large processing regions, and was used to produce the results that are reported in Section 6.7.

Chapter 6

Experiment

This chapter is organized as follows:

Section 6.1 describes the experimental image data and the reference data.

Section 6.2 describes the manual tree crown recognition experiment, the results of which were compared to the automatic tree crown recognition results.

Section 6.3 describes the evaluation of the inputs to the model instance set generation procedure.

Section 6.4 describes the evaluation of the adjustable parameters of the initial matching procedure: ρ_{thr}, ρ_{sthr}, $dfrac$, and η_{group}.

Section 6.5 describes the selection of the training set and the features that were used in estimating the probability of correct recognition.

Section 6.6 describes the evaluation of the adjustable parameters to the conflict resolution procedure.

Section 6.7 presents an assessment and analysis of the automatic tree crown recognition and diameter estimation results.

6.1 Data

The experimental image data were acquired from an aircraft with the MEIS II in two flightlines over the PNFI research forest near Chalk River, Ontario, Canada (46°00′N; 77°25′W) on the morning of 16 August 1988. The scene terrain has low relief. Basic flight parameters from the flight report are listed in Table 6.1.

FLIGHTLINE	ALTITUDE (metres agl)	AZIMUTH (degrees)	GR. SPEED (metres/sec)	TIME (GMT)	
				Start	End
Sturgeon	518	321	82.8	13:58:33	14:01:59
Headquarters	518	277	79.7	14:11:52	14:13:46

Table 6.1: Flight parameters for the MEIS II data acquired over the PNFI research forest on 16 August 1988.

The geographic coordinates of the flightline extrema were estimated by locating them on 1:25,000 scale National Topographic Series map sheets. The sun positions corresponding to the flightline location and time extrema were calculated using the method described in Chapter 1 of Iqbal (1983). The solar declination value for 16 August 1988 of +13°44′20.8″ listed in Section C of *The Astronomical Almanac for the Year 1988* was used in the calculation. These results are listed in Table 6.2.

LOCATION	LATITUDE	LONGITUDE	SUN AZIMUTH (degrees)	SUN ZENITH ∠ (degrees)
Sturgeon Start	45°56′52″N	77°20′46″W	111.93	52.03
Sturgeon End	46°00′55″N	77°25′20″W	112.71	51.55
Headquarters Start	45°59′34″N	77°22′45″W	115.08	49.94
Headquarters End	45°59′59″N	77°27′35″W	115.90	49.42

Table 6.2: Estimated coordinates of flightline extrema and the calculated sun position at those locations at the time of imaging.

Meteorological data collected at PNFI near the time of image acquisition and obtained from the Canadian Atmospheric Environment Service indicate a ground level visibility of 24.1 km, no cloud cover, and a ground level windspeed between 13 and 11 km/hr. Iqbal (1983, p. 119) lists horizontal ground level visibility values of 340, 28, and 11 km and indicative of "clean",

"clear", and "turbid" atmospheres, respectively.

6.1.1 Image Data

Table 6.3 lists the wavelength bands in which the experimental image data channels were sensed. Pixel values for each channel are represented by 8 bits.

MEIS II CHANNEL	CENTRE WAVELENGTH (μm)	WAVELENGTH RANGE (μm)	SIGNIFICANCE
7	0.4489	0.4336–0.4642	chloropyhll abs. peak
8	0.5484	0.5325–0.5643	chlorophyll refl. peak
4	0.6402	0.6217–0.6587	chlorophyll abs. peak
2	0.6751	0.6553–0.6949	leaf refl. slope peak
3	0.8734	0.8461–0.9007	leaf refl. peak
6	1.0175	0.9628–1.0722	leaf refl. peak

Table 6.3: Channel descriptions for the MEIS II data acquired over the PNFI research forest on 16 August 1988.

Channel 6 was not made available to this experiment because its quality was considered to be too low for it to be usable. Visual inspection of the displayed data shows that channels 3 and 8, which respectively correspond to near-infrared and visible green wavelength bands, have noticeably better contrast and less noise than the other channels. Channel 8 exhibits greater contrast within individual tree crown image regions than channel 3, which is to be expected since individual leaves generally have higher transmittance in the near-infrared wavelength region, as pointed out in Section 4.2.2.

Pixel value frequency histograms were compiled for each of the image data channels (Figure 6.22). The histogram for each channel represents the same 24 320-by-320 pixel image regions. Each image region is centred on a different ground sample plot (*cf.* Section 6.1.2). These sample plots are distributed throughout both flightlines (*cf.* Section 6.7).

The frequency histograms show that the sampled pixel values in all channels are concentrated in the lower half of the $[0, 255]$ range. The peaks at pixel value 255 are evidence that there were bright objects in the scene that saturated the sensor. In all of the bands tree crowns are generally not the brightest objects (some areas of exposed soil and pavement were brighter),

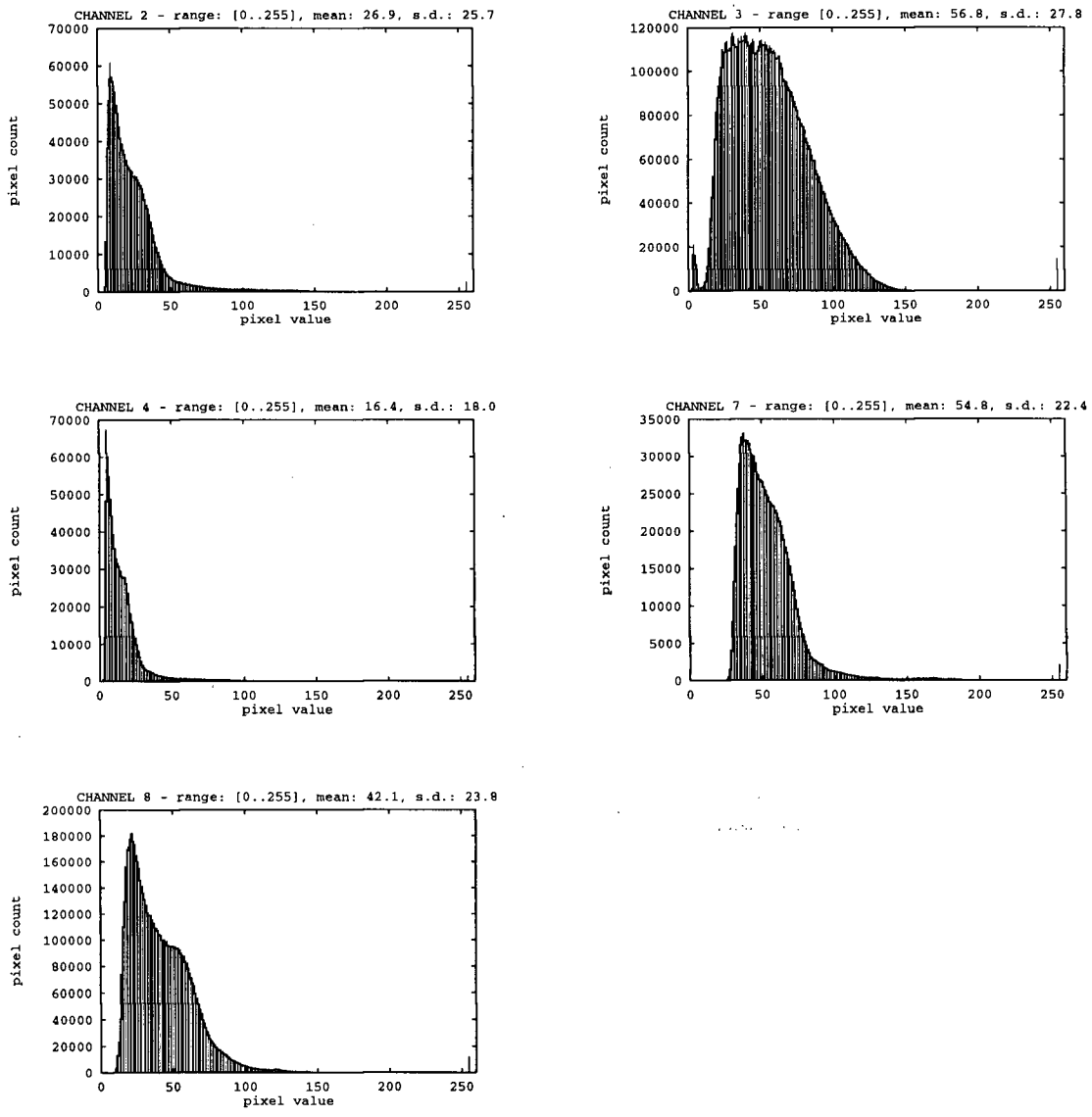


Figure 6.22: Pixel value frequency histograms for the MEIS II image channels. Each histogram is computed for the same 24 320-by-320 pixel image regions. Each image region is centered on a different ground sample plot.

so the tree crowns are effectively imaged with 7 bits of precision.

The nominal along-scanline pixel ground dimension is 0.364 m for both flightlines. This estimate is based on the nominal flight altitude above ground level (Table 6.1) and the MEIS II focal length and output pixel image-plane dimension (McColl, Neville, and Till 1983). The nominal along-flightline pixel ground dimension is 0.414 m and 0.398 m for the Sturgeon and Headquarters flightlines, respectively. These estimates are based on the data used to estimate the along-scanline dimension, the nominal flightline ground speed (Table 6.1), and a detector array scan rate of 200 Hz for both flightlines.

The image data were acquired without real-time aircraft roll correction (a MEIS II system option), and no geometric post-processing was applied to them. The orientation angles of an inertial navigation system (INS) that was mounted in the aircraft were sampled during image acquisition. The INS was not attached to, or precisely aligned with, the sensor head, so the measured orientation angles were not necessarily those of the sensor coordinate frame. However, the INS was only about 2 m away from the sensor head, so relative orientation variation due to airframe flex was negligible. Table 6.4 lists the INS orientation angle standard deviations for portions of the full-flightline images that encompass the sample plots (the sample plots are discussed in Section 6.1.2). Reliable data on groundspeed and height variation are not available.

FLIGHTLINE	# SAMPLES	ROLL σ	PITCH σ	AZIMUTH σ
Sturgeon	8883	0.407	0.359	0.200
Headquarters	6285	0.416	0.393	0.399

Table 6.4: Standard deviations (in degrees) of INS orientation angles for flightline intervals that encompass the sample plots.

Under the assumption that the $x_s y_s$-plane was horizontal at the mean roll and pitch angle for the INS, then the maximum angle between the optical and vertical axes for the Sturgeon and Headquarters flighlines was respectively 1.717° and 0.925°.

The image data were processed at PNFI in order to remove dropped scanlines (image rows correspond to scanlines), and remove some column striping. Also, an additive along-row pixel value adjustment was applied in order to reduce a gradation in average column image intensity.

This gradation was caused by along-scanline local view direction variation combined with flight directions that were not oriented exactly along the solar azimuth. Image regions corresponding to red pine stands were delineated, and for each region the mean pixel value in each band and the image location of the centroid was computed. The adjustment for each band was based on a quadratic least squares regression of region mean pixel value on region centroid scanline position. The adjustment was less than 10 pixel value increments for all bands. Although not required for this research, the adjustment was considered to be beneficial to manual image interpretation involved in other research that used these data. The image data pre-processing procedures are described in further detail in Cloney *et. al.* (1994).

6.1.2 Reference Data

During the periods of May to October 1989 and May to August 1990 various attributes of a sample of 587 overstory trees that were distributed among twenty-four 20-by-20 metre square sample plots in the scene were evaluated in the field by PNFI personnel. These attributes are (1) species, (2) stem diameter at breast height (dbh), (3) horizontal crown diameter, (4) tree height, (5) dominance, (6) crowding, and (7) species context.

The dbh was measured to the nearest 0.1 cm. The horizontal crown diameter was evaluated as the average of the maximum north-south and east-west horizontal crown dimensions and recorded to the nearest 0.1 m. Tree height was evaluated to the nearest 0.1 m. Dominance was evaluated as dominant (crown clearly higher than those of neighbouring trees) or co-dominant (crown in overstory but not clearly higher than those of neighbouring trees). Crowding was evaluated as low (tree fairly open growing and not competing with neighbours), medium (tree in light competition with neighbours), or high (tree in contact with neighbours on all sides). Species context (a measure of the species mixture of the overstory canopy in the neighbourhood of the sample tree) was evaluated as either purely softwood, a mixture of softwood and hardwood with softwood predominant (softwood/hardwood), a mixture of softwood and hardwood with hardwood predominant (hardwood/softwood), or purely hardwood.

PNFI personnel manually estimated the crown apex image location for 548 of the sample trees (these locations were confirmed through field visits with hardcopies of the image data) the image extents of the crowns of 539 of the sample trees, and the image location of each corner of each sample plot. Two of the apex locations correspond to multiple sample trees (4 each). All the other apex locations have a one-to-one correspondence with sample trees.

In a display of the image data, understory trees can be manually detected through gaps in the upper canopy stratum within each sample plot. PNFI personnel estimated the crown apex image location for these *visible understory trees*, and for trees having a stem location outside of any plot and a crown extent intersecting a plot boundary (*border trees*). This produced a total of 203 additional crown apex image locations. Attributes were not evaluated for any of the visible understory trees or border trees.

The reference data collection procedure is described in further detail in Cloney *et. al.* (1994).

A *reference display* was created for each of 24 sample plots by overlaying in white the estimated apex locations of the sample trees, visible understory trees, and border trees, and the estimated sample tree crown boundaries onto a 100-by-100 pixel channel 8 subimage centered on the sample plot. The reference displays were used in assessing the accuracy of the results produced by the recognition procedure (the accuracy assessment procedure is described in detail in Section 6.7).

6.2 Manual Recognition Experiment

An assessment of the accuracy of manual recognition for 15 of the sample plots was performed at PNFI. This includes, for each sample plot, the average number of omission errors and the average number of commission errors over 5 different interpreters who worked independently of one another. These data for 14 sample plots are tabulated in Section 6.7.1, where they are compared to automatic recognition results.

The source of the manual recognition results and the following outline of the experimental procedure is Cloney and Leckie (1995).

Four of the interpreters had extensive experience in forest inventory photo-interpretation and the remaining interpreter was a forestry student with little photo-interpretation experience. Each interpreter used the same workstation display of the same three-channel colour composite of sample plot subimages drawn from the experimental image data (linearly scaled versions of channels 3, 7, and 4 assigned to the red, green, and blue display coordinates, respectively). The colour composite was designed to also support an experiment in manual species discrimination. The workstation supported continuous zooming and roaming of the display. All of the interpreters were given the same orientation resentation in which they were introduced to the display equipment and shown some examples of correct crown recognition.

The interpreters were required to record the pixel location of the estimated centroid of each identified tree crown's image extent.

The sample plot subimages were partitioned into three subsets and a different pixel size was used in displaying each subset. The subimages for six of the plots were displayed with the original pixel dimension. The subimages for two of the plots were displayed with pixels having half the side-dimension of the original, and the subimages for the remaining seven plots were displayed with pixels having twice the side-dimension of the original (in all cases cubic-convolution resampling was used). The pixel size variation was considered to have a minor effect on the manual recognition results.

Border trees and visible understory trees (*cf.* Section 6.1.2), as well as visible ground vegetation were masked-out in the sample plot subimage, and the interpreters were told to ignore the masked-out areas.

6.3 Evaluation of Inputs to Model Instance Set Generation

This section describes the application of the procedure outlined in Section 5.1.1 to the experimental scene and image data. The results are summarized in Section 6.3.6.

6.3.1 Definition of Constraints on the Locally Variable Parameters

The scene is part of the Middle Ottawa Forest Section of the Great Lakes-St. Lawrence Forest Region (Rowe 1972, pp. 100). In this Forest Section, upland forests are usually composed of sugar maple (*Acer saccharum* Marsh.), American beech (*Fagus grandifolia* Ehrh.), yellow birch (*Betula alleghaniensis* Britton), red maple (*Acer Rubrum* L.), eastern hemlock [*Tsuga canadensis* (L.) Carr.], eastern white pine (*Pinus strobus* L.), and red pine (*Pinus resinosa* Ait.). Forests on dry ridges and sand flats are usually composed of eastern hemlock, eastern white pine, and jack pine (*Pinus banksiana* Lamb.). Varying amounts of white spruce [*Picea glauca* (Moench) Voss], balsam fir [*Abies balsamea* (L.) Mill.], trembling aspen (*Polpulus tremuloides* Michx.), white birch (*Betula papyrifera* Marsh.), red oak (*Quercus rubra* L.), and basswood (*Tilia americana* L.) are present throughout both of these general forest types. Swamp forests are common, and they are usually composed of eastern white cedar (*Thuja occidentalis* L.), eastern larch [*Larix laricina* (Du Roi) K.Koch], black spruce [*Picea mariana* (Mill.) B.S.P.], black ash (*Fraxinus nigra* Marsh.), red maple and white elm (*Ulmus americana* L.).

Descriptions of crown form in Hosie (1969) and Sayn-Wittgenstein (1978) for these tree species were compared to plots of a generalized ellipsoids that were computed from different parameter values in order to obtain the following constraints on tree crown size and shape shape for the overall scene.

Table 6.5 lists the identified constraints on $(\frac{a}{b}, n)$, which defines the crown envelope shape. Table 6.6 lists the identified constraints on $(b, \frac{a}{b}, n)$, which defines the crown envelope size and shape.

id	$\frac{a}{b}$	n
1	8.0 – 10.0	1.0 – 1.5
2	4.0 – 6.0	1.0 – 1.5
3	2.0 – 4.0	1.5 – 2.0
4	1.0 – 2.0	2.0 – 2.5

Table 6.5: Crown envelope shape intervals.

b (metres)	shape id
1.0 – 2.0	1, 2, 3
2.0 – 4.0	2, 3, 4
4.0 – 6.5	4

Table 6.6: Crown envelope size and shape intervals.

An overall horizontal radius range of 1.0 m to 6.5 m was obtained by manually inspecting the image data. The shape with id = 1 (*cf.* Table 6.5) is restricted to black spruce which has an exceptionally narrow ($b < 2\,\mathrm{m}$) elongated crown. The largest crowns ($b > 4\,\mathrm{m}$) belong to species for which the sun crown is characteristically parabolic or slightly flat-topped (eastern white pine, eastern hemlock, red pine, yellow birch, and sugar maple).

The following are the constraints that were identified for the remaining locally variable parameters:

- X_{end}: $0.7 \times \min(b, 3.0\,\mathrm{m}) \leq X_{\mathrm{end}} \leq 1.0 \times \min(b, 3.0\,\mathrm{m})$
- X_{max}: $0.2 \times X_{\mathrm{end}} \leq X_{\mathrm{max}} \leq 0.5 \times X_{\mathrm{end}}$
- D_{avg}: $0.5\,\mathrm{m}^{-1} \leq D_{\mathrm{avg}} \leq 3.0\,\mathrm{m}^{-1}$
- D_{max}: $(1.5 \times D_{\mathrm{avg}}) \leq D_{\mathrm{max}} \leq (4.5 \times D_{\mathrm{avg}})$
- k: $0.75 \leq k \leq 1.00$
- B: $0.20 \leq B \leq 0.50$

(full parameter descriptions are given in Section 4.3).

The choice of the constraints on the permeable-volume model parameters related to tree crown composition was guided by the following information:

- Average projected one-sided leaf area density values of $1.5\,\mathrm{m}^2/\mathrm{m}^3$ (for a Norway spruce) and $1.86\,\mathrm{m}^2/\mathrm{m}^3$ (for an unspecified coniferous species) are quoted in Oker-Blom and Kellomäki (1982) and Kuuluvainen and Pukkala (1989), respectively.

- In Koppel and Oja (1984) the average one-sided (not projected) leaf area density for an entire Norway spruce crown is given as $5.6\,m^2/m^3$, the maximum depth of the live foliage is given as 2.5 m, the density value for the portion of the crown above the height where the non-foliated inner region disappeared is given as $17\,m^2/m^3$, or 3.0 times the whole crown average, and the average for the sun crown is given as $8\,m^2/m^3$, or 1.4 times the whole crown average.

- In Wang *et al.* (1990) it is reported that the maximum leaf area density for a sample of radiata pine trees was up to 4 times greater than the average for the entire crown, and that the peak of the average (over the vertical crown extent) horizontal radial leaf area density distribution tended to occur between one-third and one-half of the crown radius in from the crown surface.

The constraint for B was based on a sample of ratios of canopy gap image region pixel values to tree crown image region average pixel values (in the MEIS II, the gain for each channel is selected by the operator, but the only offset is a real time subtraction of dark current values from detector outputs). The pixels were manually selected from the channel 8 (green wavelength band) image and were restricted to locations near the middle image column. Pixels near the edge of tree crown image regions where foliage gaps are evident were avoided.

A constraint for the k parameter of the opaque surface model cannot be derived from measurements of physical tree crown characteristics. For values of k that are less than 1 and incident angles greater than zero, the value of the Minnaert reflectance function (Equation 4.8) increases without bound as the exitant angle increases from 0 and approaches $\frac{\pi}{2}$. When the projection vector is vertical and the $x_s y_s$-plane is horizontal, this causes brightening at and near the edge of the direct irradiation extent of opaque surface model instances for values of n greater than 1. For k less than 0.75 the appearance of the edge brightening effect is clearly unrealistic in model instances that are generated with the other parameters evaluated to the test values. Therefore, 0.75 was selected as the low limit for k.

6.3.2 Pattern Correlation Sensitivity Tests

The pattern correlation for pairs of different tree crown image model instances was computed in a series of tests in order to support the following investigations:

1. quantify the sensitivity of model instance pattern to full-scene variation and measurement error in θ_s, ϕ_s, ϕ_f, and K, and of variation in the locally variable parameters within the specified constraints

2. compare the pattern variation that is possible with the permeable volume model to that which is possible with the opaque surface model within the specified parameter constraints

For all tests, certain sensing geometry parameters were evaluated as follows: $f = 0.02461$ m, $p = 1.729 \times 10^{-5}$ m, $H = 518$ m, $\gamma = 0.0°$, $\beta = 0.0°$, $\phi_f = 277°$, $d_r = 0$, and $d_c = 0$.

The specified values for γ, β, d_r, and d_c imply a horizontal $x_s y_s$-plane and a vertical projection vector. Consequently, two model instances that were generated from these values and from the same value for b had the same extent and apex location, which permitted the computation of their pattern difference. Also, a horizontal $x_s y_s$-plane and a vertical projection vector permitted the effect of a change in the value of ϕ_f to be simulated with a change in the value of ϕ_s.

The specified values for H and ϕ_f were taken from the Headquarters flightline flight parameters (see Table 6.1). The values for f and p are those that are specified for the MEIS II (McColl, Neville, and Till 1983).

Tables 6.7 and 6.8 list parameter values on which the tests were based. For convenience, the pair (a, n) is treated as a single locally variable parameter in the tests. A range of natural variation or natural variation and measurement error, depending on the parameter, is defined by the low and high values. The parameter descriptions and value units are given in Section 4.3.

The intermediate test values for θ_s and ϕ_s are averages of the Headquarters flightline extrema values for those parameters (see Table 6.2). The difference between the θ_s extrema values is 0.52°, and we assume that the precision of the calculation of θ_s from the specified image

test param.	low	intermediate	high
θ_s	48.42	49.68	50.94
ϕ_s	113.07	115.49	117.91
K	0.00	0.10	0.40
D_{avg}	0.50	1.75	3.00
D_{max}	$1.50 \times D_{avg}$	$3.00 \times D_{avg}$	$4.50 \times D_{avg}$
X_{max}	$0.20 \times X_{end}$	$0.35 \times X_{end}$	$0.50 \times X_{end}$
B	0.20	0.35	0.50
k	0.75	0.88	1.00

Table 6.7: Test parameter values that are independent of horizontal radius.

	low	intermediate	high
$b = 1.00$ (23 extent pixels)			
(a, n)	(1.00, 2.00)	(4.50, 1.50)	(10.00, 1.50)
X_{end}	0.70	0.85	1.00
$b = 3.00$ (220 extent pixels)			
(a, n)	(3.00, 2.00)	(12.00, 2.00)	(18.00, 1.00)
X_{end}	2.00	2.50	3.00
$b = 6.50$ (1020 extent pixels)			
(a, n)	(6.50, 2.00)	(9.75, 2.00)	(13.00, 2.00)
X_{end}	2.00	2.50	3.00

Table 6.8: Test parameter values that are dependent on horizontal radius.

acquisition times is within $\pm 1.00°$ of the actual values. These considerations lead to high and low values for θ_s of $49.68° \pm 1.26°$. The difference between the ϕ_s extrema values is 0.82°, and we assume that the precision of the calculation of ϕ_s from the specified times is within $\pm 1.00°$ of the actual values. Furthermore, we estimate that practically all (99%) of the variation in ϕ_f to be within $\pm 2.54 \times 0.399° \approx \pm 1.01°$ of 277° on the basis of the standard deviation for ϕ_f, which is listed in Table 6.4. These considerations lead to high and low values for ϕ_s of $115.49 \pm 2.42°$. These ranges for θ_s and ϕ_s are greater than those that are estimated for the Sturgeon flightline by similar reasoning.

The choice of the test values for K was guided by the following information:

- on a clear day in August 1973 near Nunn Colorado with a 50° solar zenith angle, Oliver *at al.* (1975) measured K to be 0.125 in a wavelength band centred at 0.51 μm,
- Fraser *et al.* (1975) state that shortwave noontime clear sky values for K may reach 0.40, depending on location and season.

For each test value for b and each test parameter other than b, a model instance pair was generated that represented the high and low values for that parameter and the intermediate values for the other parameters. The pattern correlation for the permeable volume model instance pairs and the opaque surface model instance pairs are listed in Tables 6.9 and 6.10, respectively. These tables also rank the pattern correlation values from lowest to highest.

Table 6.9 shows that for $b = 3.00$ and $b = 6.50$, the pattern differences caused by the variation in θ_s, ϕ_s, and K were less than those caused by the variation in any one of the locally variable permeable volume model parameters other than B. For $b = 1.00$, the X_{end}-variation as well as the B-variation caused less pattern difference than the error variation in θ_s, ϕ_s, and K. Table 6.10 shows that for all values of b, the pattern differences caused by the variation in θ_s, ϕ_s, and and K were less than those caused by the variation in any one of the locally variable opaque surface model parameters.

For each test value for b, the pattern correlation was computed for all pairs of different model instances representing all combinations of high and low test values for more than one locally

	$b = 1.00$		$b = 3.00$		$b = 6.50$	
test param.	ρ	rank	ρ	rank	ρ	rank
θ_s	1.000	8	1.000	9	0.999	9
ϕ_s	0.994	4	0.994	6	0.995	6
K	0.995	5	0.997	8	0.997	8
(a, n)	0.920	**2**	0.925	**4**	0.973	**4**
D_{avg}	0.826	**1**	0.916	**3**	0.949	**3**
D_{max}	0.995	**5**	0.859	**2**	0.906	**2**
X_{end}	0.998	**7**	0.970	**5**	0.980	**5**
X_{max}	0.957	**3**	0.743	**1**	0.829	**1**
B	0.996	**6**	0.996	**7**	0.996	**7**

Table 6.9: Pattern correlation for high and low parameter value instances of the permeable volume model, ranked from lowest to highest with the ranks associated with locally-variable parameters in bold print (non-test parameters for each model instance pair were assigned the respective intermediate values).

	$b = 1.00$		$b = 3.00$		$b = 6.50$	
test param.	ρ	rank	ρ	rank	ρ	rank
θ_s	1.000	5	1.000	5	0.999	4
ϕ_s	0.995	3	0.995	3	0.996	3
K	0.999	4	0.999	4	0.999	4
(a, n)	0.918	**1**	0.918	**1**	0.972	**2**
k	0.979	**2**	0.952	**2**	0.961	**1**

Table 6.10: Pattern correlation for high and low parameter value instances of the opaque surface model ranked from lowest to highest with ranks associated with locally-variable parameters in bold print (non-test parameters for each model instance pair were assigned the respective intermediate values).

variable parameter. For the permeable volume model, sets of 2 to 6 locally variable parameters were involved in the test, with each set consisting of the locally variable parameters associated with the lowest pattern correlation values in Table 6.9 (the opaque surface model has only two locally variable parameters if (a, n) is treated as a single parameter). Summaries of the results for the permeable volume model and the opaque surface model are listed in Tables 6.11 and 6.12, respectively.

ρ	$b = 1.00$	$b = 3.00$	$b = 6.50$
lowest 2 (4 instances, 6 pairs)			
min	0.476	0.614	0.749
max	0.943	0.998	0.998
avg	0.777	0.807	0.875
lowest 3 (8 instances, 28 pairs)			
min	0.350	0.523	0.690
max	0.969	0.997	0.998
avg	0.784	0.814	0.881
lowest 4 (16 instances, 120 pairs)			
min	0.340	0.315	0.611
max	0.994	0.997	0.998
avg	0.795	0.761	0.872
lowest 5 (32 instances, 496 pairs)			
min	0.330	0.243	0.550
max	1.000	1.000	1.000
avg	0.801	0.766	0.874
all 6 (64 instances, 2016 pairs)			
min	0.223	0.231	0.514
max	1.000	1.000	1.000
avg	0.793	0.768	0.874

Table 6.11: Summaries of the pattern correlation values for all pairs of different permeable volume model instances drawn from the set of instances representing all combinations of high and low test values for the 2 to 6 locally-variable parameters associated with the lowest pattern correlation values in Table 6.9 (non-test parameters for each model instance pair were assigned the respective intermediate values).

The comparison of Tables 6.11 and 6.12 with Tables 6.9 and 6.10, respectively, shows that for both versions of the tree crown image model, covariation of the values of two or more locally variable parameters can cause larger pattern differences among model instances than variation

ρ	$b = 1.00$	$b = 3.00$	$b = 6.50$
min	0.894	0.878	0.921
max	0.987	0.996	0.972
avg	0.931	0.921	0.957

Table 6.12: Summaries of the pattern correlation values for all 12 pairs of different opaque surface model instances drawn from the set of 4 instances representing all possible combinations of high and low test values for (a, n) and k (non-test parameters for each model instance pair were assigned the respective intermediate values).

in the value of any single parameter.

For each test value for b and both versions of the tree crown image model, pattern correlation values were computed for model instances generated with the extreme test values for (θ_s, ϕ_s) and the intermediate test values for the other parameters (Table 6.13).

b	1.00	3.00	6.50
perm. ρ	0.988	0.988	0.990
opaq. ρ	0.989	0.990	0.991

Table 6.13: Pattern correlation values for permeable volume model (perm. ρ) and opaque surface model (opaq. ρ) instances generated with extreme test values for (θ_s, ϕ_s): $(48.42, 113.07)$ and $(50.94, 117.91)$. Parameters other than θ_s, ϕ_s, and b were assigned the respective intermediate test values.

Table 6.13 shows that the pattern differences caused by natural variation and estimated measurement error in θ_s, ϕ_s, and ϕ_f over the full Headquarters scene were less than those caused by variation in the values of most of the individual locally variable parameters (*cf.* Tables 6.9 and 6.10) and covariation of two or more locally variable parameters (*cf.* Tables 6.11 and 6.12).

Table 6.14 lists pattern correlation values for model instance pairs consisting of a permeable volume model instance and an opaque surface model instance. The pattern correlation values in Table 6.14 are grouped according to the values of D_{avg}, D_{max}, X_{max}, and k that the model instances were generated from. Opaque surface model instances generated from either the high or low test value for k have a relatively high pattern correlation (consistently greater than 0.9) with permeable volume model instances that represent a high density of scattering elements close to the surface of the crown envelope, which are generated from high test values for D_{avg}

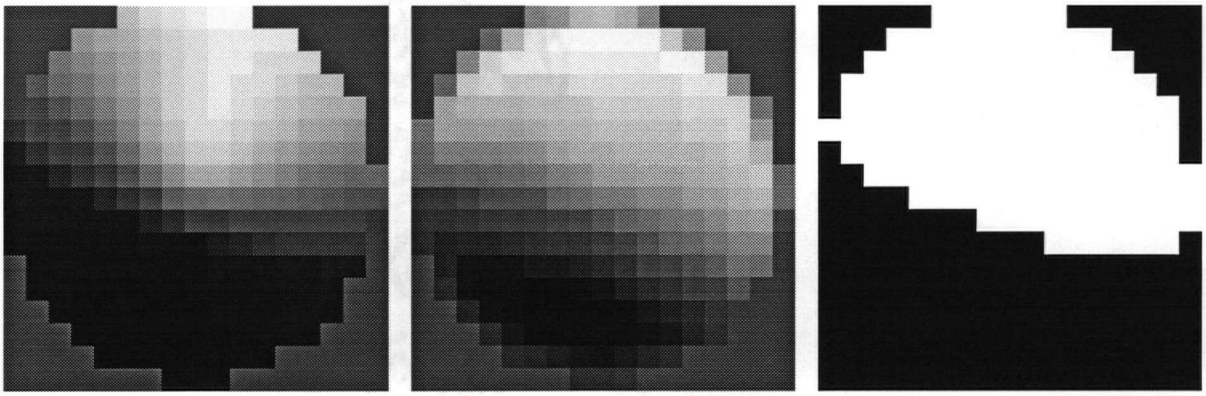
and $D_{\max}$ and the low test value for $X_{\max}$. However, permeable volume model instances that represent a low density of scattering elements far from the surface of the crown envelope, which are generated from low test values for D_{avg} and $D_{\max}$ and the high test value for $X_{\max}$, have a relatively low pattern correlation (consistently less than 0.9) with opaque surface model instances generated with either the low or high test value for k. These results show that when the values of locally variable parameters are restricted to the specified constraints, the patterns that are produced with the opaque surface model are nearly subsumed by the patterns that are produced by the permeable volume model.

(a, n)	$b = 1.00$	$b = 3.00$	$b = 6.50$
high D_{avg}, high $D_{\max}$, low $X_{\max}$; high k			
low	0.957	0.952	0.950
inter.	0.951	0.927	0.941
high	0.933	0.936	0.938
high D_{avg} high $D_{\max}$ low $X_{\max}$; low k			
low	0.967	0.983	0.985
inter.	0.947	0.976	0.987
high	0.933	0.956	0.988
low D_{avg}, low $D_{\max}$, high $X_{\max}$; high k			
low	0.525	0.741	0.902
inter.	0.542	0.818	0.894
high	0.508	0.767	0.867
low D_{avg}, low $D_{\max}$, high $X_{\max}$: low k			
low	0.368	0.607	0.847
inter.	0.433	0.785	0.876
high	0.584	0.607	0.884

Table 6.14: Correlation coefficients for pairs of permeable volume model and opaque surface model instances with the same extents (parameter values not given in the table are intermediate).

Figure 6.23 illustrates the basic differences between the two versions of the tree crown image model, and helps to explain the test results in Table 6.14. In the permeable volume model instance the non-directly irradiated portion of the extent is brighter because of penetration of light through the crown, and the modelled radiance near the edge of the extent falls off because of a decrease in light scattering element area density. Permeable volume model instances can

be generated so that they are more like opaque surface model instances by increasing the scattering element area density and concentrating it closer to the surface of the crown envelope. This reduces the light penetration through the crown and the crown edge reflectance fall-off. However, the opaque surface model cannot model any amount of light penetration through the crown or crown edge reflectance fall-off.



(a) (b) (c)

Figure 6.23: Tree crown model instances generated with intermediate test parameter values: (a) opaque surface version, (b) permeable volume version (c) direct irradiation locations.

6.3.3 Evaluation of b_i and η_i

The desired precision for estimates of the radius of the largest crowns was evaluated to 0.75 m, and η_{adj} was evaluated to 0.78 according to Equation 5.11. The minimum b-value increment was evaluated to 1.5 pixels, or $1.5 \times 0.364\,\text{m} = 1.09\,\text{m}$ and, using the algorithm described in Figure 5.11, the sequence of b-values and associated η-values listed in Table 6.15 was computed.

i	1	2	3	4	5	6	7	8	9	10
b_i (metres)	1.00	1.55	2.09	2.64	3.18	3.73	4.28	4.84	5.48	6.21
η_i	0.65	0.74	0.79	0.83	0.85	0.87	0.88	0.88	0.88	0.88

Table 6.15: Sequence of b-values and associated η-values computed from $\eta_{\text{adj}} = 0.78$, $(\triangle b)_{\text{min}} = 1.5 \times 0.364$, $b_{\text{min}} = 1.00$, and $b_{\text{max}} = 6.50$.

6.3.4 Evaluation of ρ_{diff}, SS_{E}, and SS_{M}

A set of 11 subimages centred on tree crown images with exceptionally regular appearance and representing a variety of sizes and column offsets was manually selected from the Headquarters flightline channel 8 image. Channel 8 (the visible green wavelength channel) was selected because it is one of the two channels that exhibits the least visible noise and greatest contrast, and the other such channel (channel 3, the near-infrared wavelength channel) exhibits less contrast within crown image regions than does channel 8. The tree crown apex location within each subimage and the crown radius was manually estimated. The procedure described in Figure 5.10 was used to generate a model instance set (a set of uniform-extent subsets) for each of the column offset-based recognition regions containing the sample tree crown images. Only the permeable volume version of the crown image model was used. In the generation of each model instance set, input parameters of the procedure were evaluated as follows:

- $\rho_{\text{diff}} = 1.0$
- $d_{\text{r}} = 0$ and d_{c} to the value corresponding to middle column of the column offset-based recognition region (*cf.* Section 6.3.5)
- b_i and η_i to values from Table 6.15 that bracket the estimated crown radius for the sample tree images within the recognition region
- f, p, H, γ, β, and ϕ_{f} to the fixed test values specified in Section 6.3.2
- θ_{s}, ϕ_{s}, and K to the intermediate test values specified in Section 6.3.2
- the constraints on locally variable parameters to the overall scene values specified in Section 6.3.1

For each model instance set, the procedure was run with different values for SS_{E} (in increments of 10) and $SS_{\text{M}} = 1$ in order to estimate the minimum value at which the subsets E′ of E (*cf.* step (3.2) of Figure 5.10) are of maximum size for the given inputs. This produced an

estimate of $SS_E = 60$. The procedure was then run with $SS_E = 60$ and different values of SS_M (in increments of 10) in order to estimate the minimum value at which the subsets M′ of M (*cf.* step (3.4.4) of Figure 5.10) are of maximum size for the given inputs (for $\rho_{diff} = 1.0$ this is the point where M begins to contain non-random duplicate model instances). This produced an estimate of $SS_M = 150$.

Model instance sets that were generated with $SS_E = 60$ and $SS_M = 150$ were used to derive a working value for ρ_{diff}. For each sample tree crown image, the uniform-extent subset that best coincided with the tree crown image extent (with the crown apex as the registration feature), was manually selected and the correlation coefficient for the tree crown image and each of the 150 elements of the uniform extent subset was computed. Table 6.16 lists the minimum and maximum correlation coefficient obtained for each sample tree crown image (ρ_{min} and ρ_{max}, respectively), the value of d_c that was used to generate the corresponding model instance set, and the locally variable parameter values of the model instance for which ρ_{max} was obtained.

sample	b (metres)	d_c (pixels)	ρ_{min}	ρ_{max}
1	1.55	-410.5	0.906	0.949
2	1.55	135.5	0.865	0.926
3	3.73	9.5	0.779	0.918
4	3.18	-198.5	0.776	0.910
5	3.73	10.5	0.677	0.891
6	3.18	142.5	0.643	0.864
7	1.55	38.5	0.717	0.859
8	6.21	33.5	0.745	0.858
9	2.64	219.5	0.545	0.853
10	2.09	-444.5	0.571	0.814
11	1.55	-391.5	0.581	0.751

Table 6.16: Sample tree crown image correlation coefficients.

Values of ρ_{min} and ρ_{max} that were similarly obtained for the channel 3 sample crown images were generally lower than those listed in table 6.16.

The values of the locally variable parameters (other than b) for the model instances for which $\rho = \rho_{max}$ varied widely with respect to the specified constraints (a/b: 1.32 to 8.64, n: 1.03 to 2.15, X_{end}: 1.18 to 2.45 m, X_{max}/X_{end}: 0.28 to 0.49, D_{avg}: 0.52 to 2.62 m^{-1}, D_{max}/D_{avg}: 1.61

to 4.41, B: 0.27 to 0.48). This is evidence that the constraints are not excessively loose relative to the natural variation of tree crown images.

Based on the values of $\rho_{\max}$ in Table 6.16, the working value for ρ_{diff} was evaluated to 0.95. By generating the model instance sets with different values of SS_{M} (in increments of 10), the minimum value of SS_{M} at which the subsets M′ of M are of maximum size was estimated as 60 for $\rho_{\text{diff}} = 0.95$.

A display triplet showing the sample tree crown subimage, the subimage with the boundary and apex location of the highest-ρ model instance superimposed in white, and the subimage with the pattern of the highest-ρ model instance superimposed at the same location was produced for each sample tree crown subimage (Figure 6.24). The y-intercept and slope of the straight-line least squares regression of tree crown image pixel values on coincident model instance pixel values were computed and used to transform the model instance patterns before the superimposition (the slope was applied as a gain and the y-intercept as an offset). The display triplets provide a visual impression of how well the extent and pattern of real tree crown images can be matched by instances of the permeable volume tree crown image model.

6.3.5 Definition of Recognition Regions

As illustrated in Section 4.1.2, a major geometric effect in the experimental MEIS II image data is along-row relief displacement (the angle between the projection vector and the optical axis can be as high as 20°). This indicates that the recognition regions should be at least partially based on image column offset (d_{c}) intervals. The maximum d_{c}-interval width that produces regions within which the along-row tree crown relief displacement can be considered as practically constant was estimated to be 51 pixels. This interval width divides the 1024 column MEIS II image into 20 adjacent along-column image 'strips.' Three different crown envelopes were defined with $b = 1.00$, 3.73, and 6.21, and with values for $\frac{a}{b}$ and n that produced the most vertically elongated shapes possible within the parameter constraints. For a specified whole number s of along-column strips (all of uniform width except for possibly the first and

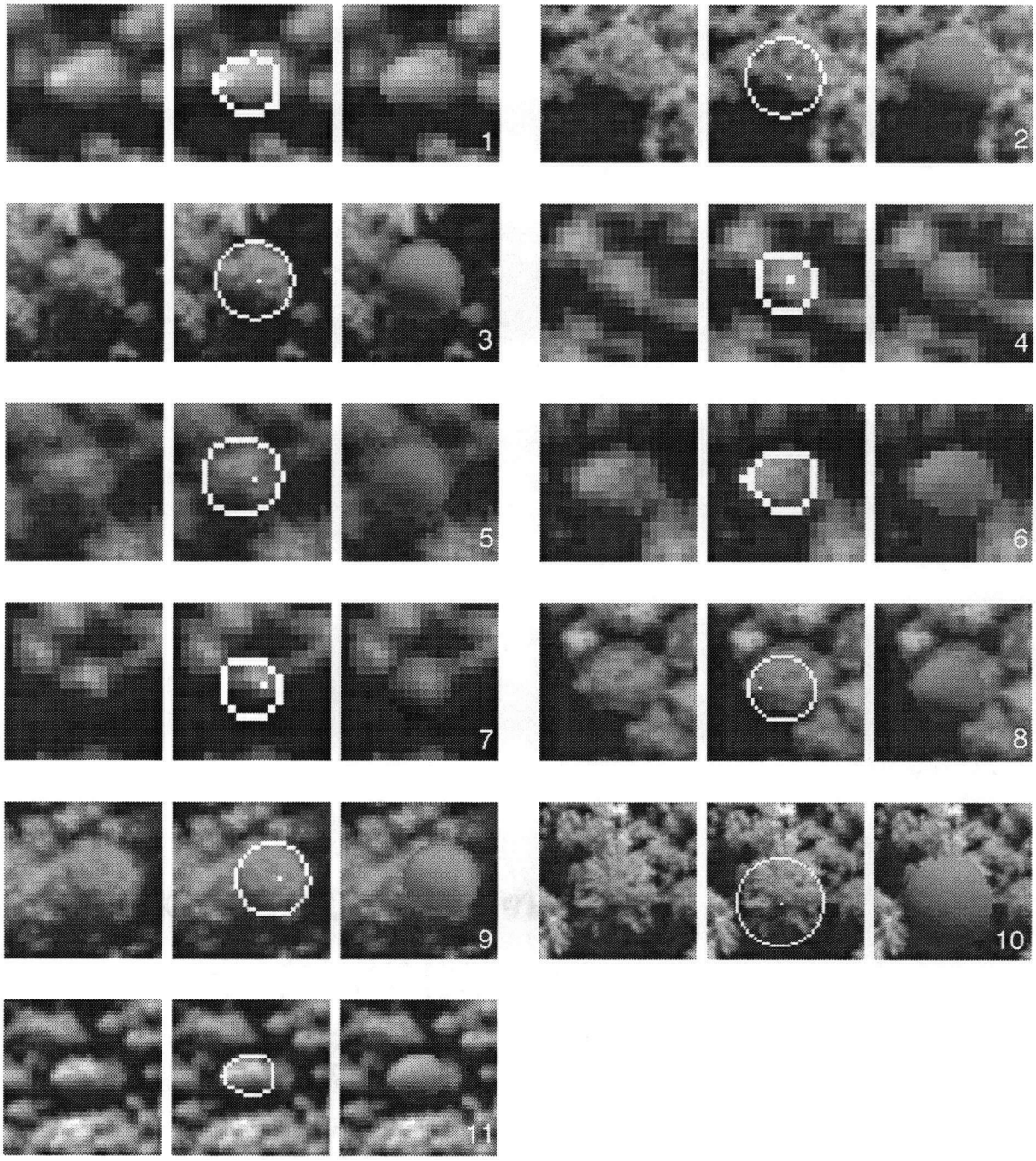


Figure 6.24: Sample tree crown subimage display triplets. Channel 8 subimage (left), subimage with boundary and apex location of highest-ρ model instance (middle), and subimage with model instance pattern (right).

last when $1024/s$ is not a whole number) a model instance was computed for each of the three crown envelope definitions and for d_c evaluated to the column coordinate of the start and middle of each strip (d_r was evaluated to 0.0 for each instance). The extent similarity values of model instances with the same crown envelope definition, adjacent d_c-values, and aligned on the apex location were computed. The value of s was incremented from 10 until all of the computed extent similarity values were no less than the corresponding η_i-values listed in Table 6.15. The extent similarity values for $s = 20$ are listed in Table 6.17.

d_c (pixels)		$(b, \frac{a}{b}, n)$		
1	2	(1.00, 10, 1.5)	(3.73, 4, 2.0)	(6.21, 2, 2.5)
0.0	25.5	1.00	0.88	1.00
25.5	51.0	0.75	0.89	0.93
51.0	76.5	0.75	0.88	1.00
76.5	102.0	0.77	0.89	0.94
102.0	127.5	0.77	0.89	1.00
127.5	153.0	0.94	0.90	0.93
153.0	178.5	0.78	0.89	0.94
178.5	204.0	0.81	0.87	1.00
204.0	229.5	0.96	0.91	0.93
229.5	255.0	0.84	0.90	0.99
255.0	280.5	0.84	0.92	0.94
280.5	306.0	0.84	0.92	0.99
306.0	331.5	0.94	0.93	0.94
331.5	357.0	0.84	0.92	0.99
357.0	382.5	0.88	0.97	0.94
382.5	408.0	0.96	0.94	0.99
408.0	433.5	0.89	0.89	0.94
433.5	459.0	0.86	0.96	0.99
459.0	484.5	0.94	0.94	0.99
484.5	512.0	0.86	0.90	0.94

Table 6.17: Extent similarity of model instances generated for different column offsets (51 pixel d_c increment).

The apex was used as the registration feature in computing these extent similarity values. The apex was also used as the registration feature in model instance generation and in the recognition procedures; the decision to do so was based on the assumption that the form and irradiation of the part of the tree crown that is near the apex is least likely to be influenced by

neighbouring tree crowns. Therefore, matches obtained for different model instances and same tree crown image would be compared to each other in the recognition procedure on the basis of the most ideal part of the tree crown.

The pattern correlation values of model instance pairs representing extreme values of θ_s, ϕ_s, and ϕ_f (*cf.* Table 6.13) are considerably larger than the sample tree crown image correlation coefficients (*cf.* Table 6.16). Therefore, we do not regard refinement of the column offset-based recognition regions defined in Section 6.3.5 on the basis of variation in θ_s, ϕ_s, and ϕ_f to be warranted.

If the d_c-value associated with a column offset-based recognition region is that of the along-row midpoint of the region, then 51-column wide regions (*i.e.*, 20 strips) represent an along-row deviation in the angle between the projection vector and the optical axis of 0.99° to 1.03°. These deviations are over twice the standard deviations of the roll (γ) and pitch (β) INS orientation angles listed in Table 6.4. Therefore, we do not regard refinement of the column offset-based recognition regions on the basis of variation in γ and β to be warranted.

Knowledge of the scene that is more specific than knowledge of its Forest Region and Section has not been obtained. Therefore, we do not have grounds for specializing the constraints on the locally variable parameters to different parts of the scene and for subdividing the recognition regions on this basis.

6.3.6 Summary

Twenty model instance sets were generated for each of the Sturgeon and Headquarters flightlines (one model instance for each 51-column wide recognition region in each flightline). Only the permeable volume model was used. The apex was used in computing the extent similarity values in the model instance set generation procedure (steps (3.2) and (3.3) in Figure 5.10). The value of d_c used to generate a model instance set was that of the along-row midpoint of the corresponding recognition region, and d_r was evaluated to 0 in each case. The other inputs to the model instance set generation procedure were evaluated as follows:

- the constraints on locally variable parameters to the overall scene values specified in Section 6.3.1
- $f = 0.02461\,\text{m}$, $p = 1.729 \times 10^{-5}\,\text{m}$, $H = 518\,\text{m}$, $\gamma = 0.0°$, $\beta = 0.0°$, (the fixed test values specified in Section 6.3.2)
- Sturgeon: $(\theta_s, \phi_s) = (51.79°, 112.32°)$, Headquarters: $(\theta_s, \phi_s) = (49.68°, 115.449°)$ (the averages of the extrema values for the corresponding flightline as listed in Table 6.2)
- $K = 0.10$ (the intermediate test value specified in Section 6.3.2)
- Sturgeon: $\phi_f = 321°$, Headquarters: $\phi_f = 277°$ (the flightline azimuth values listed in Table 6.2)
- b_i and η_i to the values specified in Table 6.15.
- $\rho_{\text{diff}} = 0.95$, $SS_E = 60$, $SS_M = 60$ (the values specified in Section 6.3.4)

The value assigned to K is relatively uncertain. However, a wide variation in the K-value produced relatively small pattern differences in the pattern correlation sensitivity tests (*cf.* Table 6.9), so we do not regard this uncertainty as being of much consequence.

The number of uniform-extent subsets in a model instance set decreases with the associated column offset. For example, the number of uniform extent subsets in the model instance sets associated with $d_c = 484.5$ pixels (a large offset from middle image column) and $d_c = 25.5$ pixels (a small column offset) are respectively 43 and 11. This occurs because as the column offset decreases the amount of modelled relief displacement also decreases and, consequently, variation in the a and n parameters has an decreased effect on the image extent of the model instances.

Conversely, the average number of model instances in the uniform-extent subsets of a model instance set increases slightly with a decrease in the associated column offset. For example, the average number of model instances in the uniform-extent subsets of model instance sets associated with $d_c = 484.5$ pixels and $d_c = 25.5$ pixels are respectively 2.7 and 3.6. This occurs because as the associated column offset and number of uniform-extent subsets for a model

instance set decreases, a larger variation in a and n is represented by each uniform-extent subset.

6.4 Tuning the Initial Matching Procedure

The tuning of the initial matching procedure consists of adjusting the value of the procedure parameters ρ_{thr} (the correlation coefficient minimum threshold), ρ_{sthr} (the correlation coefficient minimum threshold for initiating the full-resolution search of the image-space), and $dfrac$ (the extent dimension fraction), and the value of η_{group} (the extent similarity value used to define the uniform-extent subset groups).

In order to support the tuning task, six 51-by-280 subimages that represent a variety of stand types were selected from the the Headquarters flightline channel 8 image recognition regions with the associated d_c-values -229.5, 25.5, and 280.5. Figure 6.25a is a mosaic of the test subimages.

The initial matching procedure was run on the test subimages with different values for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}})$ (although η_{group} is not strictly an adjustable parameter of the matching procedure we group it with those parameters for convenience). The crown apex was used as the registration feature since it was also used in the model instance set generation (*cf.* Section 6.3.6). Figure 6.25b is a correlation coefficient map in which a non-zero value at location (x, y) is the maximum correlation coefficient for that location in the test subimage mosaic over all tested uniform-extent subsets, for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}}) = (0.00, 0.00, 0.00, 1.00)$. The map was created by linearly scaling the correlation coefficient interval $[0, 1.0]$ to the image pixel value interval $[0, 255]$. This set of parameter values implies an exhaustive search through the image-space and uniform-extent subsets (*i.e.,* the basic matching procedure outlined in Figure 5.13). The ρ_{thr} parameter was evaluated to 0.00 because a negative value would permit matches to be defined for inverse linear relationships between model instance pixel values and coincident image pixel values, and such matches would clearly be invalid.

A visual comparison of Figure 6.25a and Figure 6.25b shows that tree crown images generally

(a)

(b)

(c)

(d)

(e)

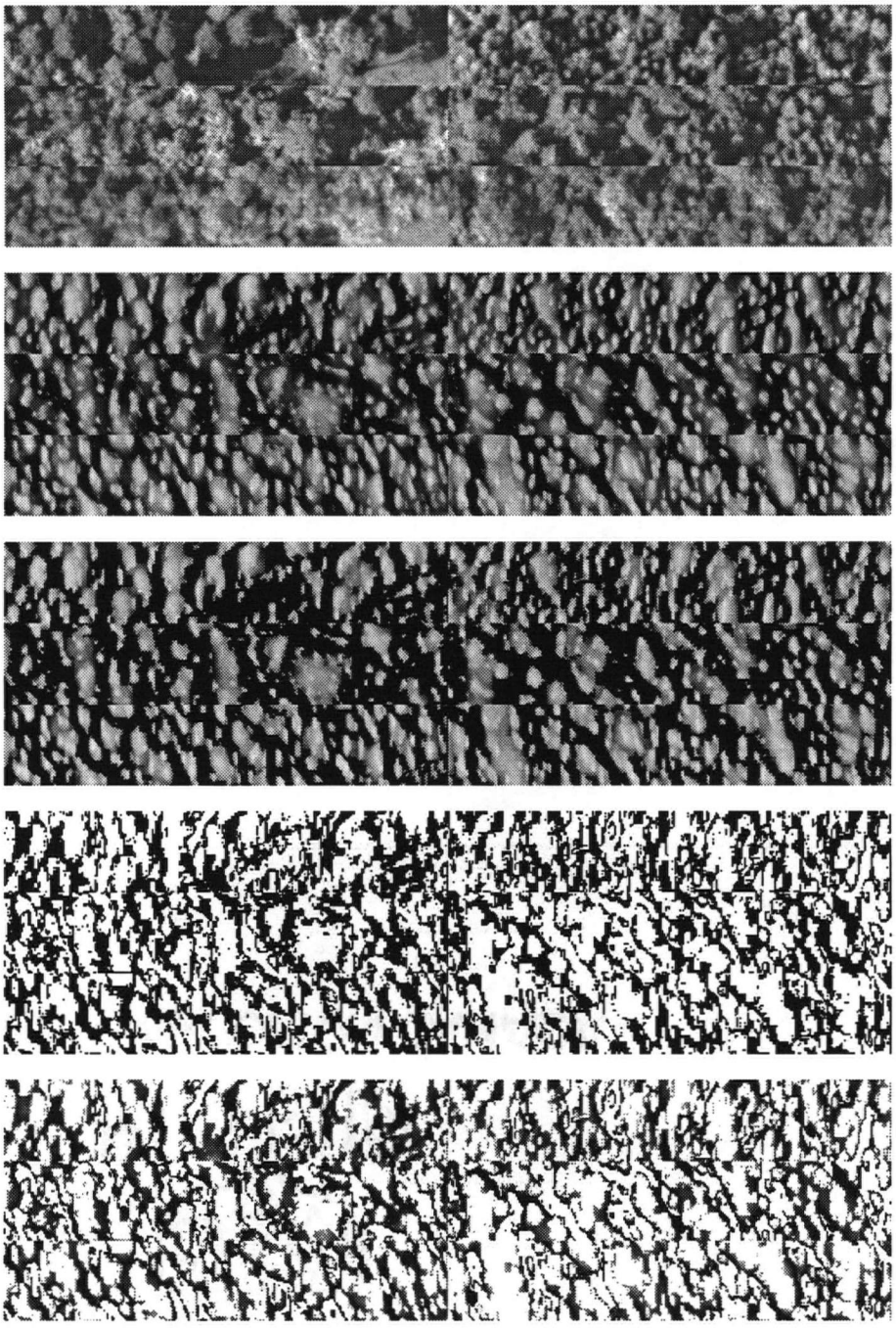


Figure 6.25: Test subimage mosaic and maps: (a) test subimage mosaic, (b) overall maximum correlation coefficient map for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}}) = (0.00, 0.00, 0.00, 1.00)$, (c) overall maximum correlation coefficient map for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}}) = (0.00, 0.50, 0.25, 0.75)$, (d) map of locations where map c pixel values are greater than zero, and (e) map b masked by map d.

coincide with local maxima in the correlation coefficient surface, and that the spatial extent of a hill in the surface is generally related to the spatial extent of the coincident tree crown image.

Figure 6.25c is the map of maximum correlation coefficients for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}}) = (0.00, 0.50, 0.25, 0.75)$. In this case, the 32, 11, and 28 uniform extent subsets associated with $d_c = -229.5$, 25.5, and 280.5 were put into 15, 8, and 15 groups, respectively. All of the local maxima in Figure 6.25b appear to be present in Figure 6.25c. This was confirmed by generating a map of all non-zero pixels in Figure 6.25c (the white pixels in Figure 6.25d) and using this map to mask Figure 6.25b. The result (Figure 6.25e) shows that the unmasked regions generally have relatively low pixel values in Figure 6.25b.

The adjustable parameter values that were used to generate Figure 6.25c were estimated to be those that minimized the amount of computation involved in the matching procedure while resulting in the retention of all of the local maxima that appear in Figure 6.25b. This estimate was arrived at by generating and inspecting maximum correlation coefficient maps that were generated from a set of value combinations for ρ_{sthr}, $dfrac$, and η_{group}.

6.5 Training Set and Feature Selection

The training set was selected from the image regions illustrated in Figure 6.26. These regions represent a wide range of tree species, tree size, and column offset, as well as some non-forested areas that contain incorrect initial recognition instances. Each training region was defined within a single recognition region and none of them intersect a ground sample plot image region.

Figure 6.27 illustrates the initial recognition instances for one of the training regions illustrated in Figure 6.26. The channel 8 subimage is pictured with the boundaries and apex locations of the matching model instances overlaid in white.

For each training region, the initial recognition instances were selected which were considered with a high degree of confidence by the author to be either correct or incorrect. Over all training regions, 725 correct recognition instances and 1503 incorrect recognition instances were selected.

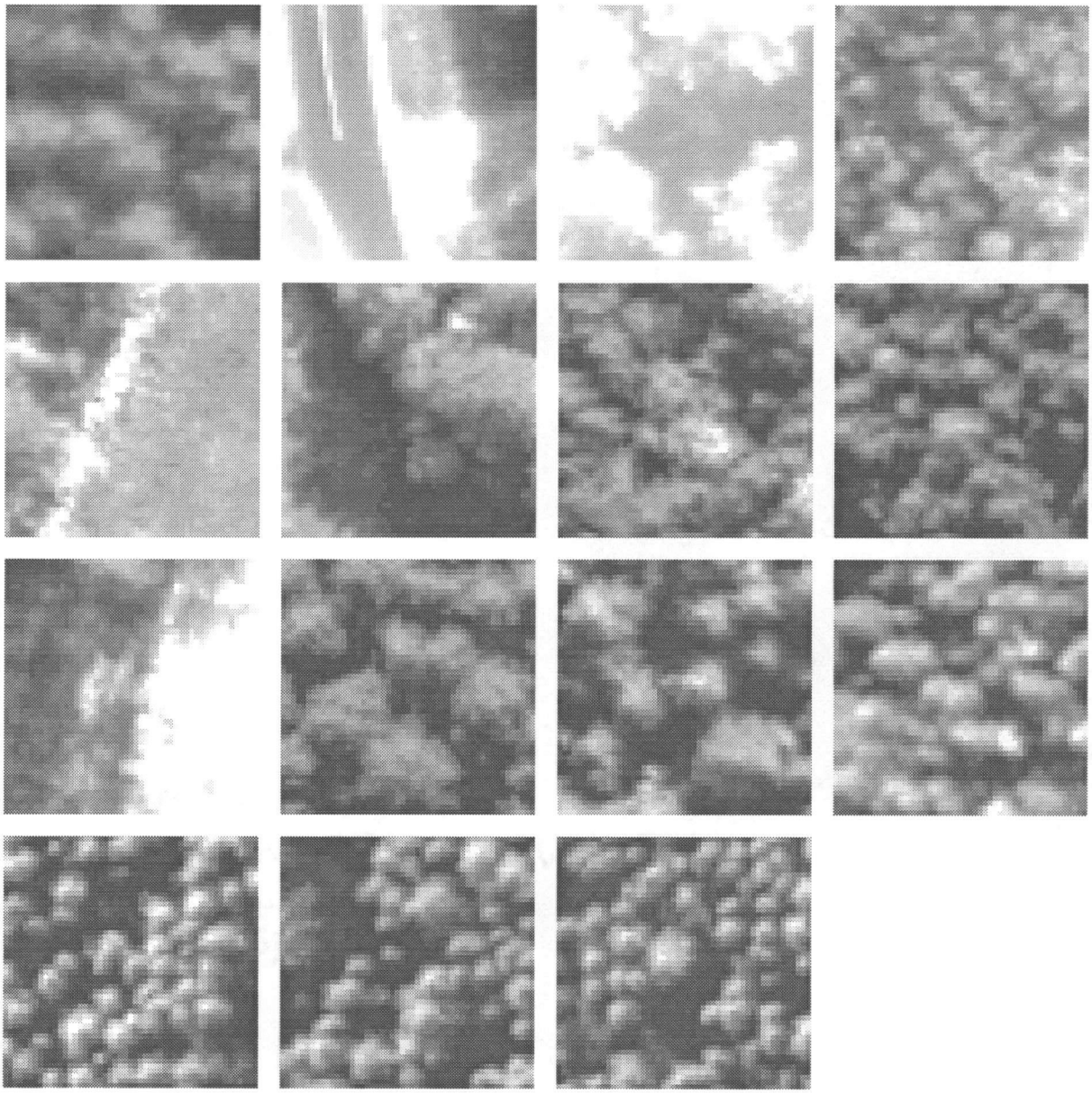


Figure 6.26: Image regions from which the training set was selected. Each region is 51-by-51 pixels in size, and was defined within a single recognition region. The displayed subimages are from channel 8.

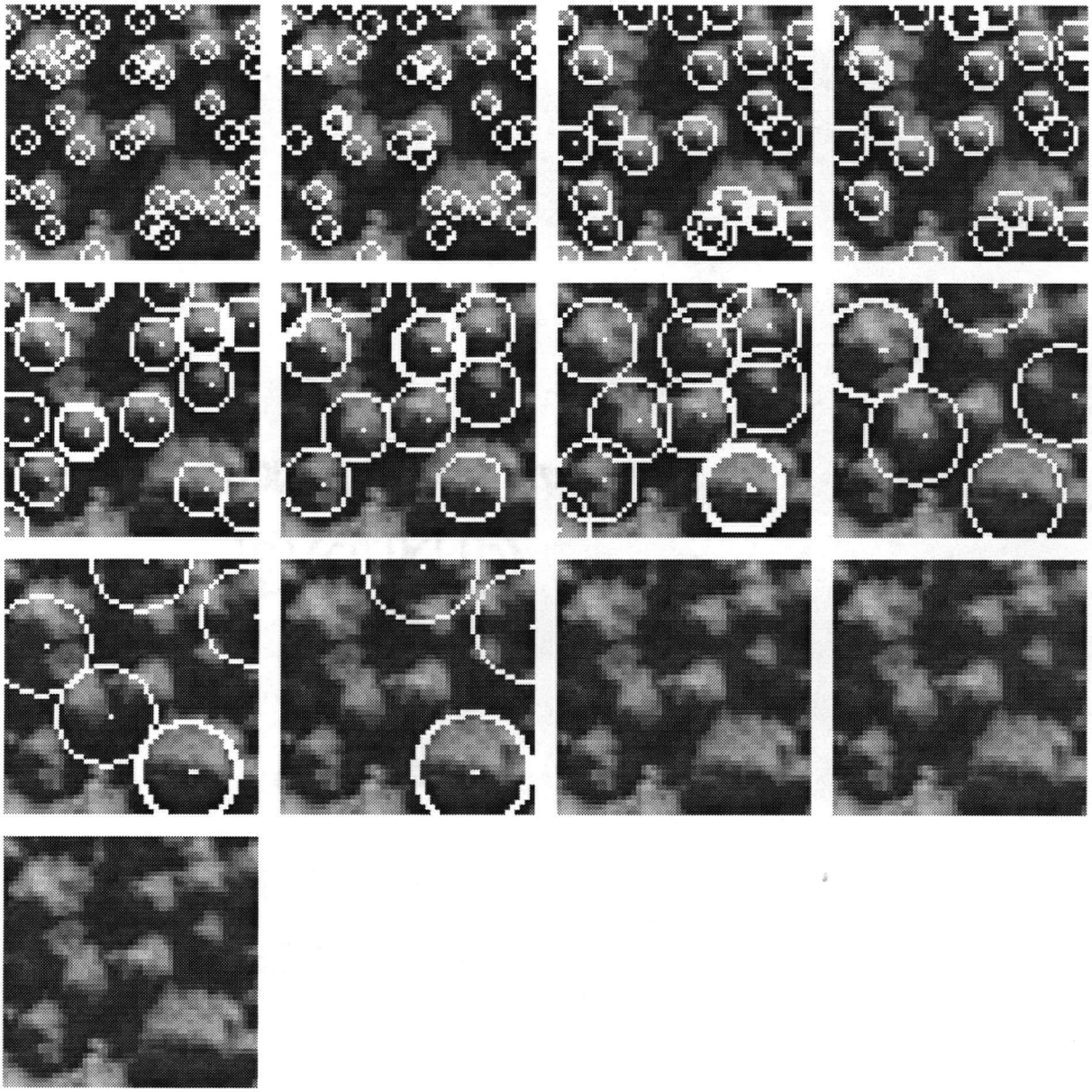


Figure 6.27: Initial recognition instances for a training set region. The region is from the $d_c = 76.5$ pixels recognition region. The recognition instances for all 13 uniform-extent subsets are pictured.

725 incorrect recognition instances were randomly selected from the initial set of 1503 incorrect recognition instances. This was done because the training regions were not randomly selected from the scene (for example, four of the training regions were deliberately selected to contain unforested areas), and so the relative numbers of correct and incorrect recognition errors do not necessarily reflect the relative *a priori* probabilities of correct and incorrect recognition.

50 recognition instances were randomly selected from each of the sets of 725 correct and incorrect recognition instances; a test set was made from these 100 recognition instances and the training set was made from the remaining $2 \times 675 = 1350$ recognition instances.

The probability of correct recognition for the correct and incorrect recognition training exemplars was evaluated respectively to 1.0 and 0.0.

The VSM learning procedure was initially run with the following 9 features: R^2_{ch8}, C_{ch8}, C_{ch3}, $\bar{D}_{ch7}$, $\bar{D}_{ch8}$, $\bar{D}_{ch4}$, $\bar{D}_{ch3}$, $\bar{D}_{ch3}/\bar{D}_{ch4}$, and $\bar{D}_{ch8}/\bar{D}_{ch7}$ (see Section 5.4.2 for the feature definitions). The feature that was assigned the smallest weight by the VSM learning procedure was removed and the procedure was then re-run. The removal of a feature and re-running of the VSM procedure was repeated as long as the cross-validation error did not increase and the classification accuracy for the test set did not decrease. The final features were the following: R^2_{ch8}, C_{ch8}, C_{ch3}, and $\bar{D}_{ch3}/\bar{D}_{ch4}$, with the respective weights 9.24, 0.21, 0.18, and 2.77. The final optimized value r was 0.73. The initial and final values for the cross-validation error E were respectively 454.52 and 344.42. The fractions of correct and incorrect recognition test exemplars that were initially correctly classified were respectively 0.74 and 0.78 (0.76 overall). After the optimization, the fractions were respectively 0.90 and 0.88 (0.89 0verall).

The cross-validation error and the test set classification accuracy were both essentially unchanged when R^2_{ch8} was replaced by $R^2_{ch8}/R^2_{\mathrm{max},ch8}$ (an X-axis interval width of 5 was used to compute $R^2_{\mathrm{max},ch8}$), and worsened when C_{ch8} and C_{ch3} were replaced by C'_{ch8} and C'_{ch3}; therefore, the final features listed in the previous paragraph and the associated weights and r-value were used in the evaluation of the correct recognition probability.

The VSM learning procedure was run with the final features and a training set consisting of

the 675 correct recognition exemplars and 1453 incorrect recognition exemplars. A final cross-validation error of 464.92 was obtained, and the fractions of correct and incorrect recognition test exemplars that were correctly classified were respectively 0.82 and 0.96 (0.89 overall). These results are no better than those obtained with the original training set and were produced with a larger training set; therefore, the original training set was used in the evaluation of the correct recognition probability.

In each run of the VSM learning procedure, all of the feature weights were initialized to 1.00, r was initialized to 0.80, the number of nearest neighbours involved in the interpolation was 10, and the number of nearest neighbours used to compute the kernel width was 3.

6.6 Tuning the Conflict Resolution Procedure

The tuning of the conflict resolution procedure consists of defining a window grid and margin dimension, and adjusting the objective function component weights (*cf.* Section 5.5).

A grid that defines 51-by-51 pixel windows was chosen because those windows are large enough to contain several of the largest tree crowns, and the along-column window boundaries can be made to coincide with the recognition region boundaries. The training set image regions pictured in Figure 6.26 each coincide with one window.

The margin dimension was initially evaluated to 25 pixels. The objective function component weights were adjusted by computing final recognition results for several subimages, using a range of values (in increments of 5) for each of the weights. The sub-images were selected to represent wide canopy type and column-offset variation. Of the values that were tested, $(w_1, w_2, w_3) = (10, 5, 1)$ produced the best results over all subimages (according to visual inspection); therefore, this weight triplet was used in the main experiment. The results were then recomputed with decreasing values for the margin width (5 pixel increments were used). A margin width of 20 pixels was found to be the smallest at which no reduction in the quality of the results was evident; therefore this margin width was selected for the main experiment.

Figure 6.28 presents some examples of final recognition instances that were computed with

different objective function component weights, including the selected final weights. The weight triplets $(w_1, w_2, w_3) = (10, 1, 1)$, $(1, 10, 1)$, and $(1, 1, 10)$ were chosen to respectively emphasize minimization of overlap among interpreted crown extents, maximization of the average correct recognition probability value, and maximization of the area of the processing region covered by interpreted crown extents. The intended emphasis for each of these weight triplets is evident in the corresponding results.

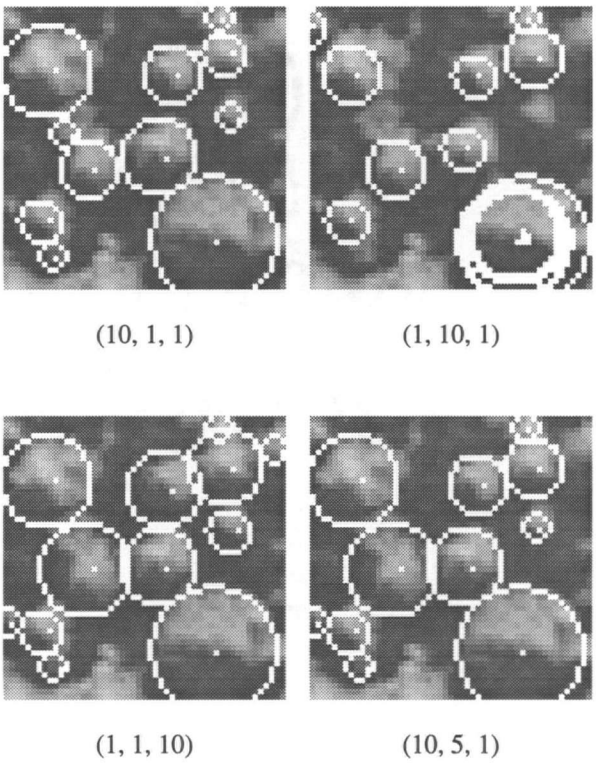


Figure 6.28: Final recognition instances corresponding to different objective function component weights. Results are shown for the training set region pictured in Figure 6.27. Each example is labelled with the corresponding value for (w_1, w_2, w_3). The weights are associated with components of the objective function as follows: w_1 - minimization of overlap among interpreted crown extents, w_2 - maximization of the average correct recognition probability value, w_3 - maximization of the area of the processing region covered by interpreted crown extents.

6.7 Results

An assessment of the automatic recognition results for 14 sample plots is presented in Section 6.7.1. These sample plots will be referred to as the *test sample plots* and the sample trees in them will be referred to as the *test sample trees*. The test sample plots were selected from the 15 sample plots for which manual recognition results are available (the manual recognition experiment is described in Section 6.2). Sample plot 3, which was included in the manual recognition experiment, was omitted from the test set because it contains the two reference locations that are associated with multiple trees (*cf.* Section 6.1.2), and it is not clear how these cases were accommodated in the enumeration of the manual recognition errors.

An assessment of the automatic crown diameter estimates for the automatically recognized test sample trees is presented is Section 6.7.2.

The analysis of each set of results includes significance tests of various hypotheses about a population of trees that is randomly sampled by the test sample trees (the *test population*). The population corresponding to the full set of sample plots or to any subset of them has not been formally defined. However, the sample plots were selected on the basis of photo-interpretation and local knowledge so that they would represent as wide a range of non-plantation stand types as possible, given the resources that were available for field work and the geographic coverage of the image data (Cloney, Leckie, and Gillis 1994). As shown in Tables 6.18 and 6.19, the values for the different reference tree attributes, which are defined in Section 6.1.2, are distributed over the 340 test sample trees much as they are over all 548 sample trees. Therefore, we assume that the test population encompasses a substantial portion of the tree attribute-value combinations that are represented by the scene.

	dbh (cm)				height (m)				crown diameter (m)			
	min	max	μ	σ	min	max	μ	σ	min	max	μ	σ
all	8.2	59.0	23.4	9.7	6.5	33.3	17.8	4.2	1.0	10.2	3.6	1.7
test	8.9	59.0	23.8	9.7	6.5	33.3	18.5	4.2	1.1	10.2	3.5	1.7

Table 6.18: Ranges, means, and, standard deviations for interval-valued reference attributes for all 540 sample trees and for the 340 test sample trees.

	dominance		crowding			species		species context			
	dom.	co-dom.	low	med.	high	soft.	hard.	soft.	s./h.	h./s.	hard.
all	0.58	0.42	0.41	0.46	0.13	0.73	0.27	0.62	0.12	0.10	0.16
test	0.58	0.42	0.36	0.49	0.14	0.74	0.26	0.64	0.12	0.08	0.16

Table 6.19: Proportions of all 540 sample trees and of the 340 test sample trees with specified values for ordinal- and nominal-valued reference attributes. Except for species, all values for each attribute are represented. "Softwood" (soft.) and "hardwood" (hard.) respectively represent 9 and 10 different species.

The results assessments and analyses are summarized in Section 6.7.3.

6.7.1 Recognition Results

Figure 6.29 provides an overview of the recognition results. The figure was created from three 275-by-275 pixel channel 8 subimages, each of which corresponds to a 1 hectare subscene that contains a single sample plot. The displays on the left show the subimages with the sample plot boundary overlaid in white. The displays on the right show the subimages with the sample plot boundary and the boundaries and apex locations of model instances overlaid in white. Each overlaid model instance boundary illustrates the computed recognition of one tree crown and the estimate of its image region. Inspection of the figure reveals that tree crowns were successfully recognized over a wide range of crown size and crown image texture, even in cases where the crown boundary is weakly defined or only implicit for much of its length. Some errors are also evident (examples of errors are discussed and illustrated later in this section). The erroneous recognition of tree crowns within unforested parts of the subscenes is largely avoided.

Table 6.20 lists SUN SPARCstation 2 running times for different components of the recognition procedure as applied to the subimages pictured in Figure 6.29 (the times are actually for the minimum number of 51-by-51 pixel windows that cover the subimage extent, so in each case results were computed for an image extent that was slightly larger than the pictured subimage).

The matching component, which produces the initial recognition instances, was run with three different sets of parameter values (A, B, and C). The times for these three runs show

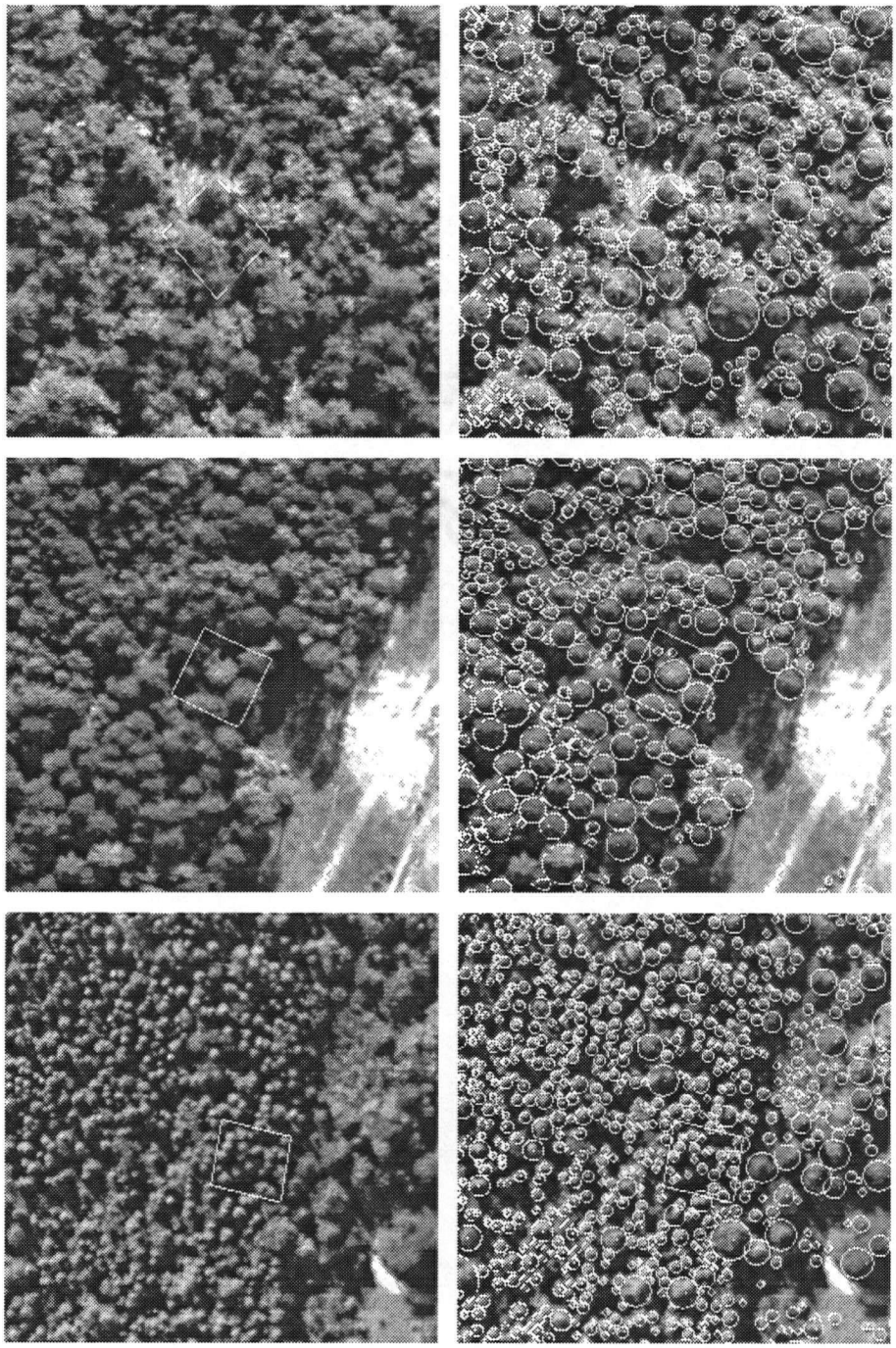


Figure 6.29: Channel 8 subimage and recognition results for 1 hectare subscenes that contain sample plots 3 (top), 7, and 21 (bottom). The column offsets of the plot centroids from the optical axis image point are respectively -8, -254, and 122 pixels.

subimage	matching A	matching B	matching C	correct rec. prob.	conflict resolution	total
plot 3	25:13	18:01	6:47	4:03	16:28	27:18
plot 7	43:11	29:14	10:27	6:32	26:30	43:30
plot 21	27:05	20:35	5:42	4:53	24:08	34:43
average	31:50	22:37	7:39	5:09	22:22	35:10

Table 6.20: Running times for different components of the recognition procedure as applied to the subimages pictured in Figure 6.29. The tabulated quantities are the elapsed (real) times in minutes and seconds on a SUN SPARCstation 2. The matching procedure was run with the following values for $(\rho_{\text{thr}}, \rho_{\text{sthr}}, dfrac, \eta_{\text{group}})$: A: $(0.00, 0.00, 0.00, 1.00)$, B: $(0.00, 0.00, 0.00, 0.75)$, and C: $(0.00, 0.50, 0.25, 0.75)$. The other components were run with the initial recognition instances generated with parameter tuple C. The conflict resolution component includes the integration of the results for the separate processing regions.

that a substantial savings in computation was obtained through a two-resolution search of the uniform-extent subsets as compared to a single-resolution search (*i.e.*, $\eta_{\text{group}} = 0.75$ vs. $\eta_{\text{group}} = 1.00$) and through a two-resolution search of the image space as compared to a single-resolution search (*i.e.*, $\rho_{\text{sthr}} = 0.50$ and $dfrac = 0.25$ *vs.* $\rho_{\text{sthr}} = 0.00$ and $dfrac = 0.00$). Parameter value set C was used to compute the recognition results that are discussed in this section and were used to obtain the running times of the other procedure components (the choice of this parameter value set is discussed in Section 6.4).

In general, the interpretation conflict resolution filter required more time than the sum of the times required for the version-C matching component and the correct recognition probability filter. The matching component required more time for the plot 7 subimage than for either of the two other subimages because of the relatively large absolute column offset for the plot 7 subimage. A relatively large column offset means that model instance sets with relatively large numbers of uniform-extent subsets are involved in the matching procedure (tree crown shape becomes a more significant variable as the absolute column offset increases). The interpretation conflict resolution component required more time for the plots 7 and 21 subimages than for the plot 3 subimage because the numbers of intermediate recognition instances were larger for the plots 7 and 21 subimages. In the case of plot 7, this was so because of a larger number of uniform extent subsets involved in the matching, and in the case of plot 21 this was so because

of the smaller size of the tree crowns and, consequently, the greater number of intermediate recognition instances in the processing windows.

For each of the 24 sample plots, a display showing the sample plot boundary and the recognition results (a *results display*), and a display of the sample plot boundary alone (a *plot boundary display*) were produced from copies of the same 100-by-100 pixel channel 8 subimage that was used to produce the reference display. The plot boundary displays are useful because they show the subimages with a minimum of annotation clutter. The plot boundary, reference, and results displays were assembled into triplets. A display triplet for each of the sample plots is presented in Appendix A (selected display triplets are reproduced later in this section).

The automatic recognition results for a sample plot were assessed by manually comparing the reference and results displays for that sample plot. A binary subimage in which only the reference apex locations are indicated was referred to in cases where the overlaid boundaries in the reference display were too cluttered to allow the coincident crown images to be clearly distinguished. In general, the tree crown image associated with a reference crown boundary or apex location could be identified within the channel 8 subimage. A false-colour composite display of channel 3, 4, and 8 subimages was also referred to in order to confirm the identification.

The manual recognition results for each test sample plot consist of a set of image locations, each of which corresponds to the interpreted location of a single tree crown. The interpreters were not required to delineate the crown image region. Omission errors were counted only in cases in which no interpreted tree crown locations were specified within the known image region of a single sample tree crown. A case in which multiple interpreted tree crown locations were specified within the known image region of a single sample tree crown never produced an omission error, and all but one of the locations were counted as commission errors.

In contrast to the manual recognition results, the automatic recognition results include estimates of crown image regions. In cases where multiple interpreted crown image regions coincide with a single sample tree crown image region, it is reasonable to consider the possibility that none of the interpreted image regions describe the sample tree crown image region reasonably

well. Therefore, the automatic recognition results were assessed twice, according to two different sets of rules.

The first assessment of the automatic recognition results, called the *liberal assessment* (Table 6.21), was intended to produce error counts that can be compared to those for the manual recognition results. In this assessment, an omission error was counted in all cases where an interpreted crown image region is not mostly included within a sample tree crown image region. In all cases where a sample tree crown image region intersects most of the area of $n > 1$ interpreted crown image regions, $n - 1$ commission errors were assessed. In a liberal assessment, example (a) of Figure 6.30 would produce zero omission errors and zero commission errors, example (b) would produce zero omission errors and two commission errors, example (c) would produce one omission error and one commission error, example (d) would produce zero omission errors and one commission error, and example (e) would produce three omission errors and one commission error. The actual tree crown would be counted as recognized only in examples (a), (b), and (d).

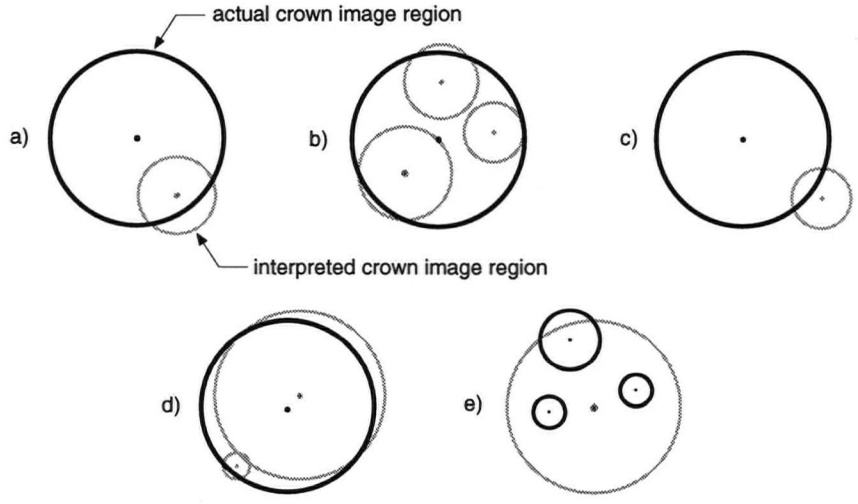


Figure 6.30: Schematic examples of relationships between actual and interpreted crown image image regions.

In the second assessment of the automatic recognition results, called the *conservative assessment* (Table 6.22), a model instance boundary that does not delineate a sample tree crown image region was counted as an omission error. A model instance boundary that coincides with an image region that is not clearly identified by the reference data as the image of a single tree crown was counted as a commission error. Some judgement was exercised in deciding whether or not a model instance boundary adequately delineates with a tree crown image region (the exact coincidence of model instance boundaries and reference boundaries was not insisted upon). In a conservative assessment, example (a) of Figure 6.30 would produce one omission error and one commission error, example (b) would produce one omission error and three commission errors, example (c) would produce one omission error and one commission error, example (d) would produce zero omission errors and one commission error, and example (e) would produce three omission errors and one commission error The actual tree crown would be counted as recognized only in example (d). The liberal and conservative assessments would be the same only for examples (c), (d), and (e).

Table 6.21 shows that the numbers of commission errors in the manual and the liberally assessed automatic recognition results are close (respectively 10% and 11% of the total number of sample trees). However, the number of liberally assessed automatic recognition omission errors is 2.8 times the number of manual recognition omission errors. Comparison of Tables 6.21 and 6.22 shows that the numbers of conservatively assessed errors are only slightly greater than the numbers of liberally assessed errors.

The automatic recognition results for six of the test sample plots are discussed below. Results that are among the best and the worst over all of the the test plots are included. The discussions explain the cause of various errors that occur in the automatic recognition results, and include descriptions of error enumeration under both assessment procedures. The display triplets for these sample plots are reproduced below, with additional annotation, to support the discussions.

The automatic recognition results for sample plots 21 and 7, which are discussed next, are

plot	man. int. pixel dim.	sample trees	manual om. errs	auto. om. errs	manual comm. errs	auto. comm. errs
1	original	15	0.6	4	5.8	4
2	original	14	0.6	3	2.6	3
4	original	20	2.8	14	4.0	11
5	double	10	0.0	1	3.0	2
6	double	27	7.2	13	0.4	2
7	double	6	0.0	0	0.4	1
8	double	16	1.2	3	2.2	2
12	half	47	3.6	24	3.6	1
15	half	30	2.6	12	6.0	2
19	original	46	2.6	16	2.4	1
20	original	24	2.4	9	1.2	0
21	double	27	3.0	3	0.6	1
22	double	30	9.0	12	1.8	6
23	double	28	10.4	17	0.6	2
	total	340	46.0	131	34.6	38
	total/340		0.14	0.38	0.10	0.11

Table 6.21: Liberal enumeration of the omission and commission errors in the automatic recognition results, accompanied by the enumeration of omission and commission errors in the manual recognition results (average values over 5 different interpreters) and the display pixel dimension used in the manual interpretation of each sample plot (extracted from Cloney and Leckie (1995)).

plot	flight-line	abs. col. offset	sample trees	omission errors	commssion errors
1	Head	378	15	6	6
2	Head	414	14	4	4
4	Head	446	20	18	15
5	Head	37	10	1	2
6	Head	80	27	13	2
7	Head	254	6	1	2
8	Head	165	16	4	3
12	Head	96	47	24	1
15	Head	12	30	13	3
19	Stur	21	46	16	1
20	Stur	17	24	9	0
21	Stur	122	27	3	1
22	Stur	131	30	13	7
23	Stur	21	28	17	2
		total	340	142	49
		total/340		0.42	0.14

Table 6.22: Conservative enumeration of omission and commission errors in the automatic recognition results. The sample flightline that each sample plot belongs to and the absolute column offset from the optical axis image point of each sample plot centroid are also indicated.

relatively good.

The 3 unrecognized crowns in sample plot 21 (Figure 6.31) all have relatively dark and low-contrast images (the images of these crowns are labelled in the reference display of Figure 6.31). An omission error was counted for each of these crowns in both assessments. The heights of the unrecognized sample trees ranged from 11.4 to 11.6 m and they are all co-dominant, while the average height of the other sample trees in the plot was 17.7 m. This suggests that surrounding tree crowns were intercepting direct solar irradiance that otherwise would have reached the unrecognized crowns. The unrecognized crowns are larger than they appear to be in the image. For example, the unrecognized black spruce crown (labelled *bs* in the reference display) had a diameter of 2.4 m, which is close to the average diameter for the recognized crowns (2.6 m), but its image has an exceptionally small reference region consisting of only four pixels.

The red pine whose crown image is labelled *rp'* on the reference display of Figure 6.32 was counted as unrecognized in the conservative assessment for plot 7, but was counted as

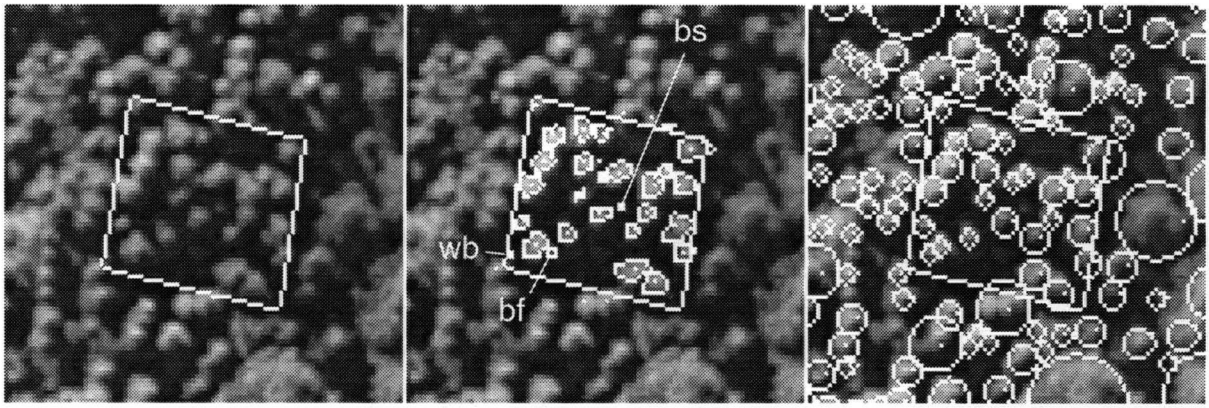


Figure 6.31: Plot boundary (left), reference (middle), and results (right) displays for sample plot 21. The images of the unrecognized sample tree crowns are labelled according to species in the reference display (*bf*: balsam fir, *bs*: black spruce, *wb*: white birch).

recognized in the liberal assessment. The two model instance boundaries that are labelled *err* in the results display each produced a commission error in the conservative assessment, but in the liberal assessment they collectively produced only 1 commission error.

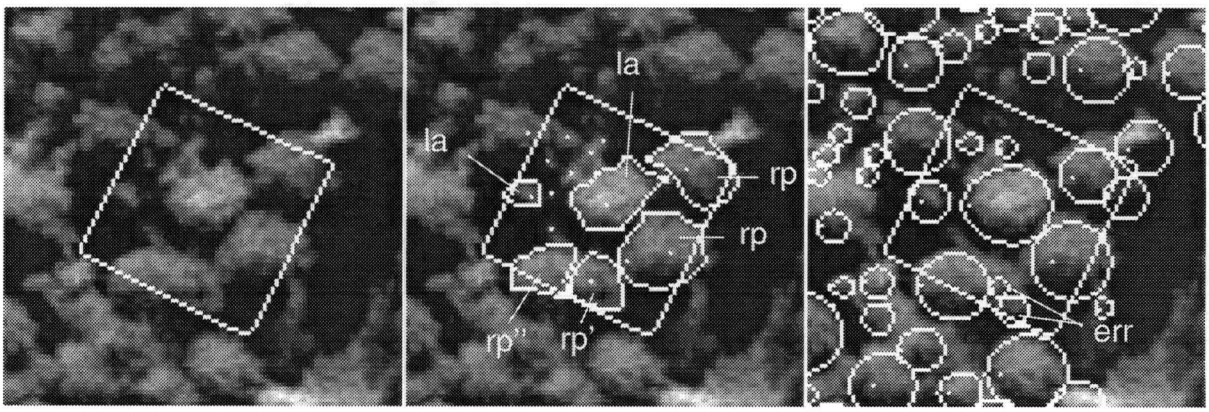


Figure 6.32: Plot boundary (left), reference (middle), and results (right) displays for sample plot 7. The sample tree crown images are labelled with their species in the reference display (la: largetooth aspen, rp: red pine). Commission errors in the conservative assessment are labelled *err* in the results display.

The image column offset for the centroid of sample plot 7 is moderate (−254 pixels), and the resulting height displacement appears to have caused the *rp'* crown to partially occlude the adjacent crown whose image is labelled *rp"*, thereby blocking the view of any between-crown

gap and explaining why the division between the two crown images is unclear (the *rp'* and *rp"* trees were respectively 27.4 and 28.2 m high, so neither tree was overtopping the other).

The *rp'* and *rp"* crowns had diameters of 6.7 m and 5.8 m, respectively, and the model instance that matches the *rp"* crown image has a horizontal radius parameter of 3.18 m, or an associated horizontal diameter value of 6.36 m. Therefore, recognition of the rp' crown should involve a model instance with approximately the same image extent as the one that matches the *rp"* crown image. However, the extents of the two positioned model instances would overlap considerably, and this is discouraged in the conflict resolution procedure (Section 5.5). Instead, two relatively small model instances were selected to match irregularities within the *rp'* crown image (these irregularities suggest a foliage gap or foliage clumping). The combined extent of these model instances covers a large part of the *rp'* crown image while producing little overlap of interpreted crown image regions.

Figure 6.33 shows the effect of changing the w_1-value from 10 (the value used to generate the results shown in Figure 6.32) to 5. The greater tolerance for overlap of interpreted crown extents (due to the smaller w_1-value) in the new results allows approximately the same amount of the combined image extent of the *rp'* and *rp"* crowns to be accounted for by two interpreted crown extents as is accounted for by three interpreted crown extents in the original results.

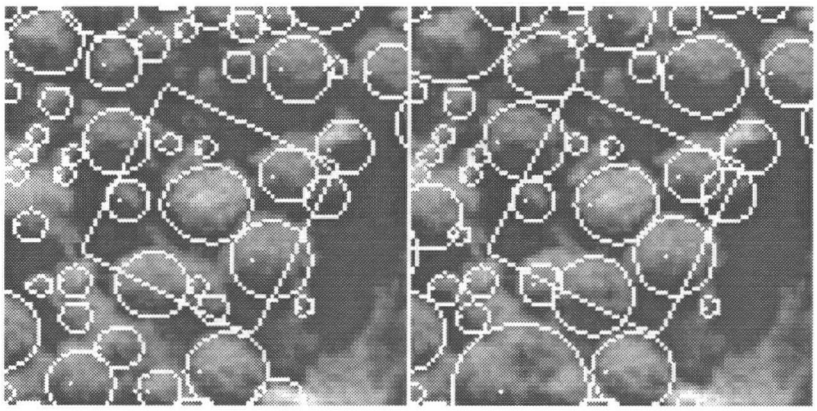


Figure 6.33: Sample plot 7 results computed with different overlap penalty weights. The display of the original results from Figure 6.32 is reproduced on the left. These results were computed with a w_1-value of 10 (*cf.* Section 6.6). The display on the right shows the results computed with the same parameters as the original results, except for an w_1-value of 5.

The automatic recognition results for sample plots 4, 12, and 22, which are discussed next, are relatively poor.

The automatic recognition results for sample plot 4 (Figure 6.34) have considerably greater numbers of both omission errors and commission errors, in either assessment, than the manual recognition results. The model instance boundaries that coincide with the 2 sample trees that are counted as recognized in the conservative assessment are labelled *rec* in the results display of Figure 6.34. The 6 sample trees that were counted as recognized in the liberal assessment are labelled *rec* in the reference display of Figure 6.35.

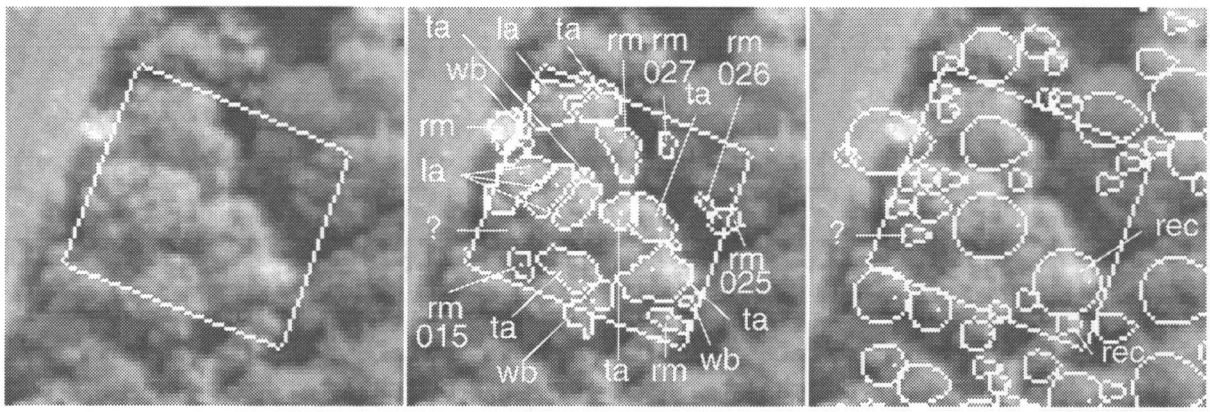


Figure 6.34: Plot boundary (left), reference (middle), and results (right) displays for sample plot 4. The matches labelled *rec* in the results display indicate the only instances of successful tree crown recognition in the conservative assessment. The sample tree crown images are labelled with their species in the reference display (*la*: largetooth aspen, *ta*: trembling aspen, *rm*: red maple, *wb*: white birch). The labels for the relatively dark and low-contrast tree crown images that are discussed in the text include the sample tree identifiers 015, 025, 026, and 027.

The automatic recognition procedure failed to establish model instance matches for most of the sample tree crown images in sample plot 4 because their spatial extents are too irregular or, in some cases, they are too dark and low-contrast. This resulted in a large number of omission errors in both assessments, relative to the number of manual recognition omission errors.

The irregular crown image extents are explained by the tendency for large hardwood trees under competition to develop irregular crown form, and by the partial occlusion of tree crowns by other tree crowns. All 20 of the sample trees were hardwoods, many of them were large

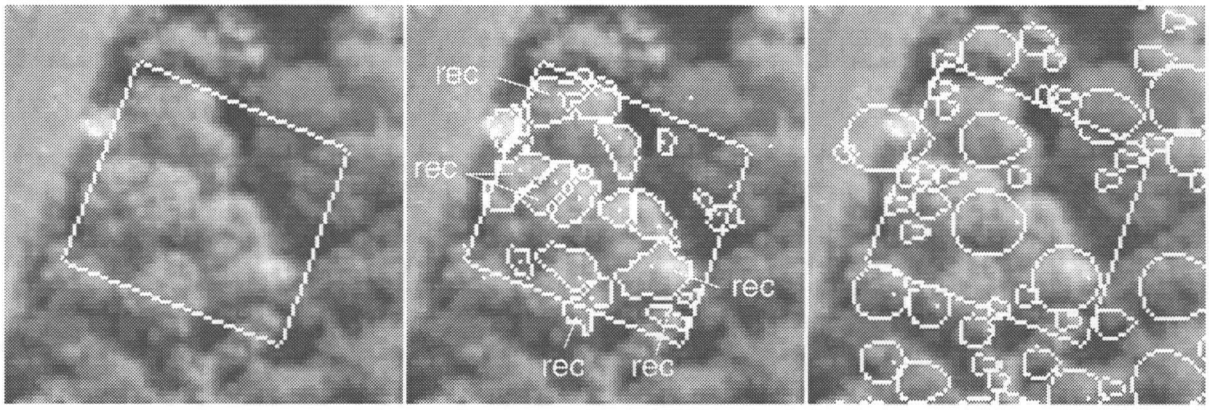


Figure 6.35: Recognized sample trees in sample plot 4 in the liberal assessment are labelled *rec* in the reference display.

(11 of them had a crown diameter greater than or equal to 5 m), they were all subjected to either medium crowding (14 trees) or high crowding (6 trees), and the offset of the sample plot centroid from the optical axis image point is large (446 pixels).

Four of the unrecognized sample tree crowns have images that are generally darker and lower in contrast than the positive training example images (these sample crown images are indicated in the reference display of Figure 6.34 with labels that include sample tree identifiers). For each of the images labelled 025, 026, and 027 one or more reasonable candidate model instance matches were initially established by the matching procedure but these were later eliminated by the correct recognition probability filter. The image labelled 015 was not even initially assigned a match. The heights of the corresponding sample trees ranged from 14.8 m to 17.9 m and they are all co-dominant, while the mean height for the other sample trees in the plot was 19.2 m and 44% of these were dominant. This suggests that the primary cause of the dark and low-contrast quality of their crown images was their relatively low height and, consequently, the interception by surrounding crowns of direct solar irradiance that otherwise would have reached them.

The matches that resulted in instances of recognition success in only the liberal assessment are matches between small model instances and within-crown image features of relatively large sample crowns. Such matches also produced most of the commission errors.

Note that the feature labelled *?* in the reference display of Figure 6.34 does not coincide

with a sample or understory tree according to the reference data, yet the appearance of this feature is at least as crown-like as the four sample tree crown images discussed above. This feature has been automatically recognized as a tree crown, which resulted in a commission error in both assessments (the coincident model instance boundary is labelled *?* in the results display).

Under both assessments, the automatic recognition results for sample plot 12 (Figure 6.36) have a considerably greater number of omission errors than the manual recognition results, but fewer commission errors. The white spruce crown images that are labelled *ws* in the reference display of Figure 6.36 were collectively recognized as part of a single relatively large crown. This produced 5 omission errors and 1 commission error in both assessments (the model instance boundary associated with the commission error is labelled *err* in the results display). These errors are explained by two factors. First the labelled crown images collectively create a pattern that resembles the image of the directly irradiated portion of a single large crown. Second, a dark image region is juxtaposed with the cluster of labelled crown images approximately along the image projection of the direct solar irradiance direction. In this context, the dark image region resembles the image of the non-directly irradiated portion of a large crown.

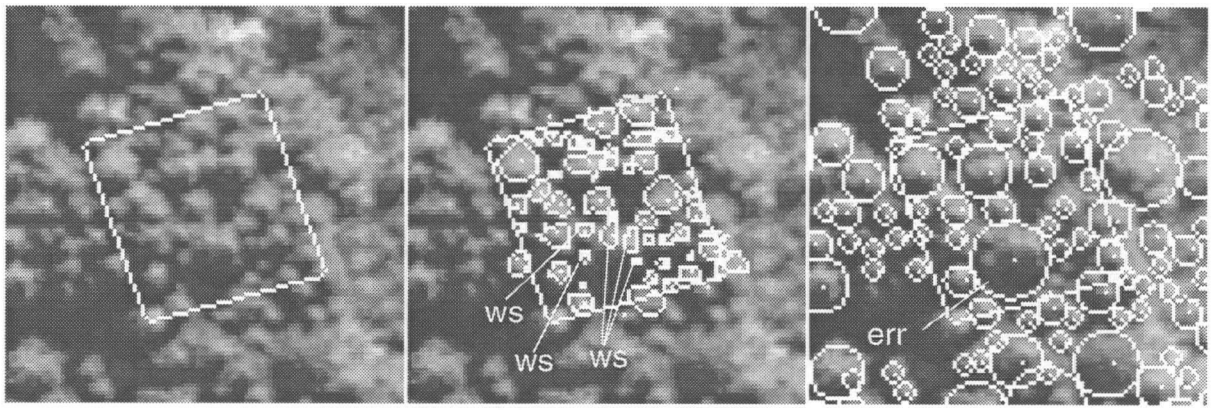


Figure 6.36: Plot boundary (left), reference (middle), and results (right) displays for sample plot 12. The images of the unrecognized white spruce crowns that are discussed in the text are labelled *ws* in the reference display and the associated commission error is labelled *err* in the results display.

The sample plot 12 results suggest that the probability of automatically recognizing a crown

with a sample or understory tree according to the reference data, yet the appearance of this feature is at least as crown-like as the four sample tree crown images discussed above. This feature has been automatically recognized as a tree crown, which resulted in a commission error in both assessments (the coincident model instance boundary is labelled *?* in the results display).

Under both assessments, the automatic recognition results for sample plot 12 (Figure 6.36) have a considerably greater number of omission errors than the manual recognition results, but fewer commission errors. The white spruce crown images that are labelled *ws* in the reference display of Figure 6.36 were collectively recognized as part of a single relatively large crown. This produced 5 omission errors and 1 commission error in both assessments (the model instance boundary associated with the commission error is labelled *err* in the results display). These errors are explained by two factors. First the labelled crown images collectively create a pattern that resembles the image of the directly irradiated portion of a single large crown. Second, a dark image region is juxtaposed with the cluster of labelled crown images approximately along the image projection of the direct solar irradiance direction. In this context, the dark image region resembles the image of the non-directly irradiated portion of a large crown.

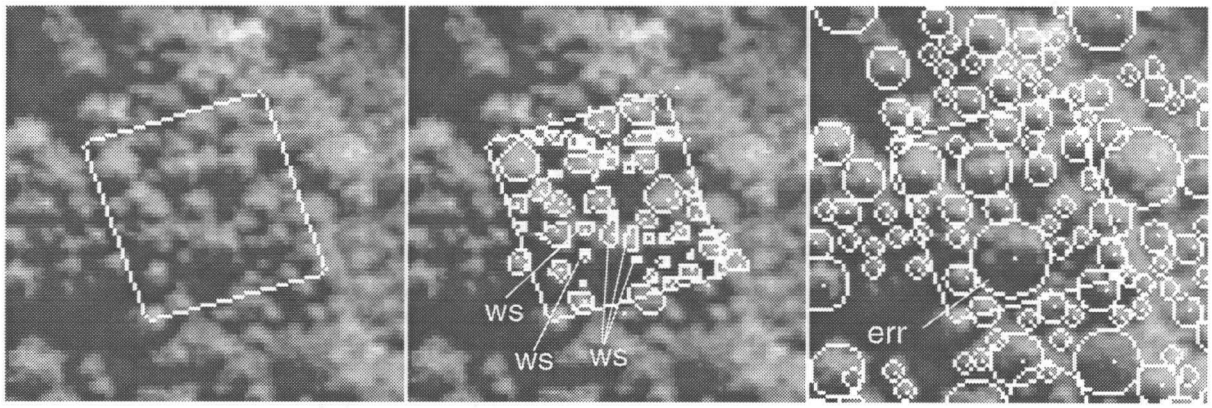


Figure 6.36: Plot boundary (left), reference (middle), and results (right) displays for sample plot 12. The images of the unrecognized white spruce crowns that are discussed in the text are labelled *ws* in the reference display and the associated commission error is labelled *err* in the results display.

The sample plot 12 results suggest that the probability of automatically recognizing a crown

is related to the size of the crown and the level of crowding that the tree was subjected to. All of the sample trees in sample plot 12 were subjected to either medium or high crowding, but 62% of the sample trees with unrecognized crowns were subjected to high crowding compared to 22% of the sample trees with recognized crowns. Also, the recognized and unrecognized crowns had an average diameter of 3.7 m and 2.5 m respectively.

The automatic recognition results for sample plot 22 (Figure 6.37) include the recognition of a single large white birch crown as multiple relatively small crowns. This example produced 5 commission errors and 1 omission error in the conservative assessment, and 4 commission errors and no omission errors in the liberal assessment (the white birch crown is labelled *wb* in the reference display of Figure 6.37, and the associated commission errors in the conservative assessment are labelled *err1* in the results display). In large- and medium-scale aerial photographs, white birch crowns tend to have a clumped appearance. This is due to the division of the main stem into several large ascending branches, each bearing its own clump of foliage (Sayn-Wittgenstein 1978, p. 69). The labelled white birch crown image is consistent with this description, so each of the associated commission errors probably coincides with a single within-crown foliage clump.

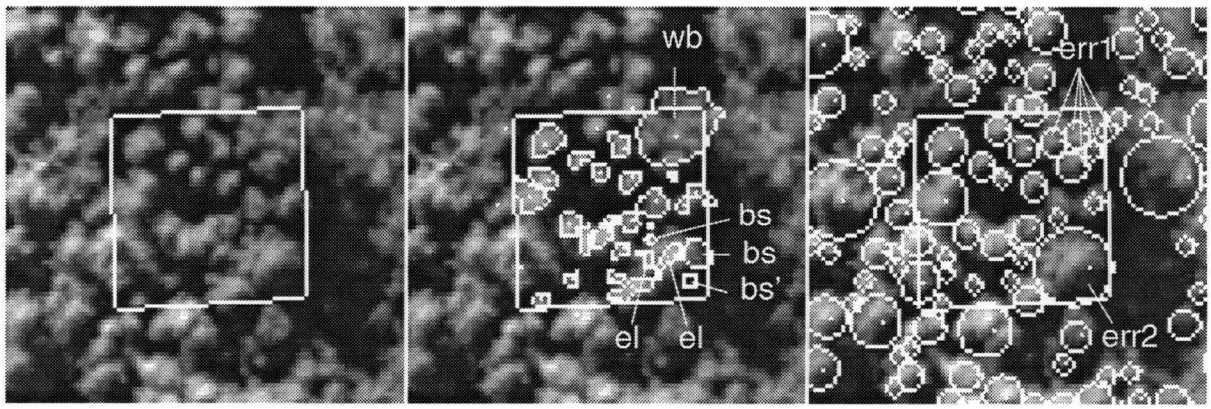


Figure 6.37: Plot boundary (left), reference (middle), and results (right) displays for sample plot 22. The unrecognized white birch crown image that is discussed in the text is labelled wb in the reference display, and the associated commission errors in the conservative assessment are labelled *err1* in the results display. The unrecognized black spruce and eastern larch crown images that are discussed in the text are respectively labelled *bs* and *el* in the reference display, and the associated commission error is labelled *err2* in the results display.

Figure 6.37 reveals a second example of the recognition of a cluster of crown images as part of a single relatively large crown. This example produced 5 omission errors and 1 commission error in both assessments. The model instance boundary associated with the commission error is labelled *err2* in the results display. Three of the crown images are of black spruce crowns (labelled *bs* in the reference display) and two are of eastern larch crowns (labelled *el* in the reference display). As is case with the example in sample plot 12 (Figure 6.36), a dark image region is juxtaposed with a bright image region that coincides with multiple poorly separated crown images, and both image regions are juxtaposed approximately along the image projection of the direct solar irradiance direction (the direct solar irradiance direction projects to different axes in the images of sample plots 12 and 22 because these images are from different flight lines).

The tree that corresponds to the dark black spruce crown image labelled *bs'* in the reference display of Figure 6.37 was 18.0 m high, while the of trees corresponding to the bright eastern larch crown images (labelled *el*) were 20.8 m and 21.8 m high. These eastern larch trees were positioned with respect to the *bs'* crown such that they probably intercepted direct solar irradiance that otherwise would have reached the *bs'* crown. This would explain the relatively dark and low-contrast appearance of the *bs'* crown image. Despite the fact that it was shorter than the eastern larch trees, the *bs'* tree was classified as dominant, possibly because it was 4 m higher than the immediately adjacent tree whose crown image is labelled *bs*.

In some cases, automatically interpreted crown regions that were computed for different processing windows overlap each other excessively. The results for sample plot 1 contain examples of this. The boundaries of the processing windows for the image of this plot are overlaid in white onto the results display of Figure 6.38. An interpreted crown region is associated with the processing window that contains the apex location of the matching model instance. The interpreted crown regions labelled *a*, *b*, and *c* in the results display respectively overlap the crown regions labelled *a'*, *b'*, and *c'* The *a*-*a'* and the *c*-*c'* examples each produced one commission error in both assessments (two of the *a* extents are outside of the sample plot and

were therefore not counted as commission errors).

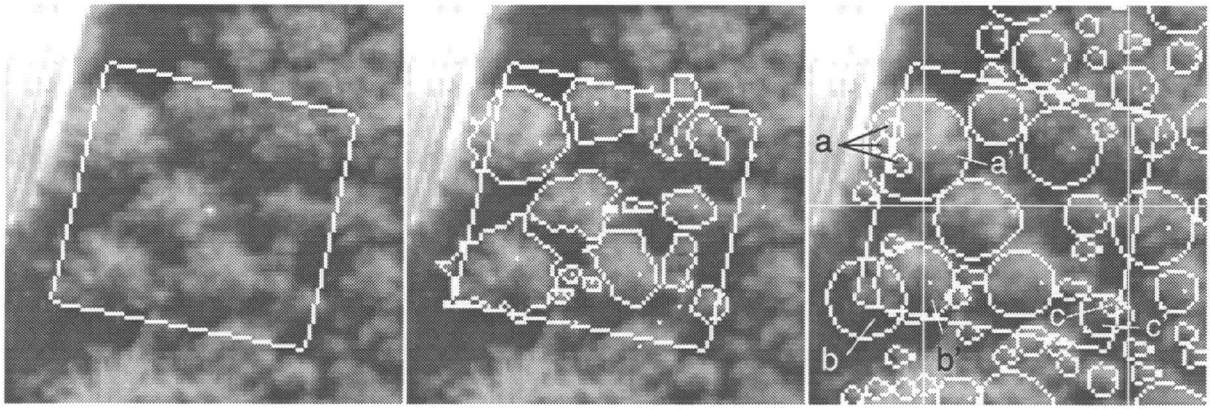


Figure 6.38: Plot boundary (left), reference (middle), and results (right) displays for sample plot 1. The processing window boundaries are overlaid onto the results display. Three examples of excessive overlap between automatically interpreted crown extents for different processing windows are labelled in the results display (*a vs. a'*, *b vs. b'*, and *c vs. c'*).

The processing step that integrates the results for different processing windows eliminates an interpreted crown region if its centroid is within a second interpreted crown extent associated with a greater correct recognition probability value (*cf.* Section 5.5). The *a*-regions were not eliminated because their match probability values (0.825, 0.878, and 0.840), are greater than that of the *a'*-region (0.668). The *c*-region is a similar case: its match probability value (0.760) is greater than that of the *c'*-region (0.691). The *b*-region has a lower match probability than the *b'*-region (0.619 *vs.* 0.812), but it was not eliminated because its centroid lies outside of the *b'*-region.

In the preceding discussions, examples were presented in which automatic crown recognition omission errors could be explained by their relationship to column offset, crowding, crown diameter, relative tree height or dominance, and, in the case of large crowns, species class (hardwood or softwood). In the following discussion, we investigate the significance of these factors with respect to the probability of automatic crown recognition over the test population.

Pairs of opposing properties that are related to the factors that we wish to investigate were defined and associated with subsets of the set of 340 sample trees. Each subset contains

exactly those sample trees with the associated property. The two subsets corresponding to any property pair are disjoint, but their union does not necessarily contain all of the sample trees. The proportion of each subset that consists of sample trees with automatically recognized crowns (in the conservative assessment) was calculated. The properties, subset sizes, and recognized proportions are listed in Table 6.23. The crown diameter threshold (3.5 m) is the mean crown diameter over all 340 sample trees. Each row of this table represents a pair of opposing properties, and the property under the heading "property 1" is associated with the greater recognized proportion.

property 1	property 2	n_1	n_2	P_1	P_2	Z	s.l. for acc. H_1	
							0.05	0.01
abs. col. off. $\leq$ 50 and med. or high. crowding	abs. col. off. $\geq$ 250 and med. or high crowding	78	42	0.462	0.333	1.367		
abs. col. off. $\leq$ 100 and med. or high. crowding	abs. col. off. $\geq$ 250 and med. or high crowding	145	42	0.469	0.333	1.564		
abs. col. off. $<$ 250 and med. or high. crowding	abs. col. off. $\geq$ 250 and med. or high crowding	175	42	0.480	0.333	1.719	$\surd$	
low crowding	med. crowding	123	168	0.813	0.470	4.683	$\surd$	$\surd$
med. crowding	high crowding	168	49	0.470	0.388	0.645		
dominant	co-dominant	197	143	0.675	0.455	4.061	$\surd$	$\surd$
diameter $\geq$ 3.5 m	diameter $<$ 3.5 m	134	206	0.634	0.549	1.553		
soft. sp. and diameter $<$ 3.5 m	hard. sp. and diameter $<$ 3.5 m	172	34	0.552	0.529	0.246		
soft. sp. and diameter $\geq$ 3.5 m	hard. sp. and diameter $\geq$ 3.5 m	81	53	0.765	0.434	3.890	$\surd$	$\surd$

Table 6.23: Automatically recognized proportions of sample tree subsets associated with specified properties. The subsets associated with property 1 and property 2 are of size n_1 and n_2 and have recognized proportions P_1 and P_2. Z is the value of the standardized variable for $P_1 - P_2$. The hypothesis H_1 that the automatically recognized population proportion associated with property 1 is greater than that associated with property 2 can be accepted at the 0.05 and 0.01 significance levels where indicated with a $\surd$.

For each pair of properties, we define the H_0 to be the hypothesis of equal automatically recognized test population proportions and H_1 to be the hypothesis that the automatically recognized test population proportion associated with property 1 is greater than that associated with property 2. The results of a test that was performed to decide between H_0 and H_1 for each pair of properties are listed in Table 6.23. The standardized variable Z used in the test

was calculated according to

$$Z = \frac{P_1 - P_2}{\sigma_{P_1 - P_2}}$$

where

$$\sigma_{P_1 - P_2} = \sqrt{P(1-P)(\frac{1}{n_1} + \frac{1}{n_2})}.$$

The assumed uniform test population proportion P was evaluated to the weighted mean of the sample subset proportions:

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}.$$

H_1 was accepted in a one-tailed test at the 0.05 and 0.01 significance levels for $Z > 1.645$ and $Z > 2.33$ respectively. This test relies on the assumption that the sample trees with a specified property represent a random sample of the test population trees with that property, in which case $n_i P_i$ has a binomial distribution and the density function of Z is very nearly standard normal for large values of n_i ($n_i \geq 30$) (Hogg and Craig 1978, § 5.4).

The results of the significance tests provide strong evidence that H_0 can be rejected for crowding variation between low and medium levels, and for dominance variation. There is also strong evidence that H_0 can be rejected for species class variation when consideration is limited to trees with average or greater than average crown diameter, but not when consideration is limited to trees with smaller than average crown diameter.

There is little evidence that H_0 can be rejected for absolute column offset variation (even when consideration is limited to trees under medium or high crowding), for crowding variation between medium and high levels, or for crown diameter class variation (diameter greater than or equal to average *vs.* diameter less than average).

Table 6.24 shows that 69% of the automatic recognition commission errors are associated with model instances that have the 1.00 m horizontal radius value. Of these 34 commission errors, 31 coincide with parts of images of relatively large crowns, as do 5 of the 6 commission errors associated with the 1.55 m horizontal radius value. The single commission error associated with the 6.21 m horizontal radius value occurs in the plot 15 results where it coincides with a

dense cluster of irregular hardwood crowns. Only 3 of the commission errors occurred because the results integration step was ineffective (2 in sample plot 2 and 1 in sample plot 8).

model instance hor. radius (m)	1.00	1.55	2.09	2.64	3.18	3.75	4.28	4.84	5.48	6.21
number of commissions	34	6	0	2	3	3	0	0	0	1

Table 6.24: Number of conservatively assessed automatic recognition commission errors associated with each model instance horizontal radius parameter value.

6.7.2 Crown Diameter Estimation Results

The horizontal diameter value associated with the matching model instance for an automatically recognized tree crown provides an *automatic diameter estimate* for that tree crown (the horizontal diameter value associated with a model instance is simply twice the horizontal radius parameter value that was used in computing the model instance). This section presents an assessment and analysis of the accuracy of the automatic crown diameter estimates for each of the 198 tree crowns that were considered to be automatically recognized in the conservative recognition accuracy assessment.

For 197 of these tree crowns, a crown diameter estimate was also derived from the manually-delineated reference crown image region by computing the average of the overall pixel-unit dimension of the region along four different axes (along-column, along-row, and both diagonals) and multiplying the results by a GPPD of 0.364 m. This provided a set of *manual diameter estimates* for those tree crowns whose accuracy could be compared to that of the automatic diameter estimates. One of the automatically recognized tree crowns had only a reference apex location and no reference image region, and was therefore excluded from the manual classification.

The automatic and manual diameter estimates were used to assign the recognized crowns to diameter classes. Two different sets of crown diameter class intervals, were considered. The endpoints of the "full-resolution" intervals were placed halfway between the full set of ordered

horizontal diameter values for the model instance set (see Table 6.25). Except for the first and last intervals, which are open-ended, the widths of these intervals range from 1.0 to 1.4 m.

index	1	2	3	4	5	6	7	8	9	10
$d \geq$		2.6	3.6	4.7	5.8	6.9	8.0	9.1	10.3	11.7
$d <$	2.6	3.6	4.7	5.8	6.9	8.0	9.1	10.3	11.7	

Table 6.25: Full resolution horizontal crown diameter (d) intervals (metres).

The endpoints of the "half-resolution" intervals were placed halfway between every second model instance set horizontal diameter value (see Table 6.26). Except for the first and last intervals, which are open-ended, the widths of these intervals range from 2.2 to 2.3 m.

index	1	2	3	4	5
$d \geq$		3.6	5.8	8.0	10.3
$d <$	3.6	5.8	8.0	10.3	

Table 6.26: Half resolution horizontal crown diameter (d) intervals (metres).

Errors in the classifications derived from the manual and automatic diameter estimates (the *test classifications*) were assessed on the basis of classifications derived from the crown diameter measurements that were made on the ground (the *reference classifications*). The correct and erroneous components of each of the test classifications are enumerated in an *error matrix*. The error matrices for the full-resolution estimates are presented in Tables 6.27 and 6.28, and those for the half-resolution estimates are presented in Tables 6.29 and 6.30. In each error matrix, a crown assigned to class a by the test classification and assigned to class b by the reference classification is counted in the matrix cell at row a and column b.

The following quantities are involved in the accuracy assessment, several of which are included in Tables 6.27 to 6.30:

N: total number of sample trees (both recognized and unrecognized) in the 14 test plots ($N = 340$)

m: the number of classes ($m = 10$ for the full-resolution classifications and $m = 5$ for the half-resolution classifications)

N_i: number of sample trees (both recognized and unrecognized) in the 14 test plots that are assigned to class i by the reference classification

P_i: (N_i/N) an estimate of the *a priori* probability of assigning a sample tree to class i by chance

n: total number of sample trees involved in the accuracy assessment ($n = 198$ for the automatic classification and $n = 197$ for the manual classification)

n_{ii}: number of sample trees counted in row i and column i of the error matrix (*i.e.*, the number of sample trees correctly assigned to class i by the test classification)

n_{i+}: sum of entries in row i of the error matrix, which is the number of sample trees assigned to class i by the test classification

n_{+i}: sum of entries in column i of the error matrix, which is the number of sample trees assigned to class i by the reference classification

P_{+i}: (n_{+i}/n) components of the marginal distribution of the reference classification

u_i: (n_{ii}/n_{i+}) *user's accuracy* for class i (Story and Congalton 1986), which is an estimate of the probability that a tree assigned to class i by the test classification actually belongs class i

p_i: (n_{ii}/n_{+i}) *producer's accuracy* for class i (Story and Congalton 1986), which is an estimate of the probability that a tree actually in class i is assigned to class i by the test classification

x_i: (nP_i) an estimate of the number of sample trees that would be assigned to class i by chance

Inspection of the error matrices reveals a tendency for each test classification to underestimate crown diameter. This has caused the class 1 producer's accuracy to be considerably larger than the class 1 user's accuracy in each error matrix, since the crown diameter for a tree that is actually in class 1 cannot be underestimated, while underestimation of the crown diameter for a tree that is actually in one of the other classes may cause it to be assigned to class 1 by the test classification.

A tendency to underestimate crown diameter has also been noted in the manual interpretation of medium scale aerial photographs (Howard 1991, p. 299). This occurs because many thin branches that help define the maximum extent of a crown and are considered in field measurements are not resolved in the photographs, or are hidden because they penetrate into adjacent crowns. These may also be significant factors in the manual and automatic diameter classifications presented here since the spatial resolution of the experimental image data is clearly not high enough to reveal individual branches. The examination of the recognition results for sample plot 21 in Section 6.7.1 suggests that the degree to which the tree crown was shaded by other tree crowns when the image data were acquired may also be a significant factor.

	1	2	3	4	5	6	7	8	9	10	n_{i+}	u_i
1	53	37	10	1	·	·	·	·	·	·	101	0.52
2	5	16	15	3	1	·	·	·	·	·	40	0.40
3	1	4	10	6	·	1	·	·	·	·	22	0.45
4	·	1	3	7	1	1	·	·	·	·	13	0.54
5	·	·	·	1	3	·	1	·	·	·	5	0.60
6	·	·	·	1	4	3	·	·	·	·	8	0.38
7	·	·	·	1	·	1	3	·	·	·	5	0.60
8	·	·	·	·	·	·	1	·	·	·	1	0.00
9	·	·	·	·	·	·	·	1	·	·	1	0.00
10	·	·	·	·	·	·	1	·	·	·	1	0.00
n_{+i}	59	58	38	20	9	6	6	1	0	0	197	
p_i	0.90	0.28	0.26	0.35	0.33	0.50	0.50	0.00	—	—		
x_i	63.158	60.834	31.284	22.596	8.116	5.220	4.058	1.734	0.000	0.000		

Table 6.27: Error matrix for the full-resolution horizontal crown diameter classification based on manual crown image region delineation. A crown estimated to be in class a and known to be in class b is counted in the cell at row a and column b.

A quantity that is often used to measure the overall accuracy of a test classification is the *overall agreement* $P_o = \sum_{i=1}^{m} n_{ii}/n$, which is the correctly classified fraction of the total number of samples (Story and Congalton 1986). The value of P_o associated with each error matrix is listed in Table 6.31. In both the manual and automatic cases, the overall agreement value for the half-resolution classification is 1.6 times that of the full-resolution classification.

For random samples of trees, nP_o has a binomial distribution; therefore, the standard deviation of P_o can be estimated according to

$$\sigma(P_o) = \sqrt{\frac{1}{n} P_o(1 - P_o)}.$$

	1	2	3	4	5	6	7	8	9	10	n_{i+}	u_i
1	43	28	8	3	·	·	·	·	·	·	82	0.52
2	16	27	19	3	2	0	1	·	·	·	68	0.40
3	·	3	7	3	·	2	·	·	·	·	15	0.47
4	·	1	3	5	3	1	·	·	·	·	13	0.38
5	·	·	1	3	2	2	1	·	·	·	9	0.22
6	·	·	·	3	2	1	4	·	·	·	10	0.10
7	·	·	·	·	·	·	·	·	·	·	0	—
8	·	·	·	·	·	·	·	1	·	·	1	1.00
9	·	·	·	·	·	·	·	·	·	·	0	—
10	·	·	·	·	·	·	·	·	·	·	0	—
n_{+i}	59	59	38	20	9	6	6	1	0	0	198	
p_i	0.73	0.46	0.18	0.25	0.22	0.17	0.00	1.00	—	—		
x_i	63.479	61.142	31.442	22.711	8.158	5.247	4.079	1.742	0.000	0.000		

Table 6.28: Error matrix for full-resolution automatic horizontal crown diameter classification. A crown estimated to be in class a and known to be in class b is counted in the cell at row a and column b.

	1	2	3	4	5	n_{i+}	u_i
1	111	29	1	·	·	141	0.79
2	6	26	3	·	·	35	0.74
3	·	2	10	1	·	13	0.77
4	·	1	1	4	·	6	0.67
5	·	·	·	2	·	2	0.00
n_{+i}	117	58	15	7	0	197	
p_i	0.95	0.45	0.67	0.57	—		
x_i	123.992	53.879	13.317	5.792	0.000		

Table 6.29: Error matrix for the half-resolution horizontal crown diameter classification based on manual crown image region delineation. A crown estimated to be in class a and known to be in class b is counted in the cell at row a and column b.

	1	2	3	4	5	n_{i+}	u_i
1	114	33	2	1	·	150	0.76
2	4	18	6	·	·	28	0.64
3	·	7	7	5	·	19	0.37
4	·	·	·	1	·	1	1.00
5	·	·	·	·	·	0	—
n_{+i}	118	58	15	7	0	198	
p_i	0.97	0.31	0.47	0.14	—		
x_i	124.621	54.153	13.385	5.821	0.000		

Table 6.30: Error matrix for the half-resolution automatic horizontal crown diameter classification. A crown estimated to be in class a and known to be in class b is counted in the cell at row a and column b.

error matrix	P_r	P_o	$\sigma(P_o)$	P_o 99% conf. int.	T	$\sigma(T)$	T 99% conf. int.
man. full-res.	0.233	0.482	0.0356	(0.390, 0.574)	0.325	0.0464	(0.205, 0.445)
auto. full-res.	0.233	0.434	0.0352	(0.343, 0.525)	0.262	0.0459	(0.144, 0.380)
man. half-res.	0.461	0.766	0.0302	(0.689, 0.844)	0.566	0.0560	(0.421, 0.710)
auto. half-res.	0.461	0.707	0.0323	(0.624, 0.790)	0.456	0.0600	(0.304, 0.614)

Table 6.31: Error matrix assessments.

For large n, the 99% confidence interval for P_o can be estimated as $P_o \pm 2.58\,\sigma(P_o)$. The values of $\sigma(P_o)$ and the associated 99% confidence interval for each error matrix is listed in Table 6.31. The confidence intervals show that the hypothesis that the test population overall agreement value is greater than zero can be accepted at the 0.005 significance level for every error matrix.

Overall agreement does not take into account the chance agreement between a reference and test classification, and may therefore give a misleading impression of high classification accuracy. The Tau coefficient (T) is an alternative measure of classification accuracy that does consider chance agreement. It is calculated according to

$$T = \frac{P_o - P_r}{1 - P_r}$$

where $P_r = \sum_{i=1}^{m} P_{+i} P_i$ is measure of random agreement (Klecka 1980; Ma and Redmond 1995). The Tau coefficient can be interpreted as an adjustment of overall agreement by the number of classes and the *a priori* probability of correct classification: if $P_i = 1/m$ then the upper limit of T as m increases is P_o.

Table 6.31 includes the value of T for each error matrix, and the estimates for the standard deviation of T, which are calculated according to

$$\sigma(T) = \sqrt{\frac{P_o(1 - P_o)}{n(1 - P_r)^2}}.$$

(Ma and Redmond 1995). Although the values for T are generally lower than the corresponding values for P_o, the associated confidence intervals show that the hypothesis of a test population value of zero for T can be rejected at the 0.005 significance level for every error matrix. Therefore, we can reject the hypothesis that the classification results only represent chance agreement.

In both the manual and automatic cases, the Tau coefficient for the half-resolution classification is 1.7 times that of the full-resolution classification.

Let T_1 and T_2 be Tau coefficients associated with test classifications 1 and 2, respectively, define H_0 to be the hypothesis of equal test population Tau coefficients for the two classifications, and define H_1 to be the hypothesis of a larger test population Tau coefficient for classification 1. A test was performed to decide between H_0 and H_1 for the pairs of test classifications listed in Table 6.32. The standardized variable Z that was used in the test was calculated according to

$$Z = \frac{T_1 - T_2}{\sqrt{\sigma^2(T_1) + \sigma^2(T_2)}}.$$

test classifications	Z	s.l. for acc. H_1	
		0.05	0.01
1. man. full-res. 2. auto. full-res.	0.965		
1. man. half-res. 2. auto. half-res.	1.340		
1. man. half-res. 2. man. full-res.	3.314	√	√
1. auto. half-res. 2. auto. full-res.	2.568	√	√

Table 6.32: Standardized variable values for differences of Tau coefficients for various pairs of crown diameter classifications. The hypothesis H_1 that the population Tau coefficient for classification 1 is greater than that of classification 2 can be accepted at the 0.05 and 0.01 significance levels where indicated with a √.

The values of Z for different pairs of test classifications are listed in Table 6.32. H_1 was accepted in a one-tailed test at the 0.05 and 0.01 significance levels for $Z > 1.645$ and $Z > 2.33$, respectively.

The results of the significance tests do not support the rejection of the hypothesis that manual and automatic crown diameter classifications are equally accurate over the test population; however, they do support the rejection of the hypothesis that tree crowns are no more accurately assigned to approximately 2.0 m wide diameter classes than to approximately 1.0 m wide diameter classes under either classification method.

6.7.3 Summary

The errors in automatic recognition results were mostly omission errors. Automatic recognition errors were evidently caused by non-ideal crown irradiation, irregular crown form (especially for some hardwood trees), partial occlusion of tree crowns due to height displacement, faulty interpretation of tight clusters of tree crowns as relatively large tree crowns, and faulty integration of the results for adjacent processing windows. The errors in the automatic and manual image-based crown diameter estimates were comparable. Both sets of estimates tended to be low relative to ground measurements, probably because of the intersection of tree crowns and the failure to resolve branches at the outer crown extent which are visible from the ground.

A more detailed summary of the results assessments and analyses follows:

Recognition results:

- The automatic recognition results were assessed twice, according to two different sets of rules. The liberal assessment, which de-emphasizes correct image-space localization, was intended to produce error counts that can be compared to the manual recognition results. The conservative assessment placed more emphasis on correct image-space localization.

- The numbers of commission errors in the manual and the liberally assessed automatic recognition results are close: respectively 10% and 11% of the 340 sample trees. However, the number of liberally assessed automatic recognition omission errors (38%) is 2.8 times the number of manual recognition omission errors (14%) (*cf.* Table 6.21). The numbers of conservatively assessed errors are only slightly greater than the numbers of liberally assessed errors (*cf.* Table 6.22).

- Inspection of the results for select test sample plots revealed evidence of the following:

  - some tree crowns were not recognized because direct solar irradiation that would otherwise have reached them was intercepted by neighbouring trees.

  - some recognition errors occurred because of the requirement for a large amount

of overlap in the interpreted crown boundaries in cases where height displacement caused tree crowns to partially occlude each other

  - some large hardwood trees were unrecognized because they had an irregular crown form that was caused by growing under competition, and some because they had an basic crown architecture that does not correspond well to the simplifying assumptions of the tree crown image model
  - tight clusters of tree crowns were sometimes recognized as the directly irradiated portion of a single relatively large tree crown when the cluster image was juxtaposed with a dark image region along the image projection of the direct solar irradiation direction
  - some errors were produced by faulty integration of the results for adjacent processing windows

- Most commission errors were due to the recognition of portions of tree crown images as separate relatively small tree crowns.

- Pairs of opposing properties that correspond to pairs of disjoint test population subsets were defined, and for each pair the null hypothesis that equal proportions of those subsets would be automatically recognized (H_0) was tested using the test sample tree recognition results (*cf.* Table 6.23).

- The results of the significance tests provide strong evidence that H_0 can be rejected for crowding variation between low and medium levels, and for dominance variation. There is also strong evidence that H_0 can be rejected for species class variation (softwood *vs.* hardwood) when consideration is limited to trees with average or greater than average crown diameter, but not when consideration is limited to trees with smaller than average crown diameter.

- The results of the significance test provide little or no evidence that H_0 can be rejected for absolute column offset variation (even when consideration is limited to trees under

medium or high crowding), for crowding variation between medium and high levels, or for crown diameter class variation (diameter greater than or equal to average *vs.* diameter less than average).

Crown diameter estimation results:

- Automatic and manual image-based crown diameter estimates were compared to field measurements of crown diameter on the basis of two different sets of crown diameter classes: the full-resolution set (with 1.0 to 1.4 m wide class intervals) and the half-resolution set (with 2.2 to 2.3 m wide class intervals) (see Tables 6.27, 6.28, 6.29, and 6.30 for the error matrices).

- Both the automatic and manual image-derived crown diameter estimates tended to be low for both sets of classes.

- The overall agreement of the automatic image-based crown diameter classification with the field measurement-based crown diameter classification is 0.434 for the full-resolution classes and 0.707 for the half-resolution classes.

- For each error matrix, the Tau coefficient permits the rejection of the hypothesis that the image-based crown diameter classification results only represent chance agreement, and do not permit the rejection of the hypothesis that the automatic and manual crown diameter image-based classifications are equally accurate.

Chapter 7

Conclusion

This dissertation discusses the design, implementation, and testing of a strategy for recognizing individual tree crowns in high spatial resolution aerial images of forests. This work makes a contribution to the overall effort to increase the level of automation in forest type mapping.

The strategy is based on the use of readily available scene and image formation knowledge to constrain the recognition task, which makes it unique within the problem domain. The knowledge is used to instantiate a tree crown image model, and the model instances are used to generate a set of initial recognition instances from which a training set of positive and negative exemplars is manually extracted. A recognition probability filter applies the training set to the selection of a subset of the initial recognition instances, which is then passed through a conflict resolution filter. The idea of using a pre-defined image model to support manual training is novel within the overall context of automated visual object recognition. Another novel idea that is incorporated into the strategy is the use of constraints on the spatial relationships among multiple neighbouring candidate recognition instances to resolve interpretation conflicts.

The strategy was implemented and tested with a scene of mixed uneven-aged forests in which the trees represent a wide variety of species, size, and growing conditions. The test results were assessed and analyzed in detail and compared to human performance in the same task. The scene represented a level of difficulty that is significantly above the level that has been addressed by previous attempts at automating the tree crown recognition task. Also, the detail of the results assessment and analysis is greater than that of previous related work.

This work demonstrated that the implementation was largely able to accommodate the

variation represented by the test scene. Recognition was successful in many cases where the tree crown image regions are not clearly delimited by a contrasting background region. Clear tree crown image region boundaries are relied on by previously implemented tree crown recognition strategies. However, this work also demonstrated that the recognition results that were produced by the implementation are less accurate than manually generated recognition results.

The commission error levels in the automatically and manually generated recognition results for 340 sample trees were practically the same (respectively 11% and 10%); however, the omission error level of the automatically generated results (38%) was almost three times that of the manually generated results (14%). Irregularities in tree crown form and irradiance that were caused by tight vertical and horizontal spacing of the crowns were the most significant sources of error in the automatic recognition results. Human interpreters were better able to accommodate these irregularities than the implementation was able to.

The automatic and manual crown diameter estimation results for the automatically recognized trees do not differ significantly. Crown diameter estimation was significantly better for approximately 2.2 m wide diameter class intervals than for approximately 1.2 m wide diameter class intervals (the overall agreement values for the automatically generated results for these classifications and the field-collected reference data are respectively 0.71 and 0.43). Crown diameter was usually underestimated, probably because the full crown extents are usually not visible in the experimental image data.

7.1 Further Research

The operational utility and relative merits of different automatic tree crown recognition strategies, including the one described in this dissertation, is still undefined primarily because operational standards for the accuracy of individual tree crown recognition do not exist. Since the primary application of forest type mapping is to the estimation of mechantable timber volume in forest stands, further research should place a high priority on defining the level of tree crown recognition accuracy that adequately supports this task.

The conversion of scene and image formation knowledge into tree crown model instance sets, the definition of recognition regions, and the tuning of the components of the recognition procedure require considerable user input and experimentation. Further systemization of these tasks and possibly the development of support tools would significantly increase the operational potential of the tree crown recognition strategy.

The training set that was used by the correct recognition probability filter in the main experiment consists of initial recognition instances that a human interpreter considers, with a high degree of confidence, to be either correct or incorrect. Further experiments should investigate the benefit of including in the training set initial recognition instances with less clearly visible evidence of correctness or incorrectness.

Evidence of image contrast between tree crowns and background should be investigated as a cue for crown detection that may in some cases offset the effects of irregular crown form and irradiance. Such evidence may be incorporated as a feature that is used to estimate the correct recognition probability. The feature value could be computed as a function of the image values near the boundary of the initial recognition instance extent.

The correct recognition results that are generated by the implementation provide a good basis for tree species classification. The image-space localization for a recognized tree crown includes the separation of the image regions for the directly irradiated and non-directly irradiated portions of the visible crown surface. This separation has been found to be important in experiments in automatic species identification from manually delimited tree crown image regions. It should be possible to investigate supervised tree species classification by extending the existing training set collection and nearest-neighbours interpolation procedures.

The current implementation of the interpretation conflict resolution filter seeks to maximize a function of the correct recognition probability values and spatial relationships for the intermediate recognition instances within uniform sized image windows. However, a window that is large enough to contain the image of a cluster of several large tree crowns may define an extremely large search space in parts of the image that predominately contain images of

relatively small tree crowns. This means that exact combinatorial optimization techniques such as backtracking are not generally practical for instances of the optimization problem. An implementation of the filter that is based on variable sized windows may make such techniques practical and may possibly be more efficient than the current filter. A fundamental problem that must be addressed in following this approach is the development of a method for determining the best window size for a particular part of the image.

In the main experiment, the component weights of the conflict resolution objective function were fixed over all processing regions. However, the experimental results suggest that match extent overlap should be penalized less in image regions where there is a large amount of relief displacement (and therefore partial occlusion of tree crowns by other tree crowns) than in regions where there is little or no relief displacement. The benefit of varying the component weights (in particular, the weight that controls the the match extent overlap) with image location should be investigated.

A method of measuring the degree of confidence with which a final recognition instance can be regarded as correct should be developed. Such a confidence measure would be used to flag the recognition instances that should be manually verified. The correct recognition probability value for a final recognition instance might be one component of such a confidence measure. However, the level of initial ambiguity that is associated with the final recognition instance and was resolved by the conflict resolution filter might also be considered. This would implement the notion that a greater level of initial ambiguity implies less confidence in the results. The level of initial ambiguity might be defined as a function of the correct recognition probability values and the image extents of the rejected intermediate recognition instances that conflict with final recognition instance: a rejected recognition instance with a relatively low correct recognition result or with an image extent that is relatively close to that of the final recognition instance would make a relatively small contribution to the level of initial ambiguity. The effectiveness of the confidence measure would be evaluated according to its effectiveness as a predictor of recognition errors.

Bibliography

Ahern, F. J. and D. G. Leckie (1987). Digital remote sensing for forestry: Requirements and capabilities, today and tomorrow. *Geocarto International 2*(3), 43–52.

Ahern, F. J., D. G. Leckie, and D. Werle (1993). Applications of RADARSAT SAR data in forested environments. *IEEE Transactions on Geoscience and Remote Sensing 19*(4), 330–337.

Albers, B. J., A. H. Strahler, X. Li, S. Liang, and K. C. Clarke (1990). Radiometric measurements of gap probability in conifer tree canopies. *Remote Sensing of Environment 34*(3), 179–192.

Aldred, A. H. and G. M. Bonner (1985). Applications of airborne lasers to forest surveys. Information Report PI-X-51, Petawawa National Forestry Institute, Chalk River, Ontario.

Biging, G. S., R. G. Congalton, and E. C. Murphy (1991). A comparison of photointerpretation and ground measurements of forest structure. In *Technical Papers of the 1991 ACSM-ASPRS Annual Convention*, pp. 6–15.

Brickell, J. E. (1992). Delineation of map units for vegetation management. In H. G. Lund, E. Landis, and T. Atterbury (eds.), *Proceedings of Stand Inventory Technologies '92*, Bethesda, Maryland, pp. 74–82. American Society for Photogrammetry and Remote Sensing.

Cloney, E. and D. G. Leckie (1995). Analysis of MEIS interpreted tree count. sixth draft of an unpublished report prepared at the Petawawa National Forestry Institute, P.O. Box 2000, Chalk River, Ontario, K0J-1J0.

Cloney, E. E., D. G. Leckie, and M. Gillis (1994). *The PNFI Data Set* (second ed.). P.O. Box 2000, Chalk River, Ontario, K0J-1J0: Petawawa National Forestry Institute.

Cosandier, D., T. Ivanco, and S. Mah (1992). The geocorrection and integration of the global positioning system with the compact airborne spectrographic imager. In J. K. Hornsby, D. J. King, and N. A. Prout (eds.), *Fifteenth Canadian Symposium on Remote Sensing*, Toronto, Ontario, pp. 385–390.

Draper, N. R. and H. Smith (1981). *Applied Regression Analysis* (second ed.). New York, New York: John Wiley & Sons, Inc.

Fitch, J. P. (1988). *Synthetic Aperture Radar*. New York, New York: Springer-Verlag.

Fraser, R. S., N. E. Gaut, E. C. Reifenstein, and H. Sievering (1975). Interaction mechanisms — within the atmosphere. In *Manual of Remote Sensing*, Chapter 5, pp. 181–233. Falls Church, Virginia: American Society of Photogrammetry.

Friedman, J. H., J. L. Bentley, and R. A. Finkel (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software 3*(3), 209–226.

Gardiner, M. (1965). The "superellipse": a curve that lies between the ellipse and the rectangle. *Scientific American 213*(3), 222–232.

Gausman, H. W., W. A. Allen, C. L. Wiegand, D. E. Escobar, and R. R. Rodriguez (1971). Leaf light reflectance, transmittance, absorptance, and optical and geometrical parameters for eleven plant genera with different leaf mesophyll arrangements. In *Proceedings of the Seventh International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1599–1625.

Gibson, J. R. (1994). Photogrammetric calibration of a digital electro-optical stereo imaging system. *Geomatica 48*(2), 95–109.

Gillis, M. D. and J. A. Edwards (1988). Volume compilation procedures in forest management inventories. Information Report PI-X-79, Petawawa National Forestry Institute, Chalk River, Ontario.

Gillis, M. D. and D. G. Leckie (1993). Forest inventory mapping procedures across Canada. Information Report PI-X-114, Petawawa National Forestry Institute, Chalk River, Ontario.

Gougeon, F. A. (1993). Individual tree identification from high resolution MEIS images. In D. G. Leckie and M. D. Gillis (eds.), *Proceedings of the International Forum on Airborne Multispectral Scanning for Forestry and Mapping (with Emphasis on MEIS)*, Chalk River, Ontario, pp. 117–128. Forestry Canada.

Gougeon, F. A. (1994). A valley following approach to the automatic delineation of individual tree crowns in high spatial resolution MEIS images. unpublished draft.

Gougeon, F. A. (1995). Comparison of multispectral classification schemes for tree crowns individually delineated on high spatial resolution MEIS images. *Canadian Journal of Remote Sensing 21*(1), 1–9.

Guyot, G., D. Guyon, and J. Riom (1989). Factors affecting the spectral response of forest canopies. *Geocarto International 3*, 3–18.

Hakkila, P. (1989). *Utilization of Residual Forest Biomass.* Berlin: Springer-Verlag.

Hershey, R. R. and W. A. Befort (1995). Aerial photo guide to new england forest cover types. General Technical Report NE-195, USDA Forest Service, Radnor, Pennsylvania.

Hogg, R. V. and A. T. Craig (1978). *Introduction to Mathematical Statistics* (fourth ed.). New York: Macmillan Publishing Co., Inc.

Hooper, F. C., A. P. Brunger, and C. S. Chan (1987). A clear sky model of diffuse sky radiance. *Transactions of the ASME Journal of Solar Energy Engineering 109*(1), 9–14.

Horn, H. S. (1971). *The Adaptive Geometry of Trees.* Princeton, New Jersey: Princeton University Press.

Horvath, R., J. G. Braithwaite, and F. C. Polcyn (1970). Effects of atmospheric path on airborne multispectral sensors. *Remote Sensing of Environment 1*, 203–215.

Hosie, R. C. (1969). *Native Trees of Canada* (seventh ed.). Canadian Forestry Service, Department of the Environment.

Howard, J. A. (1991). *Remote Sensing of Forest Resources.* London, England: Chapman & Hall.

Hughes, J. S., D. L. Evans, and P. Y. Burns (1986). Identification of two southern pine species in high-resolution aerial MSS data. *Photogrammetric Engineering and Remote Sensing 52*(8), 1175–1180.

Iqbal, M. (1983). *An Introduction to Solar Radiation.* Don Mills, Ontario: Academic Press Canada.

Jack, S. B. and J. N. Long (1992). Forest production and the organization of foliage within crowns and canopies. *Forest Ecology and Management 49*(3–4), 233–245.

Jacobs, D. M., D. L. Evans, and J. C. Ritchie (1993). Laser profiler and aerial video data for forest assessments. In *ASPRS Technical Papers in Remote Sensing: Proceedings of the 1993 ACSM/ASPRS Annual Convention and Exposition*, New Orleans, LA, pp. 135–142.

Johnson, D. S. (1974). Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences 9*(3), 256–278.

Klecka, W. R. (1980). *Discriminant Analysis.* Beverly Hills, California: SAGE Publication, Inc.

Koppel, A. and T. Oja (1984). Regime of diffuse solar radiation in an individual Norway spruce (*Picea abies* (L.) Karst.) crown. *Photosynthetica 18*(4), 529–535.

Kuuluvainen, T. and T. Pukkala (1989). Simulation of within-tree and between-tree shading of direct radiation in a forest canopy: Effect of crown shape and sun elevation. *Ecological Modelling 49*(1–2), 89–100.

Leckie, D. G. (1990a). Advances in remote sensing technologies for forest surveys and management. *Canadian Journal of Forest Research 20*(4), 464–483.

Leckie, D. G. (1990b). Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination. *Photogrammetric Engineering and Remote Sensing 56*(9), 1237–1246.

Leckie, D. G. (1993). Application of airborne multispectral scanning to forest inventory mapping. In D. G. Leckie and M. D. Gillis (eds.), *Proceedings of the International Forum on Airborne Multispectral Scanning for Forestry and Mapping (with Emphasis on MEIS)*, Chalk River, Ontario, pp. 86–93. Forestry Canada.

Leckie, D. G. (1994). Possible airborne sensor, processing and interpretation systems for major forestry applications. In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition*, Strasbourg, France, pp. II:159–169.

Li, X. and A. H. Strahler (1985). Geometric-optical modeling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing 23*(5), 705–720.

Lowe, D. G. (1995). Similarity metric for a variable-kernel classifier. *Neural Computation 7*, 72–85.

Ma, Z. and R. L. Redmond (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing 61*(4), 435–439.

McColl, W. D., R. A. Neville, and S. M. Till (1983). Multidetector electro-optical imaging scanner, MIES II. In *Proceedings of the Eighth Canadian Symposium on Remote Sensing*, pp. 71–77.

McDowell, D. Q. (1974). Spectral distribution of skylight energy for two haze conditions. *Photogrammetric Engineering 40*(5), 569–571.

Minnaert, M. (1941). The reciprocity principle in lunar photometry. *Astrophysics Journal 93*, 403–410.

Murgu, D. (1996). Individual tree detection and localization in aerial imagery. Master's thesis, Department of Computer Science, University of British Columbia.

O'Brien, R. and D. C. Chojnacky (1992). The potential for describing structural elements of bird habitat with forest inventory data. In H. G. Lund, E. Landis, and T. Atterbury (eds.), *Proceedings of Stand Inventory Technologies '92*, Bethesda, Maryland, pp. 204–212. American Society for Photogrammetry and Remote Sensing.

Oker-Blom, P. and S. Kellomäki (1982). Theoretical computations on the role of crown shape in the absorption of light by forest trees. *Mathematical Biosciences 59*(2), 291–311.

Oliver, R. E., J. K. Berry, and J. A. Smith (1975). A portable instrument for measuring apparent directional reflectance. *Optical Engineering 14*(3), 244–247.

Otten, R. H. J. M. and L. P. P. P. van Ginneken (1989). *The Annealing Algorithm.* Boston, Massachussets: Kluwer Academic Publishers.

Pinz, A. J. (1991). A computer vision system for the recognition of trees in aerial photographs. In J. C. Tilton (ed.), *Multisource Data Integration in Remote Sensing*, pp. 111–124. Greenbelt, Maryland: National Aeronautics and Space Administration.

Pratt, W. K. (1991). *Digital Image Processing* (second ed.). New York, New York: John Wiley & Sons, Inc.

Rowe, J. S. (1972). *Forest Regions of Canada.* Ottawa, Ontario: Department of the Environment, Canadian Forestry Service. Publication No. 1300.

Sayn-Wittgenstein, L. (1978). Recognition of tree species on aerial photographs. Information Report FMR-X-118, Forest Management Institute, Ottawa, Ontario.

Schreuder, H. T., T. G. Gregoire, and G. B. B. Wood (1993). *Sampling Methods for Multiresource Forest Inventory.* New York, New York: John Wiley & Sons, Inc.

Schroer, G. L. (1992). Methods for sampling wildlife habitats. In H. G. Lund, E. Landis, and T. Atterbury (eds.), *Proceedings of Stand Inventory Technologies '92*, Bethesda, Maryland, pp. 204–212. American Society for Photogrammetry and Remote Sensing.

Siala, F. M. F., M. A. Rosen, and F. C. Hooper (1990). Models for the directional distribution of the diffuse sky radiance. *Transactions of the ASME Journal of Solar Energy Engineering 112*(2), 102–112.

Silva, L. F. (1978). Radiation and instrumentation in remote sensing. In P. H. Swain and S. M. Davis (eds.), *Remote Sensing: The Quantitative Approach*, Chapter 2, pp. 21–135. New York, New York: McGraw-Hill Book Company.

Sproull, R. F. (1991). Refinements to nearest-neighbor searching in k-dimensional trees. *Algorithmica 6*(4), 579–589.

Story, M. and R. G. Congalton (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing 52*(3), 397–399.

Thomasson, J. A., C. W. Bennet, B. D. Jackson, and M. P. Mailander (1994). Differentiating bottomland tree species with multispectral videography. *Photogrammetric Engineering and Remote Sensing 60*(1), 55–59.

Turk, M. and A. Pentland (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3*(1), 71–86.

Wang, Y. P., P. G. Jarvis, and M. L. Benson (1990). Two-dimensional needle-area density distribution within the crowns of *pinus radiata. Forest Ecology and Management 32*(2–4), 217–237.

Warbington, R. and J. Levitan (1992). How to estimate canopy cover using maximum crown width/DBH relationships. In H. G. Lund, E. Landis, and T. Atterbury (eds.), *Proceedings of Stand Inventory Technologies '92*, Bethesda, Maryland, pp. 319–328. American Society for Photogrammetry and Remote Sensing.

Williams, D. L. (1991). A comparison of spectral reflectance properties at the needle, branch, and canopy level for selected conifer species. *Remote Sensing of Environment 35*(2 & 3), 79–93.

Woodham, R. J. and M. H. Gray (1987). An analytic method for radiometric correction of satellite multispectral scanner data. *IEEE Transactions on Geoscience and Remote Sensing 25*(3), 258–271.

Young, A. T. and S. A. Collins (1971). Photometric properties of the Mariner cameras and selected regions on Mars. *Journal of Geophysical Research 76*(2), 432–437.

Appendix A

Sample Plot Display Triplets

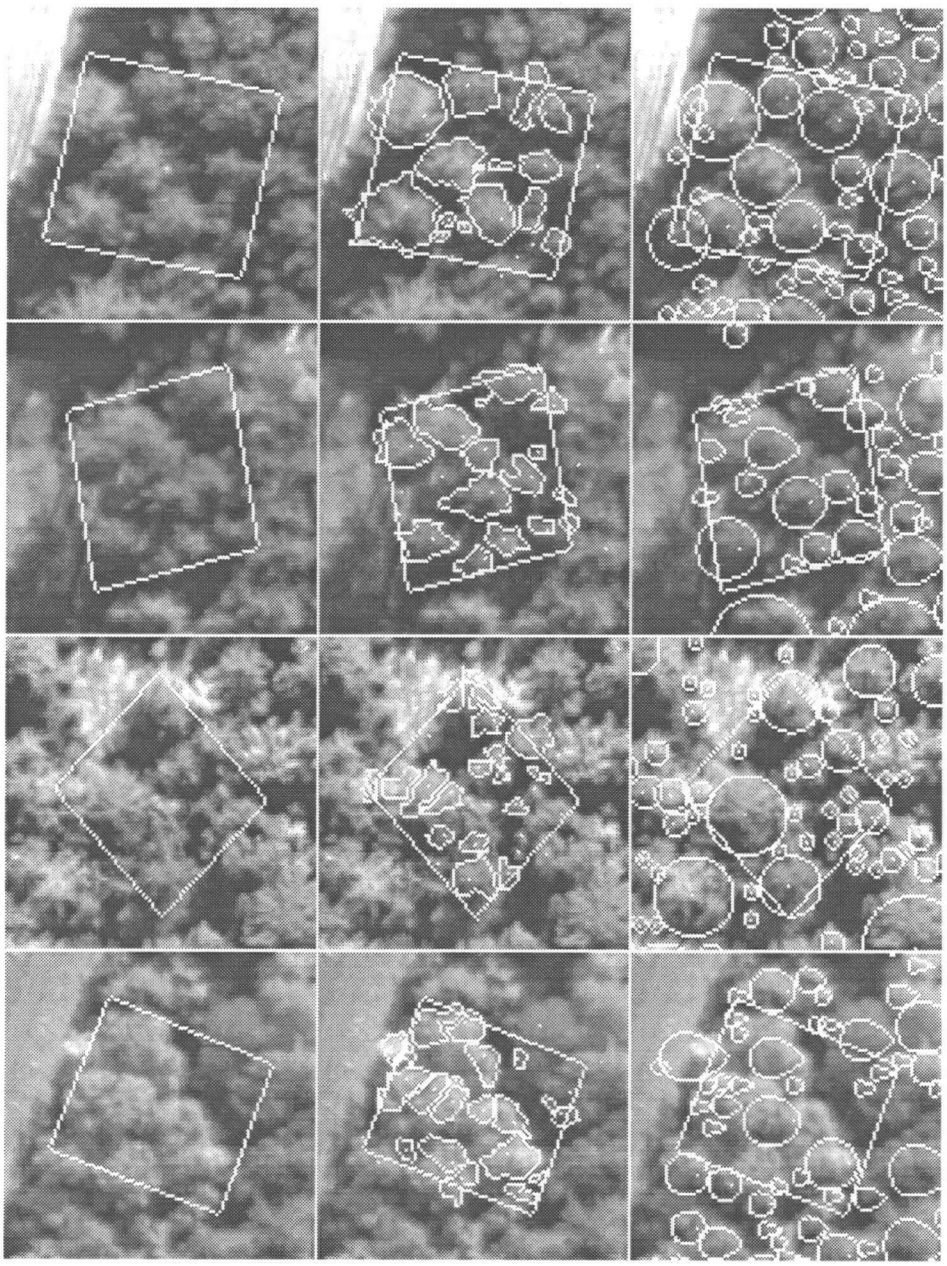


Figure A.39: Plot boundary, reference and results displays for sample plots 1 (top) to 4 (bottom).

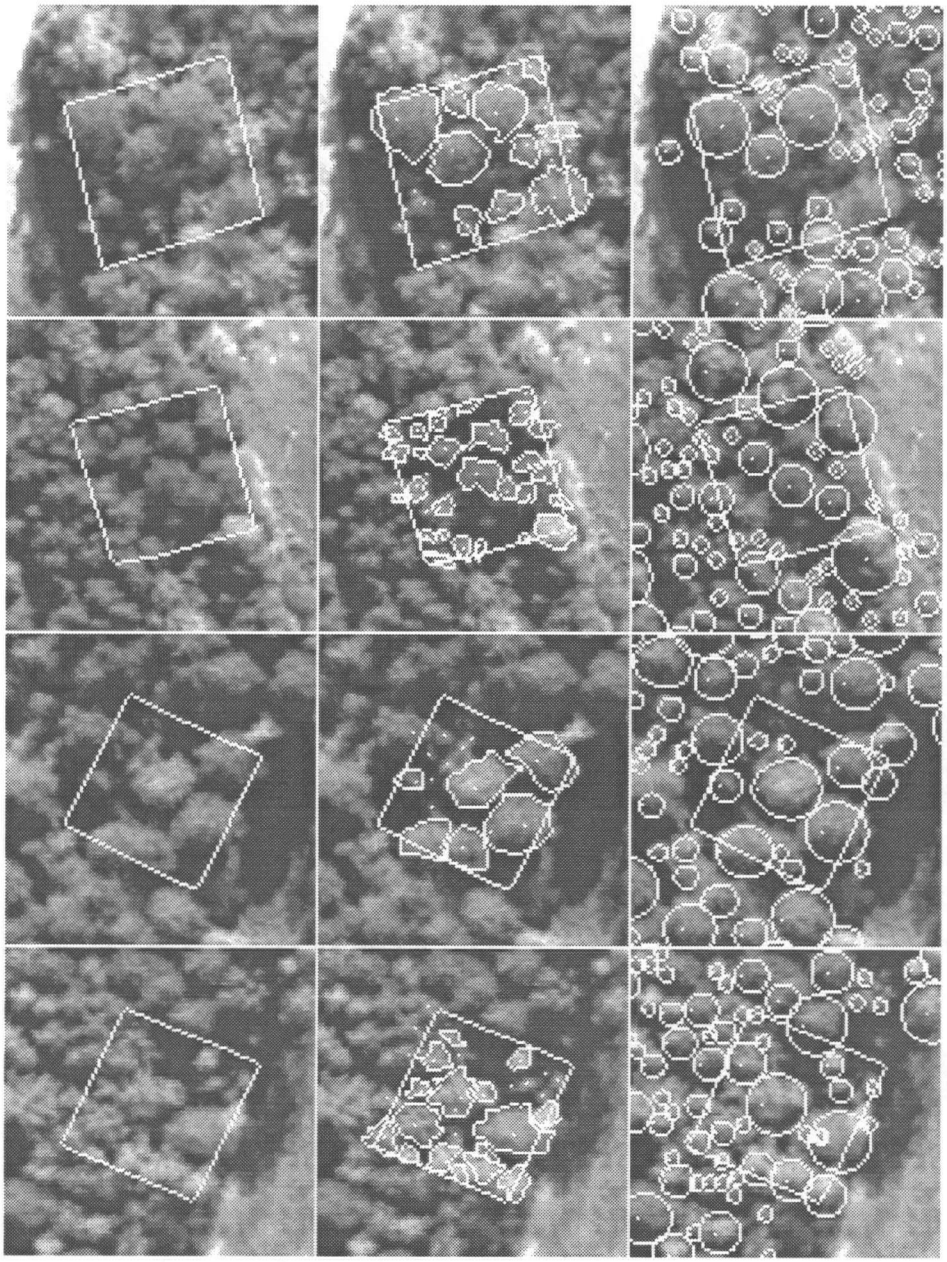


Figure A.40: Plot boundary, reference and results displays for sample plots 5 (top) to 8 (bottom).

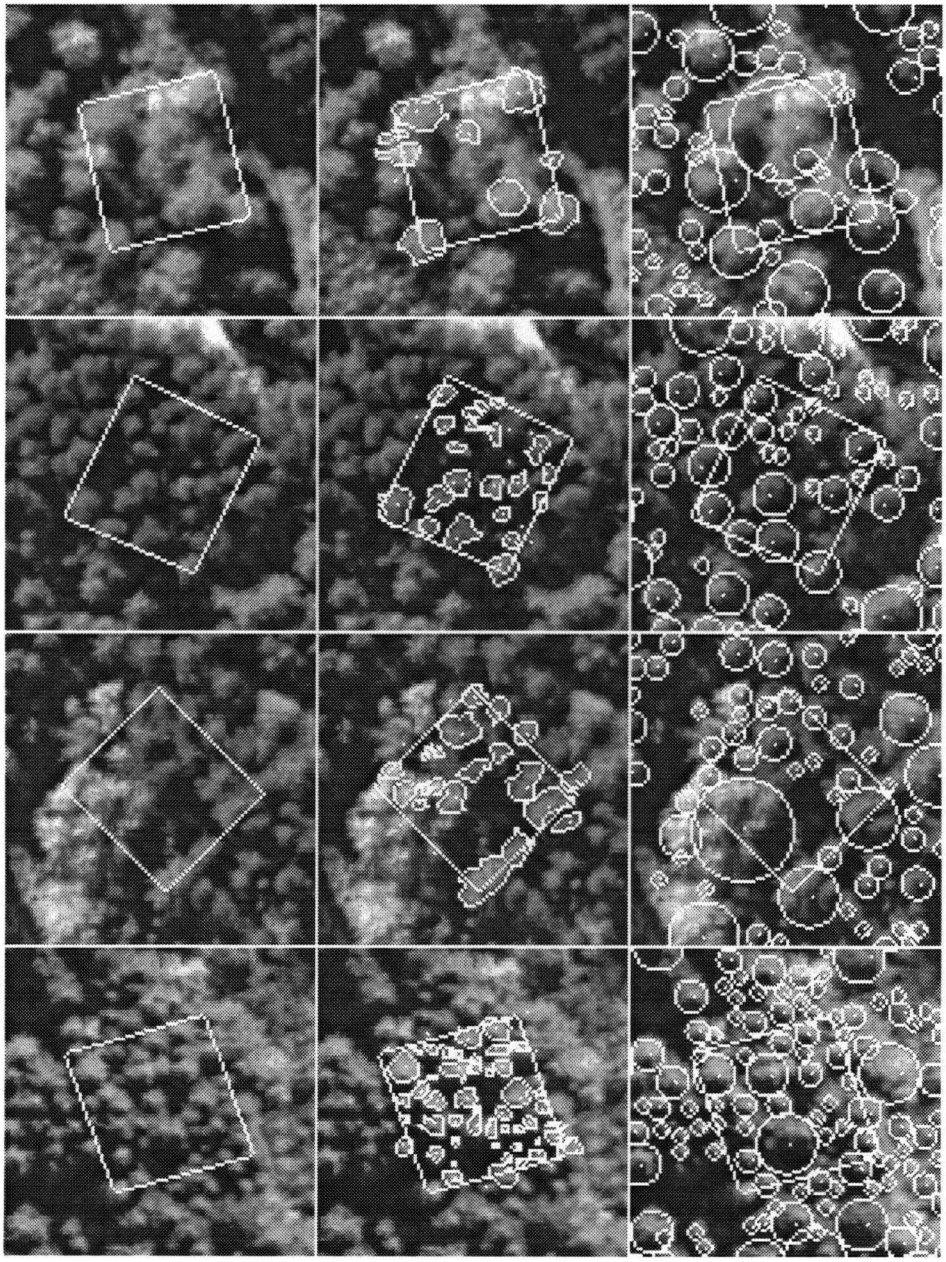


Figure A.41: Plot boundary, reference and results displays subimages for sample plots 9 (top) to 12 (bottom).

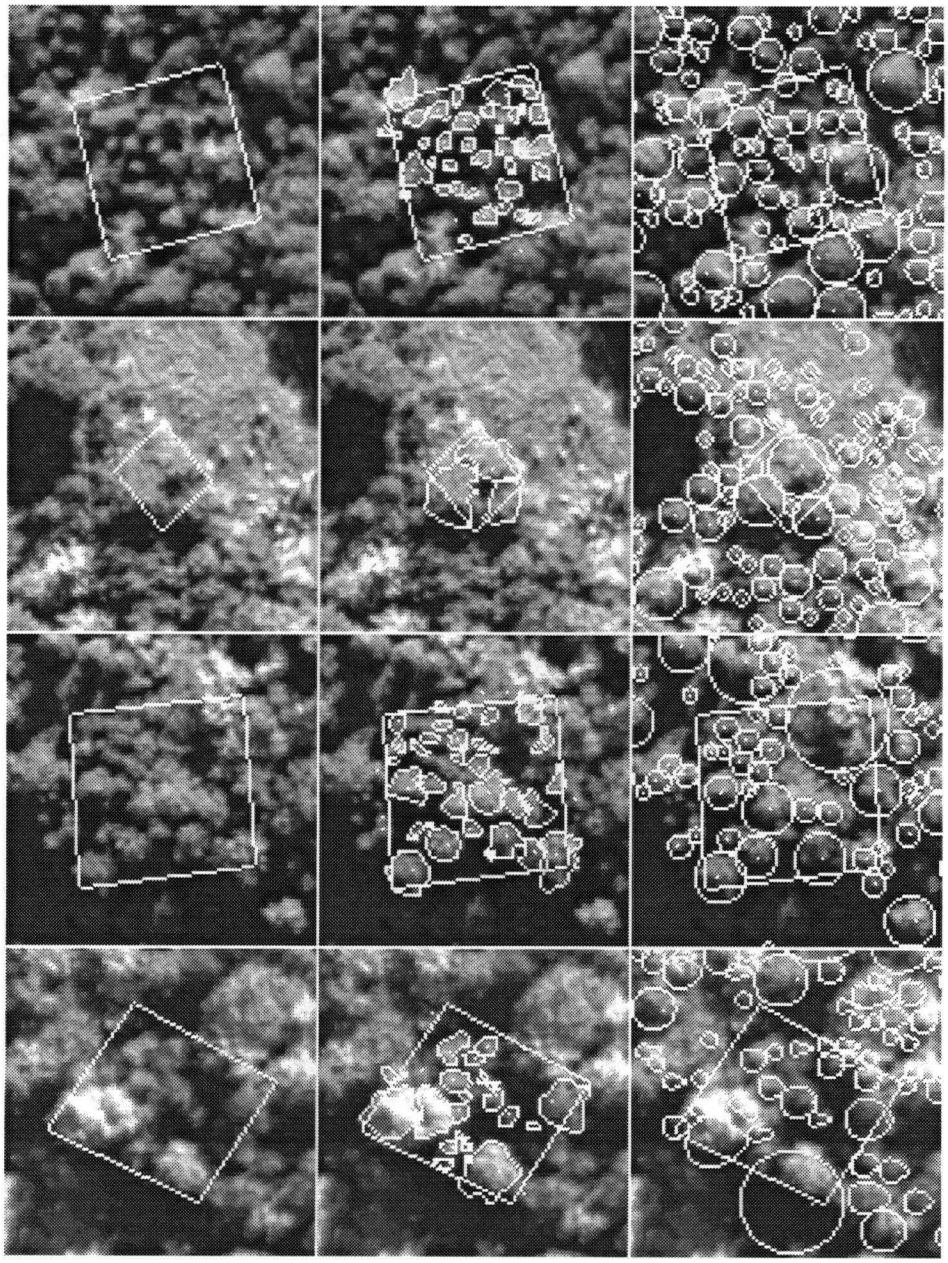


Figure A.42: Plot boundary, reference and results displays subimages for sample plots 13 (top) to 16 (bottom).

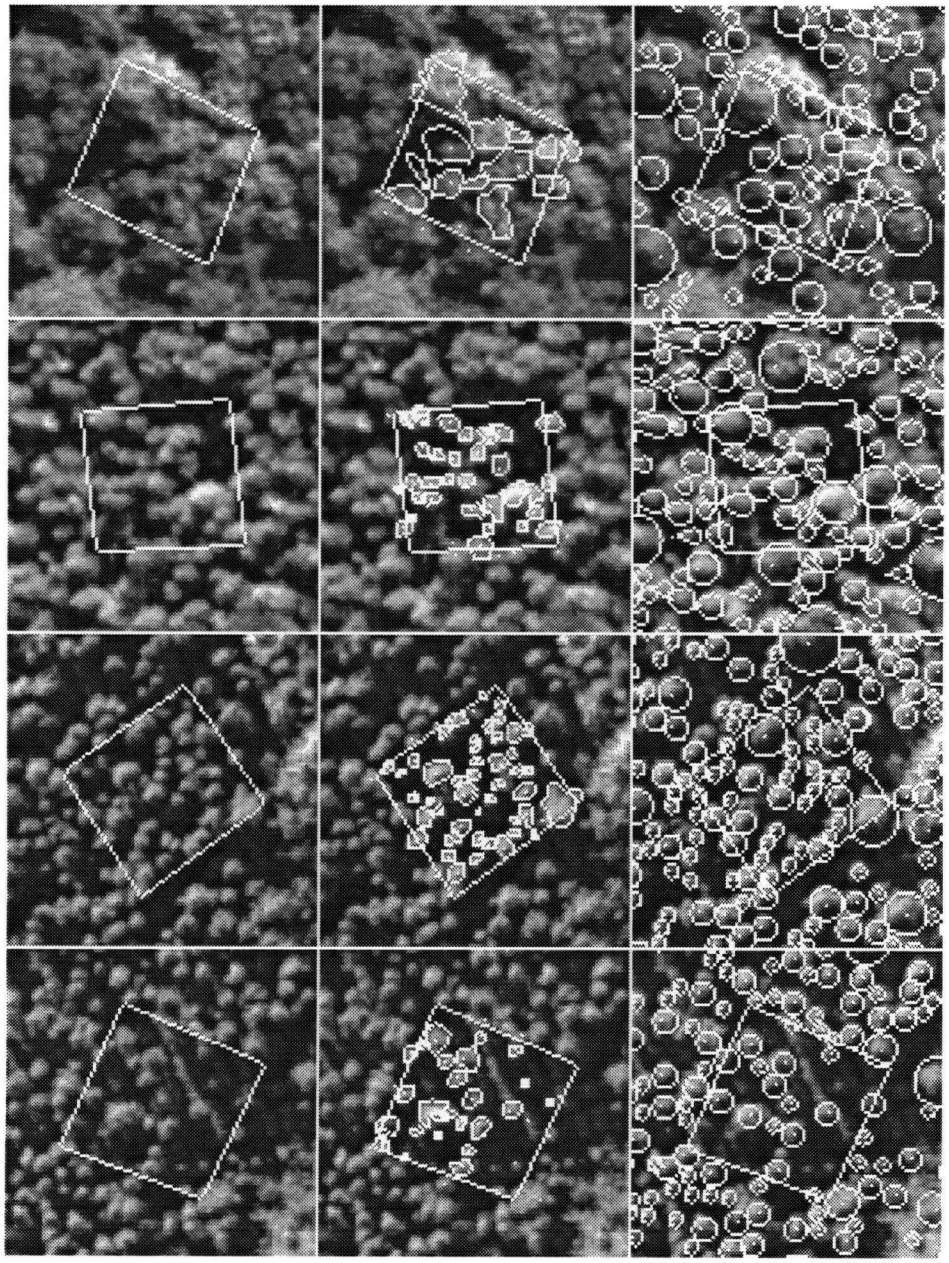


Figure A.43: Plot boundary, reference and results displays for sample plots 17 (top) to 20 (bottom).

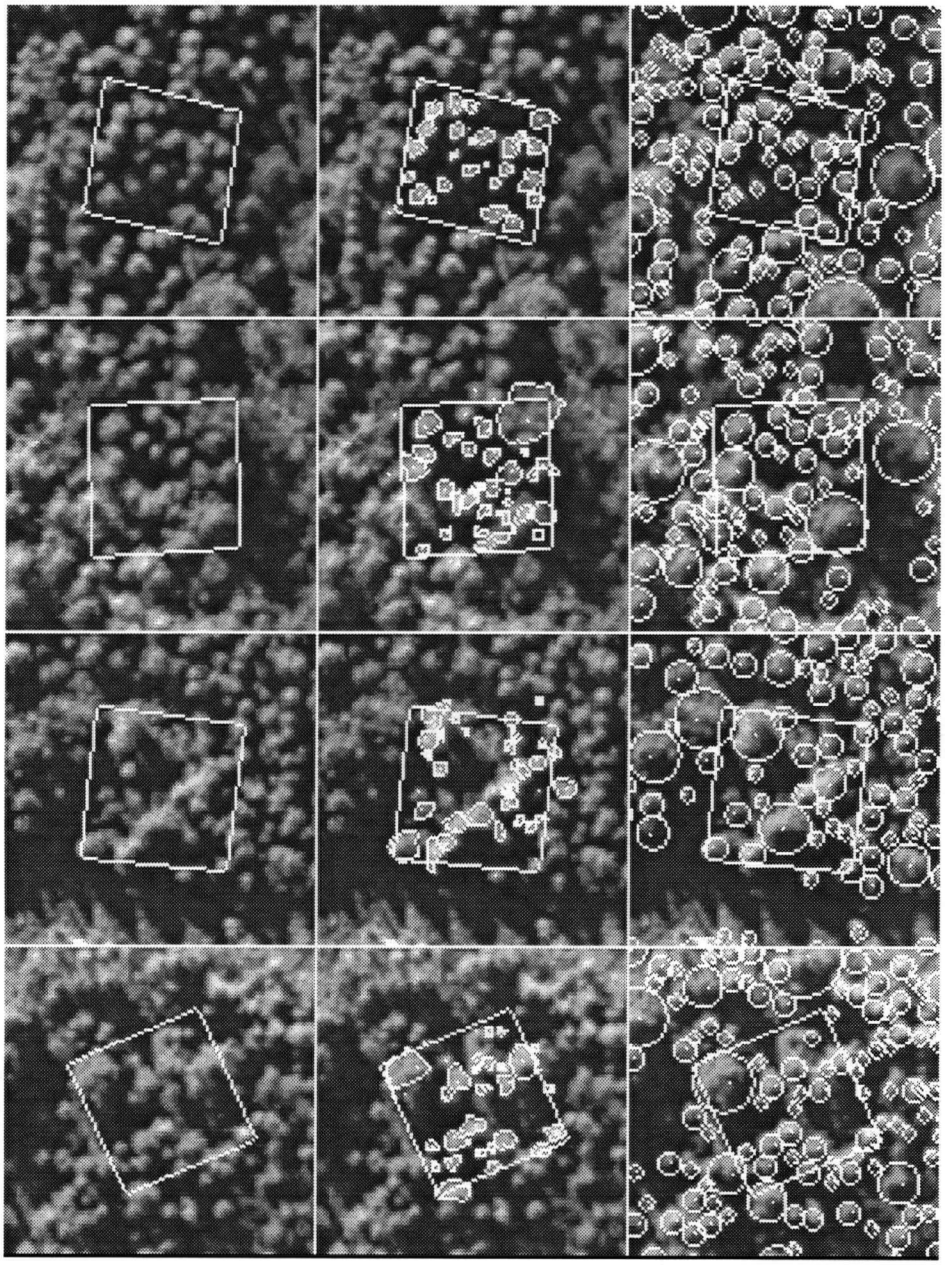


Figure A.44: Plot boundary, reference and results displays for sample plots 21 (top) to 24 (bottom).

Appendix B

Index to Term and Acronym Definition Page Numbers